Character Variation in Early Chinese Manuscripts

William G. Boltz | Seattle

In reading any Chinese manuscript, irrespective of whatever may or may not be its relation to texts known from the received tradition, the first step is to determine what words the characters of the manuscript stand for.[1] For transmitted texts, particularly those that have been written or edited in the mediaeval or modern periods, by which time the writing system had become in large measure standardized in the form familiar to us as the received orthography, the process of identifying what words the characters stand for is so automatic and so unconsciously done that we generally lose sight of the fact that what we are really doing is matching characters to words.[2] Formally, this is what 'reading a text' means. It is entirely natural in such cases to refer to the characters themselves as the 'words' of the text. They are not, of course, *sensu stricto* the words of the text; they are the written representations of those words.[3] But we can safely ignore this technical distinction and read transmitted texts

'automatically' precisely because these texts use characters in the standard, conventionally established way. That is to say, the texts are written in the standard writing system of the time, and we have learned to read that writing system. Having learned to read, we are free to remain unconscious of that part of the process involving the matching of character to word and consequent understanding of meaning. In most of what we read we do not expect, nor do we encounter, many deviations from the conventionally accepted standard, and so we do not have to think about the actual process of reading. This happy innocence does not extend to the matter of reading pre-Han or early Han manuscripts. To be sure, this does *not* mean that the orthography of those manuscripts is not systematic or conventional, that is, 'standard' within its own framework; it means instead that the framework itself, i.e., the set of rule-governed conventions of the orthography that makes it a workable writing system, is different to some degree from the standard framework with which we are familiar on the basis of the transmitted, received writing system.[4] The differences may be numerous and substantial or few and inconsequential, or something in between, varying from manuscript to manuscript. But whatever their extent, determining what those differences are lies at the heart of reading the manuscripts. Whatever idiosyncrasies, irregularities and apparent aberrations we may think we see in the orthography of early manuscripts, we must suppose that these actually are consistent with the rules and conventions to which the writing adheres overall.

---

[1] I am grateful to Matthias Richter and Michael Friedrich for very helpful comments, suggestions and corrections on earlier drafts of this paper. Remaining mistakes and infelicities are of course my own responsibility.

[2] Note that nothing in the process described as 'matching characters to words' restricts it to a 'one character to one word' isomorphism. While most Classical Chinese words are in fact monosyllabic and written with one character, not all are, and as the language evolves the proportion of two-character 'compounds' used to write what would linguistically be considered a single bisyllabic word increases. Conversely, in early manuscripts it is not uncommon to find instances of one character used to write a two-word phrase. These are the so-called *hé wén* 合文 'ligature graphs'. In the Baoshan manuscripts, for example, we find a single multi-component graph that can be transcribed as 戠, standing as a *hé wén* for the two characters 之歲 and presumably read as the two morphemes *zhī sùi* (in whatever the Chu Old Chinese pronunciation would have been.) See Zhang Shouzhong 1996, 234. Similarly, in the Guodian *Tai yi sheng shui* manuscript (among others) we find the two-word grammatical phrase 之所 regularly written with the *hé wén* graph 𠫭.

[3] There is an important sense in which the distinction between 'words, and 'written representation of words' is complex and multifaceted when we are concerned with the multiple functions of writing and the linguistic psychology and neurology of reading. From those linked perspectives 'written words' may well have a significance in their own right as durable, visual representations of language (not limited simply to speech) that distinguishes them from mere representations of spoken words. See, for example, the now classic work by David Olson, *The World on Paper* (Olson 1994), and the more recent work by Stanislas Dehaene, *Reading in the Brain* (Dehaene 2009). Important as this consideration is, for the analytical purpose of the present discussion we can set it aside.

