FRANÇOISE BOTTÉRO AND CHRISTOPH HARBSMEIER

The *Shuowen Jiezi* Dictionary
and the Human Sciences in China

INTRODUCTION

In this paper we aspire to reconstruct whatever we can of the process of the composition of the *Shuō wén jiě zì* 說文解字 (100 AD).[1] Our ambition is not to interpret the textual product, but to explain the process of the production of the text; like Wilhelm von Humboldt, we are interested not only in the *ergon* but in the *energeia*. Moreover, we intend to define the important place of the *Shuōwén* in the history of scientific thought in China.

The *Shuowen* is not a dictionary of basic meanings of words. It is a dictionary of graphic etymology, and etymology needs to be carefully distinguished from semantic analysis. It provides only meanings that are relevant to the explanation of the graphs used to write words. Similarly, the *Shuowen* only refers to pronunciations of graphs insofar that these are relevant to the explanation of phonetic constituents in those graphs.

---

[1] The most useful editions we have consulted are the following: Xu Xuan 徐鉉 (917–992), *Shuowen jiezi* 說文解字 (Beijing: Zhonghua shuju, 1963); Xu Kai 徐鍇 (920–974), *Shuowen xizhuan* 說文解字繫傳 (Beijing: Zhonghua shuju, 1987); Duan Yucai 段玉裁 (1735–1815), *Shuowen jiezi zhu* 說文解字注 (Shanghai: Guji chubanshe, 1988); Ding Fubao 丁福保 (1874–1952), *Shuowen jiezi gulin* 說文解字詁林 (Zhonghua shuju, 1988); Tang Kejing 湯可敬, *Shuowen jiezi jin shi* 說文解字今釋 (Changsha: Yuelu shushe, 1997). In addition, see these studies: Roy Andrew Miller, "Problems in the Study of the *Shuo-wen chieh-tzu*," Ph.D. dissertation (New York: Columbia University; rpt. Ann Arbor, 1953), Françoise Bottéro, *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, Institut des Hautes Etudes Chinoises, 1996), Richard Sterling Cook, "說文解字, 電子版 *Shuo Wen Jie Zi, Dianzi Ban*: Digital Recension of the Eastern Han Chinese Grammaticon," Ph.D. dissertation (Berkeley: University of California, 2003). For a classified bibliography on *Shuowen* see Dong Xiqian 董希謙 and Zhang Qihuan 張啓煥, *Xu Shen yu Shuowen jiezi yanjiu* 許慎與說文解字研究 (Kaifeng: Henan daxue chubanshe, 1988), pp. 175–227, and also Cook, *Shuo Wen Jie Zi*, pp. 451–92. For the more general context in the history of Chinese linguistics, see Christoph Harbsmeier, *Language and Logic*, in Joseph Needham *Science and Civilisation in China*, vol. 7, part 1 (Cambridge: Cambridge U.P., 1998).

When the author, Xǔ Shèn 許慎,[2] adds alternative pronunciations, all of these are construed as possibly graphologically relevant, and when he adds alternative meanings with or without sources for them, these are again construed as being in principle similarly graphologically relevant. Such lists *never* attempt to provide any comprehensive overview of the different meanings of words written by the graph under discussion. And moreover, the meanings indicated in his glosses are very often not the basic meanings of words in question, but those meanings that Xu Shen considers serve best to explain the structure of the graph. When *suǒ* 所 is explained as the sound of an axe as it hits a tree, Xu Shen was surely aware that this is less than helpful for a proper understanding of the character in texts, but he wishes to insist that the structure of the graph is best understood when one considers this exceedingly rare meaning of the word. (We shall revert to the methodological importance of this strategy, below.)

Words have many meanings, and in Xu Shen's time Chinese characters were very often used to write several words with different pronunciations. *Shuowen* is not concerned with this phonetic and semantic variety and it is thus neither a dictionary of pronunciation nor indeed a dictionary of the meanings, let alone the basic meanings, of characters.

For example, the character 說 has three common readings which come out in modern Mandarin as *shuō, shuì* and *yuè*. But it appears that as far as Xu Shen is concerned, the graphologically relevant pronunciation is *shuō*, and the relevant meaning is the speech act of explaining rather than the psychological state of satisfaction or delight. Nonetheless, in his book he remains free to use the character 說 as everyone else does – to write the word later standardly spelt 悦 *yuè* ("be satisfied/delighted") or even to write the word *shuì* ("persuade").

A much more complex example of a very different kind and from another radical may further illustrate this often overlooked point: the modern graph 賣 has the reading *yù* (Dà Xú běn *fǎnqiè* 大徐本反切 : 余六切) as a phonetic constituent always throughout *Shuowen*. Xu Shen must have been aware that another very common reading of this character was *mài,* but this was of no concern to him in the phonetic analysis presented in his dictionary. In the small-seal script the word *yù* ("sell in the street as a hawker") is similar but clearly different from the graph *mài* 賣 ("offer for sale"). These were different words written with dif-

---

ferent small-seal graphs, but the *kǎishū* 楷書 graphs became conflated already in Hàn times to create the misleading impression on first sight that the same character has two separate pronunciations and two similar but not identical meanings.

We shall revert below to the theoretical importance of the seal script for Xu Shen's analysis of Chinese characters, and complex stories like that of the graph 賣 illustrate better than anything else why Xu Shen needed to base his graphological dictionary on the small seal script and not on the current clerical script of his time.

Each *Shuowen* entry is compiled according to a set of rules, the implicit *tǐlì* 體例 ("editorial procedure") of *Shuowen*. (We have no *fánlì* 凡例 "explicit statement of the editorial procedure".) Xu Shen is not always consistent, but it is possible to reconstruct the overall methodological *régime* which he tried to impose on the whole work. His introduction (or Postface) can be used as a point of departure for the reconstruction of his methodological *régime*, but, as we shall see, the rules he followed were more elaborate than those stated explicitly in his Postface.

In writing his dictionary, Xu Shen built on the Chinese commentarial tradition. The numerous quotations from canonical and noncanonical texts in *Shuowen* provide ample evidence for this historical link. However, his dictionary is about characters as such and not about characters in context. In modern terminology: Xu Shen was concerned with the system of the *langue*, and not with *parole*. His interest was with the writing system of the language as such and no longer, as in the commentarial tradition, with individual occurrences of characters in given texts. In its focus on the system of the *langue*, *Shuowen* was preceded by the *Ěryǎ* 爾雅 (ca. 3d c. BC), which already discussed decontextualized words rather than occurrences of words in given contexts, although its occasional concern, in one chapter, with the *Shījīng* is evident enough.[3] We might call this theoretical analytic concern with the structure of graphs "graphological" in analogy to the notion of "phonological" analysis. In order to avoid a confusion with the ordinary meaning of "graphology" we have often decided in favor of the term "graphemic" (and the derivative noun "graphemics"). By contrast, we would call "graphic" the perspective of an epigrapher or a historian of calligraphy. We reserve the awkward term "graphological" for an analysis that goes beyond identifying graphemes and goes on to discuss

---

[3] See particularly the "Shixun 釋訓" section of the *Erya* and its close connections with the Mao tradition of the *Shijing*. For *Erya* one may now consult the splendid Zhu Zuyan 朱祖延, *Erya gulin* 爾雅詁林 (Wuhan: Hubei jiaoyu, 1996).

the nature and structural interaction in characters. Thus for example, explaining a graph as "referring (pictorially) to something" is not a graphemic analysis but a graphological interpretation.