[4] The term 'standard' can by definition be used only relative to something that is 'non-standard'. It is important to recognize that the orthography of early Chinese manuscripts is not imbued with a great measure of graphic arbitrariness or capriciousness, but adheres to a set of graphic conventions and rules just as any other writing system does, even if we cannot immediately discern all of those conventions and rules. Those conventions and rules in fact define a 'standard' for that writing system. Nevertheless, for convenience in our discussions here we will reserve the term 'standard' to refer exclusively to the transmitted, received Chinese writing system that we are familiar with from the Han on. Relative to this standard so defined, the writing system of the early manuscripts is 'non-standard,' i.e., not the same standard that we automatically assume when reading everyday texts. The differences between the received script and the variant forms found in pre-Han manuscripts may be slight enough to be called 'sub-standard' variations relative to the standard. See Haeree Park, who first draws explicit attention to this fact about Warring States manuscripts and their writing (Park 2009, *passim* and especially 318-24).

The extent to which we can take the standard, received writing system as a basis for identifying what words are intended by what characters in a given pre-modern manuscript will vary from manuscript to manuscript, but in most cases the orthography of pre-Han and early Han manuscripts will likely be somewhat unfamiliar in comparison with the received standard. Many characters will have the same graphic structure and same internal arrangement of components as characters known from the transmitted writing system and may be unfamiliar only to the extent that the manuscript shapes of the graphic components differ in the outward appearance of their formal execution from how the same components appear in the standard *kaishu* 楷書 script known from the Han on.

Examples:

是： (GD LZ.A, str 08, pos 19) *shì* 'this'[5]

祭： (*Baoshan*, str 237, pos 47) *jì* 'sacrifice'

君： (SH 3.4 *Peng zu*, str 04, pos 19) *jūn* 'lord'

命： (*Baoshan*, str 12, pos 25) *mìng* 'fate'

登： (*Baoshan*, str 27, pos 14) *dēng* 'ascend'

樂： (SH 5.7 *San de*, str 16, pos 30) *lè* 'pleasure' ~ *yuè* 'music'.

In these cases the task is simply to recognize the pre-*kaishu* form of the script. All other things being equal, such characters can be presumed to stand in their manuscript usage for the same words that they stand for in the received writing system, including so-called 'loan character' usages. Other manuscript characters will have structural forms that differ from anything known in the transmitted writing system and will therefore not be immediately graphically identifiable with standard characters. Identifying what word such characters stand for generally calls for ad hoc analysis and sometimes extended investigation, and is often speculative.

Examples:

 ：遣 (GD *LZ*.A, str 08, pos 14) for *dá* 'penetrate, pervade' 〈達〉?

 ：孚 (SH 5.2, *Bao Shuya yu Xie Peng zhi jian*, str 02, pos 43) for *miǎn* 'evade' 〈免〉?

 ：女 (SH 5.8a, *Gui shen zhi ming*, str 04, pos 16) for *ān* 'thereto, thereon' 〈安〉?

 ：旻 (SH 5.8a, *Gui shen zhi ming*, str 02 recto, pos 40) for *mò* 'sink, drown' 〈沒〉?

 ：殊 (GD *Yucong* 4, str 03, pos 09) for *shì* 'era' 〈世〉?

 ：緩 (SH 4.3, *Zhao wang hui shi*, str 02, pos 20) for *biǎn* 'ritual burial' 〈窆〉?[6]

And there will be numerous characters that fall somewhere between these two extremes; characters where, relative to their transmitted equivalents, the components are familiar but re-arranged, or where one or more components are missing, or where the manuscript character has one or more additional components in comparison with its presumed transmitted counterpart, etc.

Examples:

 (豊, GD *Yucong* 1, str 27, pos 02) *lǐ* 'ritual', for transmitted 禮.

 (邽, SH 4.3 *Zhao wang hui shi*, str 09, pos 26) *bāng* 'state', for transmitted 邦.

---

[5] The following set of abbreviations will be used in referring to published collections of early manuscripts: GD: Guodian 郭店 (Jingmen Shi 1998), SH: Shanghai Bowuguan上海博物館 (Ma Chengyuan 2001-2010), LZ: Laozi 老子.

[6] See Yang Zesheng (2005). I am grateful to Yang Li for pointing this out to me.

(胃, SH 2.1 *Min zhi fumu*, str 05, pos 05)
wèi 'refer to', for transmitted 謂.

(鼎, SH 5.8 *Gui shen zhi ming*, str 01, pos 03)
gŭi 'revenant', for transmitted 鬼.

(犀, SH 5.2, *Bao Shuya yu Xie Peng zhi jian*, str 3, pos 09) *xī* 'sacrificial ox', for transmitted 犧.

For all of these categories we must find ways to analyze the characters such that we can identify with some degree of confidence what word the character is intended to write. To do this we are in effect determining for ourselves what the users of the script knew implicitly, that is, what the rules and conventions of their orthographic system were. The 'rules and conventions' that govern any writing system apply basically at two levels: (i) those rules that pertain to the structure of individual graphs themselves (*wén zì* 文字), or to the combination of individual graphs into a single orthographic unit (*hé wén* 合文, often called a 'ligature graph'), and (ii) those that govern what word is linguistically allowable in a given context.[7] The 'system' part of any writing system is a 'given' simply by virtue of the fact that the writing is intended to be effectively shared by a community of users. What those users know in common that allows them to use writing to communicate with one another is nothing other than the set of rules and conventions that govern the structure and usage of the elements in their orthography, in other words, the features of the system that prescribe how characters may or may not be structured and the rules, based on the language being written, that determine what characters may or may not occur in a given written context. Depending on what the context is, the number of allowable characters may be relatively high, but it is limited all the same. This means fundamentally that a part of the way that any character X conveys its meaning is the fact that it fits with the characters of its immediately surrounding context to write a meaningful word, phrase or sentence. If a particular character X in isolation happens to be graphically difficult to distinguish from Y, that has only

a minor impact, if any at all, on the effective functioning of the writing system. Isolated occurrences are comparatively infrequent and atypical generally; writing systems evolve, it seems safe to assume, chiefly to write connected texts, not isolated words or letters. Most ambiguities or uncertainties about the identification of a given character will likely be resolved in context by virtue of the fact that only X and not Y is allowable, i.e., only X and not Y 'makes sense'.[8] We might call this the *context constraint* on the use of a graph X or Y within a given writing system.

Writing systems, no matter how seemingly complex, are finite in the number of their constituent graphic units. Any graph operates within the system in contrast to all other graphs. This means that a given graph does not have to be absolutely identifiable, but identifiable only relative to the other graphs in the system. In principle this means that 'X means X because it isn't Y' where Y is any other graph in the system. A character X in other words conveys its meaning, apart from whatever implications or indications its own graphic structure may carry, chiefly by not being Y. All Xs need not be precisely identical in their graphic shape or execution; they have only to be more like other Xs than like any Y in the system. Phrased somewhat more formally, we would say that a writing system includes a finite number of graphic *types*, any one of which we will call *X*. Any instance of a written example of *X* is a *token*, called for illustrative purposes *x*. The tokens are the visible, physical instances of a given character type; we can see them, count them, analyze them and copy them. Any *x* need not be precisely identical to any other *x*; it must only be identifiable as a token of *X* rather than of some other type *Y*. We can call the extent of variation among tokens of a single type the *token / type allowable latitude*. The notion of *allowable latitude* is a kind of functionally allowable free variation from two distinct, but complementary perspectives, that of the scribe's execution and that of the reader's perception of a character. Viewed in the light of what we called the *context constraint* above, recognizing this *allowable latitude* among variant tokens of a single type releases us from the burden of trying to identify minute graphic details as making a significant difference in what word is intended. We can in principle narrow the choices down, so to speak, in any given case by relying on the combined effect of the context constraint and the allowable latitude of the token / type relation together as setting out the limiting conditions for identifying graph with word. In practice,

---

[7] In English, for example, the frame 'What she _____ is good' allows (when restricted to a single word) only verbs in the blank space, and the frame (similarly restricted) '_____ person did not come' allows only articles, demonstrative pronouns, relative pronouns, the number 'one', the negative 'no', or the somewhat legalistic and formal 'said' in the blank. There are many more possible words that can fill the blank in the former frame than in the latter, but the number is all the same limited in both cases, and a writing system can take advantage of these kinds of contextual constraints just as usefully as the language does.