THE COMPOSITION OF THE LEXICAL ENTRIES

The entries of the *Shuowen* follow an invariant schema:

1. Head Graph. (Obligatory, always in seal script.)
2. Semantic gloss as relevant to the graphological analysis. (Obligatory. This often provides a marginal, by no means basic, meaning of the word glossed. Optionally, the gloss may be followed by alternative glosses Xu Shen has found in the literature and wishes to record.)
3. Graphological analysis into semantic and phonetic constituents. (Obligatory. Only optionally attention is paid to the dual semantico-phonetic function of phonetic constituents: many obvious cases go unidentified.)
4. Note on graphemically distinct allographs. (Optional.)
5. Supplementary/encyclopaedic sundry material. (Optional.)
6. Subsumption formula. (Obligatory in "radicals," even when no other characters are in fact subsumed under them, but never present anywhere else.)
7. Notes on pronunciation. (Optional. These are present in approximately 10% of the characters and are regularly at the end of the entry.)

The sequence of the obligatory elements listed above tends to be invariant throughout. In general, elements that are obligatory for all entries precede optional elements, but occasionally encyclopaedic material may be entered directly after the semantic gloss, presumably because encyclopaedic remarks, being concerned with semantics, are naturally attached to the graphological semantic gloss.

Xu Shen's general discussion of the graphological system is well-known under the title *liù shū* 六書 "Six (Categories of) Scribal Acts" which he takes over from earlier Han tradition, and which he discusses briefly in the Postface of *Shuowen*. The translation "scribal act" is awkward here, because no reference to any scribes is intended. What we mean by "category of scribal act" is this: "the type of act involved in creating a graph."[4] These acts are referred to by verbal expressions. Graphs are said to:

---

[4] Concerning the complex issues raised by such terms as *liu ben* 六本, *liu shu* 六書 and *liu wen* 六文, see F. Bottéro "A New Perspective in the Six Ways of Graphically Representing Words" (forthcoming).

1. *zhǐ shì* 指事 "refer (pictorially) to something" as in the characters *shàng* 上, *xià* 下;

2. *xiàng xíng* 象形 "symbolize physical shape" as in the characters *rì* 日, *yuè* 月;

3. *xíng shēng* 形聲 "indicate shape and sound" as in the characters *jiāng* 江, *hé* 河;

4. *huì yì* 會意 "associate ideas" as in the characters *wǔ* 武, *xìn* 信;

5. *zhuǎn zhù* 轉注 "turningly gloss > reinterpret (?)."[5] Traditional examples: *kǎo* 考, *lǎo* 老;

6. *jiǎ jiè* 假借 "borrow [one graph for another]" as in the characters *lìng/lǐng* 令, *cháng/zhǎng* 長 (this probably refers to different meanings being attributed to a single character under different readings, and then more broadly for the use of a character current for a word X, for a phonetically similar word Y).

The literature on the *liu shu* is extensive, and received opinion on this crucial matter takes it for granted that *shū* 書 refers to characters or kinds of characters. See, for example, Táng Lán 唐蘭,[6] and the still authoritative English version of Qiú Xīguī 裘錫圭.[7] In fact, as far as we know, *shu* 書 never refers to written objects other than documents. Graphs or characters are called *wén* 文 and *zì* 字, respectively.[8]

Xu Shen's descriptions of the various *shu* 書 are verbal rather than nominal, but the detailed interpretation of several types of scribal acts remains unclear: for example we have yet to see any convincing account of *zhuan zhu.* There is no need here to go into controversial detail concerning the *liu shu* at this point, except to note that these six categories are not in any way mechanically or even sporadically imposed in the body of the dictionary itself, where only occasional explicit mention is made *en passant,* for example, that a character "symbolizes physical shape" (*xiang xing* 象形). Suffice it to say that in the case of what was traditionally said to *jia jie* (that is, to "borrow [one graph for another]"), Xu Shen is simply not concerned with these in the main body of the dictionary: what he explains is the nonborrowed source character in its original meaning and only very occasionally (under *wéi* 韋, *xī* 西,

5 Our interpretation of this category is no more precise or definitive than the Chinese traditional interpretations are themselves.

6 Tang Lan 唐蘭, *Zhongguo wenzixue* 中國文字學 (Shanghai: Guji chubanshe, 1979), pp. 67 ff; and *Guwenzixue daolun* 古文字學導論 (Jinan: Qilu shushe, 1981), pp. 85 ff.

7 *Chinese Writing* (Berkeley: The Society for the Study of Early China, 2000), pp. 151–63; and the Chinese version: *Wenzixue gaiyao* 文字學概要 (Beijing: Shangwu yinshuguan, 2000), pp. 97–104.

8 For a detailed presentation of this terminology see Françoise Bottéro, "Revisiting the *wen* and the *zi*: The Great Chinese Characters Hoax," *BMFEA* 74 (2004), pp. 14–33.

néng 能, *fèng* 鳳) will he refer to loan-borrowing phenomena that are so common in ancient Chinese writing.

Writing a *Shuowen* entry was like filling in a form of the sort we have just outlined above and not a general implementation of the ideas presented in the Postface. As we have seen, in this form certain fields are obligatory and others are optional. Identifying the underlying pattern for the Shāng dynasty oracle bone inscriptions was a decisive breakthrough in oracle bone inscriptions philology.[9] Specifying such a schema for the *Shuowen* must be regarded as a *conditio sine qua non* for any disciplined and systematic study of the text.[10]

We leave open the question to what extent the schematic organization of *Shuowen* was linked to any bureaucratic organization of its compilation by a team of collaborators. Xu Shen made an intellectually decisive move from the philological interpretation of characters to a systematic science of the structure of graphs. Later, others made the similarly exciting move from the philological collection of sound glosses in the direction of a systematic science of phonology.[11] It was the systematicity and discipline of approach that has assured *Shuowen* a permanent place in Chinese history.