[8] I am tempted to claim that there is a law of 'orthographic natural selection' at work in the evolution of any writing system that would prevent the survival of an identical graphic form for two characters intended to stand for two different, unrelated words, when those two words could easily occur in the same context.

determining the allowable latitude of a token / type relation is an empirical matter, which unfortunately makes the process potentially circular.

For the practical matter of reading a manuscript, the problem amounts to this basic question:

How different can the ways of writing a given character be before they are no longer recognizable as writing the same word?

Phrased slightly differently, the question becomes:

How do we know when two differently written graphs are variant ways of writing the same word or instead write two different words?

The question has, for Chinese manuscripts at least, different implications depending on the kind of variation we confront and the context in which the variants occur. Matthias Richter has dealt at considerable length with the first of these two points. He has set out a clear distinction between two kinds of Chinese character variation that might be found within a single manuscript or among a group of relatively homogeneous manuscripts: (A) variation in the stroke shape or individual stroke configuration of the characters; this he calls 'non-structural variation', and (B) variation based on differing graphic components or different arrangements of the same graphic components, which he calls 'structural variation'.[9]

A considerable amount of effort has gone into trying to analyze and characterize this kind of variation, including such things as comparing the angles and lengths of certain strokes, noticing when matching strokes are straight or curved, hooked or not hooked, etc. in variant scriptions of the 'same' character. The goal is to determine what the orthographic rules or conventions might be and how much latitude we can expect in the writing of such characters before they no longer are recognized as writing the same word. Richter's suggestion that we try to identify what he calls a 'profile' of graphic variation for a single manuscript, while discussed initially largely in regard to instances of structural variation, can equally usefully be applied to non-structural variation as well. In this approach we might think of any graphic 'profile' that we may be able to discern in a given manuscript as a 'manuscript context', on a par with lexical context of the kind mentioned above, as an additional strategy for determining the words written. It may turn out that the hand-writing profile of a given manuscript, that is the 'manuscript context', is sufficiently distinctive to determine a particular reading in the same formal way that a lexical context often is; and further, a recognizable profile of non-structural variation may allow the identification of a single scribe as responsible for a number of different manuscripts, thus providing a basis for a historical grouping of the manuscripts in question together as coming from a common source.[11]

For example: (A) non-structural variation:
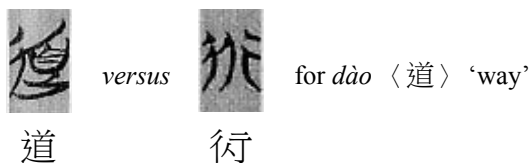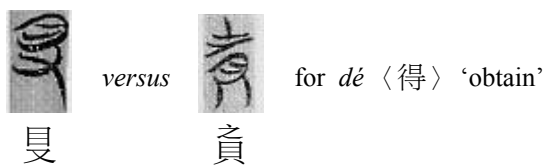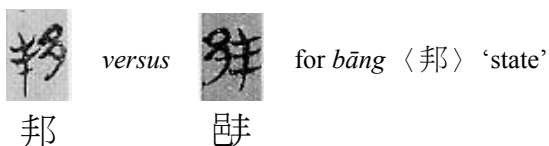
 versus for 冬 〈終〉 *zhōng* 'end',

 versus for 長 *cháng* 'long',

 versus for 才 〈在〉 *zài* 'to be located',

 versus for 為 〈為〉 *wéi* 'to act'.[10]

(B) structural variation:
i.

 versus for *tiān* 'sky'
天 　 夭

 versus for *shàng* 'above'
上 　 上

 versus for *xià* 'below'
下 　 下

---

[9] See Richter 2005.

[10] These manuscript forms and the ones in part (B), structural variation, below are taken from Matthias Richter's Guodian manuscript character tables (http://www.aai.uni-hamburg.de/MPC/datab.html).