Imperial Han dynasty taboos override the general rules on obligatory elements. When the head graph of an entry happens to be the name of a deceased emperor, Xu Shen felt obliged, as a public servant writing a book to be handed up to the current emperor, not to give any gloss but to write instead: *shàng huì* 上諱 ("Taboo because of the emperor"). Elsewhere, such absence of obligatory elements is often marked by the explicit formula *quē* 闕 ("[information] missing"). In this, Xu Shen follows the venerable and very "scientific" tradition inaugurated in *Lúnyǔ* xv/26. The story of this self-critical nonomniscient scientific pose expressed by the term *quē* is an important part of Chinese intellectual history. We should even say that it is a significant part of the history of scientific thinking, and the history of logical methods.

---

[9] See David Keightley, *Sources of Shang History: The Oracle-Bone Inscriptions of Bronze Age China* (Berkeley: U. California P., 1978).

[10] Compare the traditional schema 字義字形字音 ("meaning of the character, shape of the character, sound of the character"), which was useful enough as far as it went, but which is not detailed enough for our purposes.

[11] Incidentally, and *a propos* of the movement towards a science of phonology: we refuse to consider the expression "human sciences" as an oxymoron or even a conceptual incongruity, nor do we consider the expression "natural sciences" tautologous or redundant. This is a substantial point of conceptual analysis, and the matter cannot be resolved by discussing English idiomatic usage of the word "science." Here as always it is important to distinguish carefully between doing the semantics of words and the analysis of concepts.

THE SIGNIFICANCE OF THE STYLE OF THE HEAD GRAPH

The head graph in Xu Shen's dictionary is not the standard and predominant *lìshū* 隸書 "clerical script" character as used on the bamboo strips of his time, but another, older, style, that of the *xiǎozhuàn* 小篆, the "small-seal script." This small-seal script, though widely used on seals as well as other precious artefacts, was certainly not the standard way of writing Chinese in most other contexts. Literally hundreds of thousands of bamboo strips bear witness to current everyday scribal practice from late Warring States times down to Xu Shen's own times. Nonetheless Xu Shen decided to disregard this current scribal practice. Instead, he entered as head graphs the seal-script graphs.

The clerical-script head graphs added for convenience in all modern *Shuowen* editions are not part of the received *Shuowen* text but represent modern insertions. Since Xu Shen could have added such *lishu* forms, why did he not find it worth his while to do so? It would conceivably have been convenient for the Han readership if he had added the clerical-script graphs, but in the context of his systematic analysis of characters, the decisive fact remains that this analysis is applied not to clerical-script forms but to small-seal forms of characters. Xu Shen's analysis does go on to identify in clerical script and not in small-seal script the structural constituents of each seal graph.[12] But this inconsistency remains harmless as long as there are no graphologically significant differences between the small-seal and the clerical-script versions of constituents identified. Xu Shen must have been aware of the many problems involved in using the small-seal graphs as head graphs and then referring to the elements in these graphs in their standard clerical-script shape.

The crucial point to notice is that Xu Shen's analysis of characters is graphemic and not graphic: he never explains the distribution of strokes in a graph or any prescribed or recommended manner of their execution. His concern is more abstract. Even when, on occasion, he does discuss the location of a constituent in relation to other constituents, he never gets anywhere close to a discussion of stroke order and the like. His concern is with what we should call the immediate constituent structure of each graph and the nature of the participants in this structure. These constituents are not graphs but graphemes. Now, the graphemic structure of characters is in Xu Shen's view best brought

---

[12] No wonder the Qing-era scholar Duan Yucai is sometimes tempted to rewrite Xu Shen's text so as to bring the identification of constituents closer to the small-seal script form. Duan Yucai 段玉裁 (1735–1815), *Shuowen jiezi zhu* 說文解字注 (Shanghai: Guji chubanshe, 1988).

out in that archaizing variety of traditional Chinese script of his time, namely, the small-seal script. To take one example among many possible ones, the character *diào* 弔 in clerical script does not contain the grapheme *rén* 人, whereas the small-seal version does. Here, as often, the small-seal script is structurally more revealing than the clerical-script standard characters, and therefore of greater use to Xu Shen's analytic approach. See also *shǐ* 史: "从又持中" ("*shǐ*: has 又 'hand' as a semantic constituent, which holds 中"), and *shu* 書: "从聿者声" ("*shu*: has 聿 'brush' as a semantic constituent and 者 as the phonetic constituent"), in which the constituent structure attributed by Xu Shen is manifest in the small-seal script and has disappeared in the clerical-script versions of the characters.

弔: 𢎛   史: 㞢   書: 𦘠

Epigraphy and palaeography are giving us ever new insight into the graphic variations of Chinese characters in ancient excavated texts. Xu Shen remains fairly indifferent to the great variety of ways in which a given graphic constituent can be manifested epigraphically in the texts of his own time with which he must have been highly familiar. His variant graphs always constitute graphemic variants and not mere structurally identical allographs. The fact that a character can be written very differently does not interest Xu Shen until the variation affects the graphemic constituent structure. What we mean by graphemic analysis is the focus on abstract constituent structure of graphs and the abstraction from the epigraphic variability of the graphic realization of the same graphemic elements. And what graphological analysis within the *Shuowen* system adds to graphemic analysis is the specification of the typology of the graphemic constituents. Thus, it is part of the graphological analysis of the graph 上 that this graph "refers (pictorially) to something" (*zhi shi*), but graphemically – under this interpretation – the graph is unanalyzable.[13] Xu Shen's intellectual ambition turns out to have been not merely to provide a graphemic analysis of the characters used in the classics of his time but to use small-seal script as a means to reach back into the history, or graphic etymology, of Chinese graphs.

One might be tempted to ask why, then, he did not go as far back as to the bronze inscriptions to which he does seem to have had some limited access. To this important question there is a straightforward

---

[13] The fact that one might insist that this character should be read as *hui yi* 會意 "associating ideas" because something X is depicted as being above a level Y is irrelevant to the present discussion.

answer: as any dictionary like *Hànyǔ dà zìdiǎn* 漢語大字典 shows,[14] even today, with vastly increased excavated bronze inscriptions available, a solid majority of the characters in *Shuowen* are found to have no *jīnwén* 金文 corresponding graphs. Xu Shen can thus only refer to such earlier varieties of writing incidentally, when they are available to him, and when they provide evidence relevant to his graphological analysis. The choice of the seal script enabled Xu Shen to go as far back as one can when one is looking for a reasonably complete character set for the large number of graphs he wished to discuss.

RADICALS

The first ordering principle of Xu Shen's dictionary is the organization of all 9,353 characters under 540 "radicals," or "classifiers."[15] The next principle is that each of these radicals is given an informally structured elaborate graphological description at the head of its 540 sections. All the characters (with the important exception of the radicals themselves) are subsumed under the 540 radicals.