[11] The possibility of such groupings is up to a point analogous to identifying diviner groups of the Shang divinatory inscriptions. The big difference is that Shang inscriptions typically include the diviner's name, which can be correlated fairly directly with calligraphic style, whereas Warring States and Han period manuscripts only rarely mention scribes by name. See Richter 2005, 175-82.
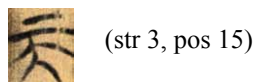
ii.

versus for *jī* 〈奇〉 'to make a lateral attack'

敧   敆

 versus *versus* for *měi* 〈美〉 'attractive'

𡵂   敽   敿

 versus for *bāng* 〈邦〉 'state'

邦   㘣

 versus for *dé* 〈得〉 'obtain'

昗   賣

 versus for *dào* 〈道〉 'way'

道   衍
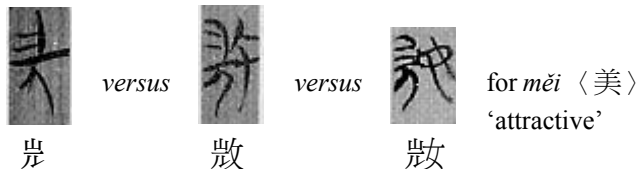
There would seem to be little doubt that the pairs of variants in the (B.i) set are different ways of writing the same word. In the (B.ii) set it is much less obvious that the variation is between two ways of writing the same word. This has to be determined generally on a case by case basis, until the orthographic conventions of the manuscript(s) in question have been identified. Structural variation also includes, in its absolute form, entirely different characters used to write the same word. This we might call absolute structural variation, exemplified here by 𡵂 / 敽 / 敿 vs. 美 for *měi* 'attractive' and 衍 *versus* 道 for *dào* 'way' in the (B.ii) set above, where the various manuscript characters seem not to have any component in common at all with the character used in the received orthography for the same word.**[12]** Such cases

---

[12] The two different ways shown here to write the word *dào* 'way' do *sensu stricto* have one component in common, viz., the left side component, 彳, of 行, which is historically the same as the three-stroke upper part of 辵 (*Kangxi* classifier 162, combining form ⻌). This is clear from the manuscript forms of the two characters, but becomes completely obscured in the received writing system. To be sure, we find unambiguous manuscript testimony to apparently unconditioned graphic variation between the component 辵 and the component 彳. This would suggest that the 彳 in 行 is tantamount to 辵 (⻌) in 道. Beyond this, the variation of the second component, 人 [*rén*] 'person' in 行 and 首 [*shǒu*] 'head' in 道 is categorically consistent
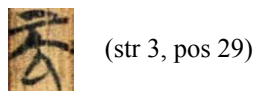
---

inevitably raise the question of whether the variation might be lexical instead of graphic. In fact there remains always the possibility, at least in principle, that any pair of variants, no matter how similar or different each is to the other, write different words and represent therefore lexical variation, not graphic. Intuitively, we are likelier to be drawn to the possibility of lexical variation in cases of absolute structural variation than in those of merely partial structural variation, an intuition that would probably be borne out statistically. For our purposes here, in these examples and in the ones to follow, we are assuming that the fact of graphic variation as opposed to lexical has been established independently of the features we discuss.

Consider next these two groups of character tokens from the Shanghai Museum manuscript called *Gui shen zhi ming* 鬼神之明 (SH 5.8):
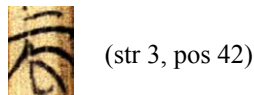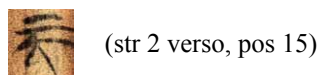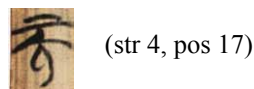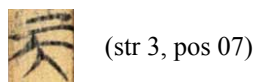
1. *ér* 而 'then':

 (str 3, pos 15)

a

 (str 3, pos 29)

b

 (str 3, pos 42)

c

 (str 2 verso, pos 15)

d

 (str 4, pos 17)

e

2. *tiān* 天 'sky, heaven':

 (str 3, pos 07)

a

---

semantically insofar as both components are 'body part' terms (understand 人 as 'body' proper). These two facts about the variation between 道 and 衍 as ways to write the word *dào* 'way' together suggest that we might not want to call this an example of 'absolute' structural variation. I am grateful to Matthias Richter for drawing my attention to the need to analyze this example fully. For further discussion of this example, including comments on what seems to be still another variant, written 衛, see Richter 2003, 5-8.