It is clear that the total number of radicals was more important in Xu Shen's eyes than their functional use. Among the 540 there are 36 under which no character has been subsumed. It was suggested that 540 was the product of the multiplication of the symbolic numbers for the Yīn and the Yáng (6 x 9 = 54) and that in order to acquire a sufficient number of classifiers for his classifying purpose Xu Shen multiplied 54 by 10, and arrived at his number of 540.[16] Xu Shen probably chose the numeral 10 because it represented what he defines as the "complete number."[17] Ten is indeed glossed as "the completion of the series of numbers" and the graph is explained as including the four cardinal points and the center (i.e. the five directions).

Not all radicals can be used to write words: for example, the radicals *kǎn* 凵, *wéi* 囗, *mián* 宀, *zhuó* 丵, *qiān* 亍 , *ruò* 叒, etc., are recurrent elements in characters, but they are not complete characters in themselves. Since they do not represent words, they are *ipso facto* never pronounced as such, and one must ask how any pronunciation could

---

[14] *Hanyu da cidian* 漢語大詞典, Luo Zhufeng 羅竹風, ed. (Shanghai: Hanyu da cidian chubanshe, 1986–1993).

[15] Xu Shen points out that he has included 1,163 graphic variants, which together with the 9,353 entries, add up to the number 10,516. This is sometimes said to come close enough to *wan wu* 萬物 ("myriad [all-encompassing kinds of] creatures)."

[16] Atsuji Tetsuji 阿辻哲次, *Kanjigaku: Setsumon kaiji no sekai* 漢字学, 説文解字の世界 (Tokyo: Tōkai daigaku shuppankai, 1985), pp. 135-72.

[17] Bottéro, *Sémantisme et Classification*, pp. 69-71.

be assigned to them. We conclude that the readings for radicals must have been invented by the lexicographers.[18]

It must be pointed out that radicals are not like elements, unanalyzable into constituents. The internal graphological structure of radicals is frequently explained in *Shuowen*, as when the speech radical *yán* 言 is plausibly said to "*cóng kǒu* 从口" ("have the mouth radical as a semantic constituent"). Such complex radicals constitute a set of graphs that are not in fact retrievable under their declared main semantic constituents. We might say that they are "*cóng ér bù shǔ* 从而不屬," in the sense that they have a radical as their main semantic constituent, but do not in fact belong under that radical. They thus refute the current generalization that graphs in *Shuowen* can be found under their main semantic constituent.

Another anomaly is this: as many as 36 radicals are nonproductive, that is, they are radicals under which no other characters are subsumed. Part of the explanation of this anomaly is in the fact that Xu Shen needed to get to his magical number of 540 radicals. But if a character is impossible or difficult to analyze under Xu Shen's system, then declaring it to be a radical does find a place for it in that system. The numbers 三, 四, 五, 六, and 七 were all entered as "daughterless" nonproductive radicals: these characters were read as unanalyzable "primes" without any further function in the graphemic system.

It is commonly held that the radicals constitute an advanced lexical retrieval system for Chinese characters, and there is no doubt that this is indeed how they have come to function much of the time, and in practice. Meanwhile, their theoretic function in *Shuowen* has to do with the systematization of the primarily nonphonetic constituents in Chinese characters. The phonetic role of constituents is specified explicitly by the technical term *shēng* 聲, whereas a nonphonetic constituent X is *not* explicitly characterized as "semantic"; although we find reason to translate the technical term "*cóng* 从 X" as "has X as a semantic constituent": as we shall discuss forthwith, the nonphonetic constituents are generally construed by Xu Shen as semantic. Constituents introduced in this way can occasionally be assigned an additional secondary phonetic function by the formula "X *yì shēng*, X 亦聲" ("X is also phonetic").

---

[18] We note in passing that the *horror vacui* from which traditional lexicographers have tended to suffer may well have induced them to provide readings for many characters without reliably attested early readings.

SHUOWEN JIEZI AND HUMAN SCIENCES

It has to be said that a fair number of the radicals, and not only those nonproductive "daughterless" ones, provide little or no insight into any semantic features of the characters subsumed under them. While it is true that characters with the fish radical tend to have fish-related meanings, the radical *zhǔ* 丶, defined as marking a pause or breaking off in discourse, has no identifiable relation to what Xu Shen takes to be the graphologically relevant meaning "wick," of the character *zhǔ* 主 (later standardly written as *zhù* 炷) which he knew currently wrote the word for "ruler": it is significant that Xu Shen only considers the meaning of this word that is relevant to the explanation of the graph. Moreover, in this instance, Xu Shen declares the radical to be at the same time phonetic: 丶亦聲 ("丶is also phonetic").

In spite of examples like this, the radicals identify a set of recurrent predominantly nonphonetic constituents that are related to a semantic classification of words. They are needed for the graphological analysis of characters, and occasionally they are useful for the integration of characters into Xu Shen's cosmological schemes. These radicals are only incidentally convenient for the retrieval in dictionaries of the graphs containing them.

The formula "X *cong* 从 Y" generally indicates that a constituent in a given graph is typically semantic and not only phonetic. Whenever Xu Shen gives reasons or explanations for using the formula "X 从 Y," it turns out that he is concerned to give semantic explanations for the natural presence of the element Y in a graph X with the meaning he has assigned to that graph in his gloss. Thus, although Xu Shen has no similarly explicit semantic term corresponding to the phonetic specifier *sheng* 聲, there are good *internal* reasons for insisting that the formula "X 从 Y" does ascribe typically semantic elements.[19]

The entries on the radicals as head graphs in *Shuowen* are generally much more discursive than other entries. They display an encyclopedic interest that is only sporadic in the rest of the dictionary. Each section of *Shuowen* is conceived like a chapter in a book to which the entry on the radical itself serves as a kind of introduction. Within each of these chapters, the arrangement of material, though often erratic and generally unpredictable, is not always arbitrary. A large number of semantic

---

19 The isolated instances of X 从 Y 聲 are limited to explanations for allographs, i.e., graphologically distinct alternative ways of writing a given character. The pattern 或从X 聲 ("is alternatively written with the constituent X which is phonetic") in cases where X is manifestly irrelevant semantically, provides explicit evidence to prove that the technical term *cong* 从 could occasionally introduce in *Shuowen* constituents of a purely phonetic kind. See 10B 408 246 *bu* 怖: "scared out of one's mind," in which an allograph is explained by the phrase: 或从 布聲 ("[the graph *bu*] is alternatively written with the constituent *bu* 布 which is phonetic").

series structure the text, although there are some striking omissions in these series, namely, characters expected in a series that are found in arbitrary places elsewhere under the same radical. Under the heart radical alone we have identified the following series: JOY SERIES (14–16) (3 items), INTELLIGENCE SERIES (31–33) (3 items), AFFECTION SERIES (40-(42)-44) (4 items), THINKING SERIES: (59-(61)-63) (4 items), Minor FEAR SERIES 1[20] (69–70) (2 items), EFFORT SERIES (84-(+86)-87-(+88)) (3 to 5 items), Minor PEACEFUL SERIES (93–94) (2 items), ANXIOUS-NESS/EAGERNESS SERIES (106-(110)-112) (6 items), STUPIDITY SERIES (123–127) (5 items), LACK OF DILIGENCE SERIES (132–136) (5 items), NEGLIGENCE SERIES (139–142) (4 items), minor INDULGENCE SERIES (143–144) (2 items), RESENTMENT SERIES (168-(177)-180) (12 items), DISSATISFACTION/RESENTMENT SERIES (181–189) (9 items), DEJECTION SERIES (194–205) (12 items), MOVEMENT SERIES (207–209) (3 items), WORRY SERIES (211-(232)-233) (22 items), FEAR SERIES 2 (238-(244)-249) (11 items), SHAME/HUMILIATION SERIES (251–256) (6 items).

In spite of all these series, there is no attempt at an overall organization of the material under each radical, except that synonym-binome characters do tend to be entered into the dictionary together, and in the order of their occurrence in the binome. Thus *chóu* 惆 will be expected to precede *chàng* 恨 in *Shuowen* because of the currency of the synonym-binome *chóuchàng* 惆恨 "feel distressed." Moreover, we have noticed that positive terms tend to precede negative terms so that a radical section rarely begins with negatively charged terminology. On the other hand, there is a striking accumulation of negatively charged terminology towards the end of our present heart radical section. Strikingly, again, the mouth radical begins with the delightful words for a baby's crying and ends with spitting, affliction, silence and animal sounds. Note also the case of the woman radical *nǔ* 女 which begins with the highly valued terms for clan names and terminology for nubile young women, and which ends on words referring to wickedness.

Subsumption under a radical is not a matter of mechanical or straightforward semantic diagnosis. For example, quite a few characters that have psychological meanings and contain the heart radical are not to be found in the heart-radical section, in spite of the fact that Xu Shen must have known that this is where they would be looked for. As we noted above, Xu Shen did not see himself as using radicals as a retrieval system. They were an analytic conceptual tool.

---

20 "Minor" means less than three entries.

If subsumption had been merely a matter of semantic diagnosis, one might suspect that under the heart radical Xu Shen would try to assemble all characters with clear psychological meanings that contain the variants of the graph *xīn* 心 representing the heart. For example, the high-currency psychological term *yōu* 憂 ("worry"), would be looked for in vain under the heart radical, and so would the high-currency psychological term *ài* 愛 ("love"). In instances like these, Xu Shen clearly gives priority to considerations of graphological analysis, and he disregards considerations of ease of lexical retrievability (if retrievability in the lexicon indeed ever was part of his motivation at all). Xu Shen's classification is primarily of graphs and only secondarily of meanings. In cases like *ai* and *you*, it turns out that *xin* is embedded in the graph and is not an immediate constituent of that graph. Thus what modern linguists refer to as immediate constituent analysis turns out to be an indispensable methodological tool for Chinese character analysis, and Xu Shen employed this tool 2,000 years *avant la lettre*.

Xu Shen, vividly aware of the current meaning of the word *ai*, glosses the character as *xíng mào* 行皃 ("descriptive of a way of walking"), and he relates *suī* 夊 ("walk slowly") to what he sees as the graphologically operative meaning of the word. The remainder of the character, *ài* 㤅, is analysed as a compound phonetic constituent, and this compound constituent does happen to contain the heart radical. *Xin* 心 is neither a phonetic nor a semantic immediate constituent in this graph according to Xu Shen's analysis. Therefore there can be no question of its being a radical in 愛. Only immediate constituents can be radicals in the characters of which they are immediate constituents in Xu Shen's system, as translated into modern linguistic jargon. When a radical is inserted into another simplex constituent, Xu Shen goes so far as to conceive of one discontinuous graphic constituent to explain the construction of the graph. In any case, the graph *ài* 愛 cannot be said to 从心 ("have *xīn* as an immediate constituent") in Xu Shen's system. Xu Shen's abstract interest in graphological analysis goes so far that he completely disregards the well-known current meanings of Chinese key words when his analytic principles lead him to counterintuitive conclusions, as in the cases of *ai* 愛, *suo* 所 and *zhu* 主, above.

We have seen that by no means all characters that one would obviously look for under the heart radical are listed under that radical in *Shuowen*. Thus everyone would look for *sī* 思 ("think") under the heart radical until he has learned that Xu Shen treats *sī* as a complex radical which itself has the heart radical as its semantic constituent but is *not* entered under the heart radical.

It is of course also true that by no means all characters with clearly psychological meanings have the heart radical in the first place. Thus, our section on the heart radical is far indeed from exhausting the repertoire of characters with psychological meanings. Leibniz would have been disappointed to find that there is no regular relation between the presence of the feature PSYCHOLOGICAL and the presence of the heart radical in characters with basic meanings involving that feature. While few kinds of fish are written without the fish radical, many kinds of psychological concepts turn out to be written without the heart radical. Only one of the four main categories of emotions, namely, *āi* 哀 ("grief"), *lè* 樂 ("joy"), *xǐ* 喜 ("delight"), and *nù* 怒 ("anger"), have the heart radical. Moreover it is interesting to note that the word *yuè* 說 ("be pleased") came to be written with the heart radical long after it had become current as written with the speech radical. *Xiào* 孝 ("love for ones parents") is written without the heart radical and so is *hào* 好 ("have a predilection for"). In Xu Shen's time, the most current verb for "to desire" was *yù* 欲, and the current graph used to write that word did not have the heart radical, nor do any other common words for "to desire" or "to hope," like *yuàn* 願, or *wàng* 望. The fact that the Chinese writing system was very far from being a conceptually based writing system in Leibniz's sense comes out very clearly in *Shuowen*.

Only a minority of Xu Shen's glosses for graphs with a heart radical are themselves written with that heart radical. On the other hand, the heart radical is often added to characters with psychological meanings.[21] Páng Pǔ 龐樸, in an unpublished manuscript of 2004, has drawn our attention to the important practice, in excavated philosophical texts as well as in the Mohist *Dialectical Chapters*, of adding the heart radical as a marker indicating abstract psychological meaning. But this limited scribal practice never began to get near to establishing any regular practice of marking psychological terms with the heart radical. Even in the case of *yù* 慾 the addition of the heart radical has come to indicate a semantic nuance of "sexual lust" that is absent in the heartless *yù* 欲 ("desire").

THE SEMANTIC NATURE OF XU SHEN'S GLOSSES

In our translation of Xu Shen's plain gloss of the form X Y 也, we sometimes add an explanatory paraphrase introduced by EP [[in double square brackets]], and we often expand this to "X is (a kind of)

---

[21] Note incidentally that the addition of the heart radical to the character *yu* 欲 tends to create a new meaning of "illicit/inappropriate/excessive desire, lust" by Warring States times.