 (str 3, pos 21)

b

 (str 1, pos 30)

c

 (str 1, pos 38)

d

and these, both from strip 4 of the Shanghai Museum manuscript called *Heng xian* 互先 (SH 3.3):

3.a (str 4, pos 26) *ér* 而 and

3.b (str 4, pos 12) *tiān* 天.

One might be excused for thinking on a first look that all nine characters of the *Gui shen zhi ming* strips and both of those from the *Heng xian* strip were tokens of the same type, but in fact the group 1 and 3.a examples are, as indicated, tokens of the character type 〈而〉 and the group 2 and 3.b examples are all tokens of 〈天〉. Once these are seen in their respective contexts, all ambiguity disappears.
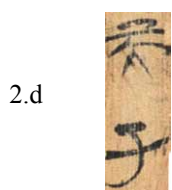
For example, 1.a occurs in the phrase
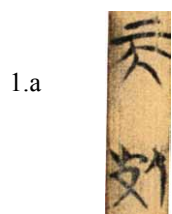
1.a 而死 'then died'

a context in which the word *ér* 'then' is natural, but where *tiān* 'sky, heaven' makes little sense, while 2.c occurs in the phrase
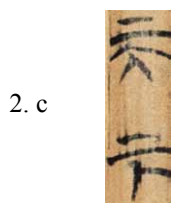
2. c 天下 'sub-celestial realm'

and

2.d 天子 'Son of Heaven'

in lexical contexts where clearly the word in question is *tiān*. In the *Heng xian* strip example 3.a occurs in the phrase

3.a 同出而異生 [for standard 性] 'having a common origin, but different nature',

and 3.b occurs in:

3.b 清嬰 [for standard 氣] 生天 'pure pneumas give rise to heaven'.

Phrase 3.a is straightforward with *ér* 'then, but', but odd to the point of incomprehensibility if read with *tiān* 'sky, heaven'; phrase 3.b is just the reverse. In particular there can be little doubt about the understanding of 3.b because it is parallel to the immediately preceding line 濁氣生地 'murky pneumas give rise to the earth'.[13] Context is the crucial determining factor in distinguishing graphically similar, if not nearly identical, characters such as these. Once that is recognized, the token / type allowable latitude constraint becomes largely inconsequential. That is, it no longer matters in these cases that the tokens of *ér* 〈而〉 and the tokens of *tiān* 〈天〉 are nearly indistinguishable in isolation.

One fairly obvious feature easily seen in the above examples is that primary horizontal strokes often show up with a secondary, slightly shorter parallel horizontal twin stroke added to the 'outside' of the primary stroke. The examples given above to illustrate the simplest kind of structural variation, those in set B (i) as well as some of the the

---

[13] All of these manuscript scans have been made from the Shanghai Museum publications of the manuscripts in question. The *Gui shen zhi ming* manuscript is the eighth manuscript in volume five, and the *Heng xian* manuscript is the third in volume three. For an exhaustive listing of tokens of *ér* 而 in the Shanghai Museum manuscripts published in volumes one through five, see Li Shoukuei 李守奎 et al. (2007), 447–56, and for *tiān* 天, op. cit., 2–5.

ér 〈而〉 tokens, are examples of this secondary horizontal stroke. This turns out to be a very common feature in third-century BCE. manuscripts, and characters written with this secondary horizontal stroke have long been recognized as in free variation with the same characters written without it. The received orthography from Han times on does not include characters with this 'extra' secondary horizontal stroke. Consider now these additional examples of the character for *ér*:[14]

     4. *ér* 而 'then':


     a   b   c   d
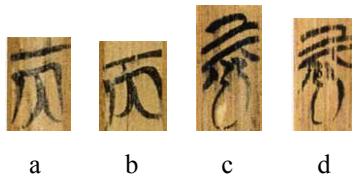
Clearly there is more than just the secondary horizontal stroke that distinguishes 4 c and d from 4 a and b. And these four examples are as a group markedly different in appearance from those given in line 1. Apart from the inconsequential secondary horizontal stroke in 4 c and d, none of these variants is structural in Richter's sense, but all of them are noteworthy as hand-writing features of these manuscripts. We could debate what type these characters are tokens of, and the question then would become a textual criticism matter. Which of the characters given in the preceding lines one through four are tokens of the graph 〈天〉 and which of 〈而〉? This kind of decision, as we said, has to be made on a case-by-case basis, generally on the basis of context. Once the context constraint has pointed us in the right direction for determining what word is intended in each case, we end up with a total of ten tokens for the character type *ér* 〈而〉 and five for *tiān* 〈天〉.