Y," "X is (a way of) Y-ing," and so on. This is because we find it impossible to believe that Xu Shen meant to say that X simply meant Y, in other words, that X and Y were synonymous. When Xu Shen glosses 息 "breathe" as 喘也 "is to pant," we assume that he was perfectly aware of the semantic difference between these two words. In such cases we have succumbed to the temptation to indicate the distinctive features that characterize X *vis-à-vis* Y. So, as in the example 慈愛也, we add a paraphrase [[EP: *cí* is (a way of) showing loving care (scil. towards one's next of kin of a younger generation).]] We are aware that these EP reconstructions do not represent interpretations explicitly indicated or even adumbrated by Xu Shen. However, we hope that these reconstructions may be found useful as provisional attempts to reconstruct the underlying semantic reasoning behind Xu Shen's often surprising glossing policy. And this much we do assume: like his commentator Duàn Yùcái, Xu Shen knew Chinese well enough not to have wanted to say what his traditional Chinese readers often took him to mean when he said "X Y 也," namely that X simply means Y, or that it is synonymous with Y.[22]

In any case, the hermeneutics of Xu Shen's decontextualized glosses poses immense questions throughout, precisely because these glosses have no disambiguating context. Translation must inevitably remain tentative in many cases. Translations of isolated glosses always have to be taken *cum grano salis*.

PHONETICS

As a premeditation on the subject of writing *vis-à-vis* pronunciation, it will be useful to take up in some detail the situation of English. There are plenty of words the writing of which preserves manifest traces of earlier pronunciations for modern words. Examples are many: there are reasons of historical pronunciation which give us both the spellings *knight* and *night*. (Compare German *Knecht* and *Nacht*.) When discussing the "spelling" of Chinese words through characters similar problems arise: the standardization of this "spelling" has to be discussed in terms of the pronunciation of the time of that standardization.

Cuī Shūhuá 崔樞華 suggests that the Old Chinese pronunciations should be used for the interpretation of the phonetic analyzes in *Shuo-*

---

22 One of the major original achievements of Duan's commentary is his intermittent but persistent attempt to explicate the different nuances of near synonyms in *Shuowen*, and in particular his distinction between *xi yan* 析言 "specific use" and *hun yan* 渾言 "generalized use" of words. See Feng Zheng 馮蒸, *Shuowen jiezi tongyici yanjiu* 說文同義詞研究 (Beijing: Shoudu Shifan daxue chubanshe, 1995).

wen.[23] Meanwhile, in our present study, we need to distinguish between three quite separate issues:

1. Do the phonological glosses we have refer to special conventions for the reading aloud of ancient texts, or do they report current readings of these characters in the colloquial language?

2. What was the phonological situation at the widely different times when the various characters in *Shuowen* were created?

3. What was the phonological situation at the particular time when Xu Shen proposed his analyses?

None of these issues has so far received sufficient focus.

In different contexts, we need both Old Chinese and Late Han reconstructions for the pronunciations of words. In order to reconstruct Xu Shen's thinking we must decide whether or not he was aware that Eastern Han pronunciation as he knew it was radically different from Warring States pronunciation and again from Late Shang and early Zhōu pronunciations. The commentator Zhèng Xuán 鄭玄 (127–200), in any case, was demonstrably aware of such historical sound changes, as is evident from his remark 古者聲栗裂同也 ("In ancient times, as for pronunciation, *lì* and *liè* were the same").[24] Xu Shen may have been aware of such sound changes. In spite of this, when interpreting Xu Shen's views, we are not entitled to make reference to details of Old Chinese reconstructions to which he would probably never have had access. Until we see convincing evidence that Xu Shen knew relevant details of Old Chinese pronunciation we must base our interpretation of his view on Eastern Han pronunciations. On this matter we are very happy to be able to use the newest drafts on a systematic reconstruction of Eastern Han pronunciations by Axel Schuessler.[25]

Having said this, however, we must be prepared for cases that on the basis of Eastern Han readings are implausible, but that in the light of what we now think we know about Old Chinese pronunciation are quite unproblematic or at least less problematic. When we comment on such matters, what we interpret is not Xu Shen's work, but the subject he is writing about: the graphological analysis of Chinese characters according to his system. And when we ask about how we need to in-

---

23 Cui Shuhua 崔樞華, *Shuowen jiezi shengxun yanjiu* 說文解字聲訓研究 (Beijing: Shifan daxue, 2000), p. 21.

24 Quoted in South Coblin, *A Handbook of Eastern Han Sound Glosses* (Hong Kong: The Chinese U.P., 1983) p. 11. Note that the issue is not whether these words rhymed, but whether they had the same pronunciation.

25 Axel Schuessler, *ABC Etymological Dictionary of Old Chinese* (Hawaii: Hawaii U.P., 2006).

terpret the composition of characters themselves (rather than how we need to understand their interpretations in the *Shuowen*), then we insist that we must base ourselves quite generally on the pronunciation of the characters at the widely different times at which these characters were produced. Thus a complex oracle-bone character with a phonetic element must be viewed in the light of Late Shang pronunciation, whereas a character first attested in Western Han inscriptions must be interpreted in terms of Western Han pronunciation, which may well have been different from Eastern Han pronunciation as Zhōu Zǔmó 周祖謨 suggests.[26]

THE VICIOUS CIRCLE OF RECONSTRUCTION ON THE BASIS OF PHONETIC SERIES

Phonetic reconstruction in Chinese can never be sure to be a complete reconstruction of the pronunciation of a word. It can only aspire to be a reconstruction of that part of the phonology of a word which our scanty surviving evidence can help us to make some more or less well-informed guesses on. Moreover, theoretical well-informedness carries its own dangers when it systematically builds into the reconstructions themselves just those elements which would later support the theories that motivated their introduction in the first place. We call this the vicious circle of reconstruction. The reconstructions themselves must never be based on the assumptions that they are then used to prove.

Moreover, apart from the problem of the vicious circle of phonological reconstruction, there is the serious problem of the phonetic specification of the realization of our necessarily phonological reconstructions. Such specifications must always remain essentially speculative as long as we have no sufficiently detailed access to phonetic descriptions of the language in question. From our point of view there are two essential points in this connection:

First, the nature of the relation between the pronunciations of phonetic elements and of the characters in which they are said to be phonetic constituents is an open empirical issue. All arguments based on assumptions concerning phonological relations between the members of phonetic series must therefore be treated as basically circular: it is the precise nature, reliability and validity of these assumptions them-

---

[26] "Liang Han yunbu lüeshuo" 兩漢韻部略說, in Zhou Zumo 周祖謨, *Wen xue ji* 問學集 (Beijing: Zhonghua shuju, 1966) 1, pp. 24–31.

selves that are at issue and must not be presupposed in the process of the reconstruction of pronunciations to be compared.