     The next question is how do we choose a graph to represent the type? How do we determine what graphic shape suggested by the available tokens we should designate as representative of the type? Our first inclination might be to identify the standard form of the character that we know from the received writing system as representative of the type, and then every instance in a particular manuscript, or in a group of manuscripts, of that character is a token of that type, irrespective (up to a point, see below) of how it may diverge from the shape or structure of the type. This recourse is often the default solution to the problem by modern editors who print transcriptions of these manuscripts in modern books. As a practical matter, this is not an unreasonable approach to the problem and is easy to defend. (It is what I have done above for 〈天〉 and 〈而〉 here.) But if our concern is with a more historically accurate

and theoretically well-informed analysis of the writing system, then this solution falls short in at least three respects: first, it is anachronistic as far as an analysis of the early writing system is concerned, because it resorts to what we know about the later, standardized script to identify the type; second, it fails to recognize that some tokens look more like others than other tokens do, – in other words, the tokens can be grouped according to their own internal graphic appearance; and third, it still doesn't tell us anything about how different two tokens can appear and still be instances of the same type.

     The first of these three shortcomings, apart from the theoretical concern, usually will not present any real problem in the majority of cases. And it is hard to imagine how *any* choice for a type could actually overcome the third shortcoming. The second is really the only substantive issue, and is the easiest to accommodate. We need only to recognize an intermediate level between type and token, which we can call *sub-type*. Consider the different ways of writing the verbal negative adjunct *bù* 〈不〉 in the Guodian manuscripts, for example, which we can array in four distinct groups as follows:[15]

     a   b

     c   d
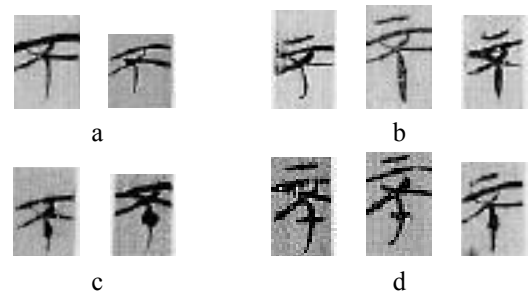
As suggested by the groupings, I would identify four sub-types: a 〈不〉, b 〈𣎴〉, c 〈不〉 and d 〈𣎴〉. A sub-type is defined by a distinctive graphic form shared by some, but not all, tokens of a given type. The relation between a sub-type and its tokens will be graphically closer than that between all tokens generally and a type, and its form is as a consequence less difficult to define or establish than the form of the type. The *ér* 'then' 'sub-type – token' relation has less indefiniteness than the 'type – token' relation. But a sub-type is, all the same, an abstraction, a mental construct, in the same way that a type is. And any visible, tangible representation or characterization of either a type or a sub-type is an artificial, in some sense even a superficial, instantiation of an abstraction. From this perspective, how we choose to represent the sub-type or type graphically is entirely arbitrary. The only caution that must be observed is

---

[14] Examples 4a and b both are from the Shanghai *'Zi yi'* ms., 4c and d are both from the Shanghai *'Wu wang'* ms.

[15] These examples are all taken from the data assembled by Matthias Richter in his Guodian Characters data base, which includes many more examples that can be classified structurally into the same four categories. See http://www.aai.uni-hamburg.de/MPC/datab.html.

not to allow our choice inadvertently to suggest a misleading historical link or developmental process.

The question remains, what does it mean to say that the 'same' character appears in another form, forms that can be written as differently from one another as seen in the examples of *ér* 而 and *bù* 不 given above? What does it mean to be the 'same' character when in visible, discernible fact the characters look different, sometimes quite different, from one another? Efforts to account for these kinds of variation by measuring stroke angles, line lengths, degree of curvature, presence or absence of hooks, etc., as useful as they may be in some respects, will not in the end alone be sufficient to answer the question of how different the characters can appear before they no longer are recognized as writing the same word.