Second, since our primary linguistic interest is not in the late history of phonological rationalization by linguists with a theoretical interest in systematic phonology, but in early pretheoretical observation, we give pervasive priority to the early evidence such as that presented in *Jīngdiǎn shìwén* 經典釋文.[27] Later systematizing and rationalizing works such as *Guǎngyùn* 廣韻 (1008), interesting and important as they are as indigenous Chinese rationalizing systems, must not be misunderstood to represent observational empirical evidence, except in those cases where we have solid reason to believe that the authors of the *Guangyun* had access to earlier or more empirically reliable sources on the early pronunciation of words than those presented in *Jingdian shiwen*.[28] We must insist: such later Chinese systematizing hypotheses – like our own historical conjectures – can be interesting and stimulating, and naturally tempting indeed to the minds of similarly systematizing historical phonologists, but they must be carefully distinguished from what native speaker informants construe as their own factual observations or empirical evidence.[29]

For example, the currently accepted reading *sì* for the character 食 is derived in current dictionaries, including *Hànyǔ dà cídiǎn* as well as Wáng Lì's justly celebrated dictionary of classical Chinese,[30] from the notoriously unreliable handbook *Jíyùn* 集韻 (1039), whose readings are routinely rejected elsewhere whenever they differ uncomfortably from those in *Guangyun*. The proper study of the readings of 食 will begin when historians of phonology as well as lexicographers begin to take care to record systematically the earliest evidence for each reading they attribute to a character, and – more importantly – the early evidence for readings other than those they choose to accept as well as their arguments for rejecting them. Our plea is for a historical and critical study of Chinese phonology.

[27] See the planned "*Jingdian shiwen* Dictionary" by Pan Wuyun 潘悟云 and Christoph Harbsmeier.

[28] We need to know what were the superior early sources of which Lu Deming 陸德明 (550?–630) was unaware in his detailed bibliographic preface to *Jingdian shiwen* (Huang Zhuo 黃焯, *Jingdian shiwen huijiao* 經典釋文彙校 [Beijing: Zhonghua shuju, 2006]).

[29] The fact that such observations can be biased in various ways deserves close study but cannot concern us in detail here.

[30] Wang Li 王力, *Wang Li gu hanyu zidian* 王力古漢語字典 (Beijing: Zhonghua shuju, 2000).

For the purposes of our analysis of *Shuowen* we need to know, for each reconstruction of the pronunciation of a character, the following documented information:

1. When and exactly where is any given graph first documented, and when are we assuming the graph was created? Without this information we do not know what pronunciations for what historical periods are relevant to the character. Moreover, the occurrence of a character in a traditional text certainly does not count as evidence to prove that the corresponding graph existed at the time. Excavated texts are crucial to address the latter question. *Grammata Serica Recensa* nowhere pretends to be a detailed study of graphs of the type that is at issue here.[31] Bernhard Karlgren had no access to the immense epigraphic evidence which must be the basis for our dating of the invention of graphs.

2. What exactly is the earliest coherent evidence attesting to the association of a given phonetic reading of a graph with a given semantic interpretation of that graph, as used in a given context. Without this information we are unable to assess the reliability of the association under discussion. If this association is abundantly and unanimously documented in early sources this is very different from a situation where the earliest attested association dates from Sòng times or after. And it is sobering to note that even what we regard as earlier commentarial evidence on the pronunciations of words must count as dangerously late when the texts at issue belong to the fifth or fourth century BC.

OBSERVATION-BASED VERSUS
CONJECTURAL RECONSTRUCTIONS

Phonetic elements that are not widely attested as independent graphs for words of the language cannot serve properly as empirical evidence in any detailed description of the relations between the pronunciation of a phonetic and the word in which it is a phonetic. When a reading of such a lexicon-word fits well into the pattern of the *xiéshēng* 諧聲 system, this is more likely due to the professional ingenuity of the theoretician rather than any coherence in observed linguistic fact: since the "lexicon-word" is never pronounced one can hardly claim to observe how it is pronounced. (See for example, the phonetic constituent in *què* 愨 which is number 12 under the heart radical.) Even when early dictionaries provide readings for such phonetic constituents, we

---

[31] Bernhard Karlgren, *Grammata Serica Recensa* (Stockholm: Museum of Far Eastern Antiquities, 1957).

must always suspect them of being based on theoretical reconstruction rather than on observed fact, even when the theoretical reconstruction is utterly convincing and uncontroversial.

Thus we must be careful to distinguish between two kinds of readings: those conjectural readings created by lexicographers to satisfy their own need for coherence of the phonological system, on the one hand, and, on the other, those empirical readings that record observed current linguistic practice. The *Jingdian shiwen* collects a wide range of early reports on linguistic practice, and it therefore often contains material that systematizing lexicographers do not like to be reminded of. The *Guangyun*, however, is the modern lexicographer's delight because it often tends to insist on providing readings that connect old readings to modern pronunciations, and on readings that seemed to its authors to fit into an overall phonological system. This dictionary tacitly – one might even say surreptitiously – dismisses many reliable observational evidences that do not fit into its overall phonological scheme. Much of the wealth of phonetic information contained in *Jingdian shiwen* is tacitly passed over in *Guangyun*. When common pretheoretical observation on the pronunciation of a word is not available, then the pronunciation assigned to that word must be treated as merely conjectural, even when the conjectures may have great plausibility. Conjecture must never be confused with observation.

The question to what extent material like *fǎnqiè* 反切 spellings and the like homogeneously aspire to represent one and only one version of the language of the time has to be discussed as an empirical open-issue without a foregone conclusion. For all we know, it may turn out that it is wrong to look for the dialect that is represented by Middle Chinese, or by Old Chinese: we may in the end need to concur with Zhū Déxī 朱德熙 when he is said to have advised against very subtle detailed linguistic analysis of a linguistic object that itself is underspecified, like the concept of modern standard Chinese, which is underspecified insofar as it covers many distinct linguistic practices.[32]

THE *DU RUO* 讀若 ("TO BE READ LIKE") PROBLEM

There are three ways to indicate the pronunciation with the *dú ruò* 讀若 formula in *Shuowen*. This formula can be followed by:

---

[32] Zhu Dexi made this point on the occasion of his *Doctorat honoris causa*, Paris, 1986.

1. a simple homophonous word (Y): 翌讀若皇,
2. a current expression (AB, AB 之 B, etc.,): 奆：讀若書卷; 卸:讀若汝南人寫書之寫,
3. or a quotation: 詅讀若《論語》"跢予之足".

In addition there are thirty-six examples of the formula *"dú yǔ Y tóng"* 讀與 Y 同 and five cases of *"dú ruò Y tóng"* 讀若 Y 同. All these formulae are usually given at the end of an entry, after the graphological analysis (with some counter-examples). Together, these represent about ten percent of the *Shuowen* entries.[33] The *Shuowen* quotations from such authors as Fù Yì 傅毅 (Eastern Han), Jiǎ Kuí 賈逵 (賈侍中) (30 AD–101 AD), Sāng Qīn 桑欽 (Western Han), and Zhāng Lín 張林 (Eastern Han), possibly speaking different dialects or belonging to different hermeneutic traditions, show that Xu Shen's notes on pronunciation are often taken from older or contemporaneous works. However, the fact remains that Xu Shen seldom specifies his sources. As a minor significant point we note that under a radical one occasionally finds whole series of entries with a *du ruo* formula (*cf. xīn* 心: SW 10B 408: 079, 082, and 083: 恣). There is no regular system of phonetic annotation in *Shuowen.* This explains the variations and inconsistencies encountered by scholars who have tried to reconstruct the *Shuowen*'s language through its *du ruo* "glosses."[34]

Xu Shen mentions diverging opinions on pronunciations when he provides alternative *du ruo* glosses (讀若 Z, 又讀若 Z'; or simply 又若 Z'), or when he indicates that "according to another source the character should be pronounced such way" (讀若 Z, 一曰讀若 Z'). This is often the case for the radicals and for the elements that are not complete characters in themselves. Most puzzling are the examples in which the character following the *du ruo* formula is the same as the phonetic constituent of the glossed character: A 从 XY 聲讀若 Y. Xu Shen was surely aware that phonetic constituents did not necessarily correspond to the pronunciation of the characters they are part of. He often gave different *du ruo* glosses for characters and their phonetic constituents to record this. By contrast with this situation, he may have wanted to indicate identity of pronunciations in certain cases: 瑂石之似玉者. 从

33 See also Lu Zhiwei 陸志韋, "*Shuowen jiezi duruo* yinding" 說文解字讀若音訂, *Yanjing xuebao* 燕京學報 30 (1946), pp. 135–278, and Miller, "Problems in the Study of the *Shuo-wen chieh-tzu*," pp. 230 ff.

34 South Coblin, "The Initials of Xu Shen's Language as Reflected in the *Shuowen duruo* Glossses," *Journal of Chinese Linguistics* 6.1 (1978), pp. 27–75, and "The Finals of Xu Shen's Language as Reflected in the *Shuowen duruo* Glossses," ibid., 7.2 (1979), pp. 181–245.

玉眉聲. 讀若眉; 噱噱也. 从口集聲. 讀若集. (We have found twenty-three cases of this sort in *Shuowen*.)

The analysis according to the "从XY聲" formula is graphological and not phonological in nature. The distinction between semantic constituents motivated by the meaning of the character, and phonetic constituents which may or may not have such semantic mnemonic motivation, does not in any way make the *Shuowen* into a dictionary concerned with the pronunciations of words.

Some of the *du ruo* usefully disambiguate a phonetic constituent. The graph 厶, for example, corresponds either to *sī* 私 or to *gōng* 肱 ("upper arm"). Xu Shen explains *sī* 玌 as 石之似玉者. 从玉厶聲. 讀與私同: "*sī* is a jade-like precious stone. [The graph] has 'jade' as a semantic constituent; 厶 is the phonetic constituent. The pronunciation is the same as that of *sī*." In this case, his *du ruo* gloss solves the problem of how to read 厶. In addition, some well-known characters have received seemingly unnecessary *du ruo* notes (宋讀若送 or 圓讀若員), whereas many rare characters where one would like to have *du ruo* notes have none. It becomes increasingly clear that the assignment of *du ruo* readings is not systematic and often seems arbitrary. They are supplementary notes added from time to time, often for reasons we can no longer reconstruct. Since *du ruo* notes are not predictable, and since quite a few of them are idiosyncratic, and since moreover they are already in the Tang manuscripts, a plausible explanation for their retention in the text might seem to be their early origin. There is no need to declare them later additions to Xu Shen's text.[35]

A CAVEAT ON TRANSLATING THE *SHUOWEN*

Translation from classical Chinese sentences depends notoriously on the pragmatic context of what is being said: sentences in texts often need to be disambiguated on the basis of context. The Chinese understood themselves in the mode the French call *à demi-mot:* readers are assumed to be able to add to what is said plenty of things that are presupposed or understood. (Compare the Latin *subaudire* and the French loan translation *sous-entendre.*) The glosses in *Shuowen* are systematically decontextualized because they are concerned with relevant meanings of decontextualized words "as such." None the less, they can only be properly understood if one is prepared to reconstruct what is "under-

---

[35] See also Miller, "Problems in the Study of the *Shuo-wen chieh-tzu,*" pp. 270–71, on the historical background of the *Shuowen.*

stood" in them. Translation from *Shuowen* thus constantly forces one to impose meaning without that indispensable aid of sufficient context. The result is that translation becomes even more underdetermined by the data in *Shuowen* than it tends to be in classical Chinese generally.

There is, nevertheless, a great need for a carefully annotated translation of the *Shuowen.* We need a philosophical annotation that focuses analytically on the text as a comprehensive, systematic enquiry into a well-defined fundamentally important subject. The annotation will also need to focus on the explicit methodology applied in the text, on its highly technical stipulative terminology, and on its pervasive reliance not on traditional opinion, not on unanimous preconceptions, even less on scholarly authorities of the past, but on independent disciplined professional analysis of carefully chosen relevant data. In other words, this annotation will have to place the *Shuowen* firmly in the tradition of Chinese scientific literature. And we take "scientific literature" to be defined not merely by its subjects belonging to the realm of natural sciences and technology, but methodologically by the sophistication of systematic analytic procedures consistently applied. What matters is the logic of the scientific method, not primarily the subject matter addressed.[36]

We sincerely hope our annotated translations from the *Shuowen* that are in progress will live up to some of those high intellectual standards which Nathan Sivin has helped to set for the Chinese history of science. And when we consider the importance of the *Shuowen* for the history of science more generally, we think that the most important part is perhaps not only the tremendous amount of precious natural science information which it does contain. There is another part, of equal significance for the history of science: the sustained discipline and systematicity of a terminologically transparent analytic perspective, brought to bear on a well-defined important subject. Science is not only about information. It is also about logically coherent systematic analysis.

The *Shuowen* is not merely an important source for the historian of natural science, it is itself a remarkable monument of scientific inquiry, and the history of science must not be reduced to the history of natural science.

36 Hu Shi 胡適 (1891–1962) wisely called his book "The Development of the Logical Method in Ancient China" (Shanghai: Oriental Book Company, 1922). Such "logical method," such sustained systematic analysis, is indeed at the heart of the history of science, and it is much in evidence in *Shuowen.*