Characters are written to be read. Apart from a reader, that is to say, in the absence of someone who recognizes what a character is supposed to write, no graph of any kind writes anything. Any graph that functions as glottographic writing functions *ipso facto* as an element in a writing system. And as we have already pointed out above, the rules of the system play a part in conveying the meaning of any given graph that is used as a part of that system. In order for a reader to recognize a character as writing a word X, he must have a knowledge of the writing system to which that character belongs and the rules that govern how it operates. This is the 'background knowledge' against which the recognition of a character as writing word X takes place. Most readers, of course, will not be conscious of this background knowledge, until someone points it out, but it is there all the same as a necessary condition for reading. Imre Galambos makes the same point in a slightly different way when he says that:

> [W]ith the absence of a standard form that could serve as an abstract form of a particular character, the solution to the problem of identification lies in determining the relationship of a character form with the word it stands for. Because written characters are graphical representations of words of a language, they are meaningful only in reference to the word they have been chosen to represent.[16]

To be sure, not all problems of graphic variation will be easily resolved. The foregoing examples have been deliberately chosen to illustrate the discussion in the clearest and least ambiguous way. Many such questions will entail more uncertainty than these examples have shown, sometimes a lot more. But that does not change the fact that writing systems work in principled ways.

We are in a better position to deal with problems of graphic variation if we recognize how the writing system operates and how graphs within it function than if we allow ourselves to overlook this aspect of what we mean by writing.

Abbreviations

GD : Guodian, i.e., Guodian Chumu zhujian, see Jingmen Shi 1998.

LZ : Laozi.

pos : position (in a strip).

SH : Shanghai, i.e., Shanghai Bowuguan cang Zhanguo Chu zhushu, see Ma Chengyuan 2001-2012.

str : strip (of a bamboo manuscript).

REFERENCES

Dehaene, Stanislas (2009), *Reading in the Brain* (New York: Viking).

Galambos, Imre (2006), *Orthography of Early Chinese Writing: Evidence from Early Chinese Manuscripts* (Budapest: Eötvös Loránd University; Budapest Monographs in East Asian Studies, 1).

Jingmen Shi Bowuguan 荊門市博物館 (1998), *Guodian Chumu zhujian* 郭店楚墓竹簡 (Beijing: Wenwu chuban she 文物出版社).

Li Shoukuei 李守奎 et al. (2007), *Shanghai Bowuguan cang zhanguo Chu zhushu (1－5) wenzi bian* 上海博物館藏戰國楚竹書 (一－五) 文字編 (Beijing: Zuojia).

Ma Chengyuan 馬承源, ed. (2001–2012), *Shanghai Bowuguan cang zhanguo Chu zhushu* 上海博物館藏戰國楚竹書, Volumes 1-9. (Shanghai: Guji chuban she 古籍出版社).

Olson, David (1994), *The World on Paper* (Cambridge: Cambridge University Press).

Park, Haeree (2009), *The Shanghai Museum Zhou yi manuscript and the Warring States writing system,* Ph.D. dissertation (University of Washington).

Richter, Matthias (2003) 'Suggestions Concerning the Transcription of Chinese Manuscript Texts – A Research Note', *International Research on Bamboo and Silk Documents: Newsletter*, 3, 1: 1–12.

Richter, Matthias (2005), 'Towards a profile of graphic variation', *Asiatische Studien / Études Asiatiques,* 49, 1: 169–207.

Richter, Matthias (n.d.), *Guodian Characters Data Base,* (http://www.aai.uni-hamburg.de/MPC/datab.html).

Yang Zesheng 楊澤生 (2005), '*Du* Shangbo si *zha ji* 讀 上博四 札記' (http://www.jianbo.org/admin3/2005/yangzesheng001.htm).

Zhang Shouzhong 張守中 (1996), *Baoshan Chujian wenzi bian* 包山楚簡文字編 (Beijing: Wenwu chuban she 文物出版社).

---

[16] See Galambos 2006, 77.