



CHAPTER 3

Phonology in the Chinese Script and Its Relationship to Early Chinese Literacy

David Prager Branner

On first consideration, the tantalizing subject of literacy in Early China might seem beyond our reach to discuss intelligently. For all that written Chinese is documented continuously back to the second millennium B.C.E., we see nothing concrete in the way of reflection on that tradition or on the systematization of it until the very end of the early period, in the first century.

No discussions of the practice of reading and writing or the relationship between the sound and the recorded form of language have survived into the received pre-Han canon. Received tools for classifying characters by shape are not seen until the *Shuowen jiezi* 說文解字 (hereafter, *Shuowen*) of the late first century (and the redaction of it dates from the tenth century).¹ Received tools for determining systematically how characters

This chapter is dedicated to the memory of Gilbert L. Mattos (1939–2002). Most of this chapter was originally read as “Crypto-phonograms in Chinese and the Ideography Debate,” at the 213th Annual Meeting of the American Oriental Society, 4 April 2003, Nashville, in memory of Mattos; the sections “Continuity of Literacy” and “Functional Homophony” were presented as “Loan-graphs and the Sound of Written Chinese,” at the 218th Annual Meeting of the American Oriental Society, Chicago, 15 March 2008; and the content of the section “Initial Types for *Xiéshēng* Groups” is adapted from an unpublished manuscript on the *xiéshēng* series. I am indebted to Mark Asselin, Jakob Dempsey, and Zev Handel for discussions about some of these matters during our student years in the 1990s and to eleven years’ worth of my own students at the University of Minnesota, National Chengchi University, and the University of Maryland, where I have taught using William Boltz’s *The Origin and Early Development*. I am grateful to a number of members of the Early China Seminar, who gave the chapter very close reading in 2009 — particularly Peter Daniels, William Boltz, Ken-ichi Takashima, and Matthias Richter. Thanks also to Keith Cunningham,

are read date from the sixth century, and their most unsystematic origins do not seem to date back more than a few hundred years. Expositions relating sound to form in the early script begin in the early nineteenth century. The oldest known Chinese glossary, the *Erya* 爾雅, says nothing about written form and seems to have been constructed without taking into consideration the obvious fact (obvious to us) that some characters represent more than one word with more than one pronunciation and meaning.²

People who wrote and read Chinese in early times must have thought critically about what they were doing from the beginning, but whatever they thought does not seem to have made its way into the received tradition until rather late.

Actually, complete silence on the subject would be strange if literacy really was common in Early China; we would expect to hear about it from some source or other and perhaps we will soon, through excavated texts. But for now, let us put our doubts aside and consider the evidence we can find in the explicit phonological content of the ancient script. Explicit phonological content is mainly found in the type of compound character known as the *xíngshēngzì* 形聲字, or “phonogram,” to use the translation adopted by Gilbert Mattos and Jerry Norman.³

Phonology is the study of relationships among the sounds of a language. Chinese historical phonology, in spite of appearances, is not mainly concerned with the reconstructed pronunciations of old words; it is concerned with the systematic relationships among and within Chinese speech sounds at points in the past and over time. In order to study the gross phonological freight of the script for the question of literacy, the motivating premise here is a major modern hypothesis about the historical structure of the Chinese writing system. The hypothesis is controversial and serves as an aid to discussing the phonological content of Chinese writing and considering what that implies about literacy in Early China. Note that this chapter restricts itself to the more passive aspect of literacy,

who in 2004 helped me collate a bibliography on experimental studies of reading and related issues. And thanks to one of the anonymous prepublication reviewers of this book, who made extensive suggestions and caught many errors.

1. The *Shuowen* is an inventory of characters, organized by structure and containing sporadic discussion of the relationship between structure, sound, and meaning. There is some question as to whether it should be described as a “dictionary.”
2. David Prager Branner, “On Early Chinese Morphology and Its Intellectual History,” *Journal of the Royal Asiatic Society*, ser. 3, 15.1 (2003): 60.
3. Qiu Xigui, *Chinese Writing*.

the process of reading.

The argument in this chapter is divided into six parts, each discussed and then summarized in its own section.

The Crypto-phonogram Theory of Chinese Character Structure

The modern hypothesis under consideration in this chapter is what may be called the “crypto-phonogram theory” of Chinese character structure. No other name exists for it now. In the words of William G. Boltz, it states that “there is no way a character can be ‘invented’ by putting together constituent elements none of which is intended to have any phonetic function” and that “many . . . characters constituted of two or more elements allegedly based only on the meaning of the elements, not the sound, are after careful analysis explicable as phonetic compounds.”⁴ Boltz’s phrasing here is unequivocal. The originator of the theory, Peter A. Boodberg (1903–1972), in his last published statement on the subject, described purely semantic compounds as “a relatively small but exceedingly important class of graphs.”⁵ Boltz’s stronger formulation, however, is more useful for discussion.

As an example, consider table 3.1, showing the graph *xìn* 信, composed of two elements, one of which serves as the “phonetic component” (or “phonetic” or “phonophore”) — the element that indicates the sound of the word the character is intended to represent — and the other as the “semantic component” or “signific,” a determinative token that disambiguates graphs for different words represented by the same phonetic. Here, *rén* 人 is the phonetic element in *xìn* 信 and, according to the theory, 言 serves to disambiguate 人 from other uses, such as those disambiguated as 仁 ‘kernel’, 千 ‘thousand’ (historically 人 with a mark, i.e., 𠂇). A “reading” of {N-in (真)} for the graph 人 is adequate to show that it serves as the phonetic element in 信 {N-in (真)}.⁶

4. See William G. Boltz, *The Origin and Early Development*, 72.

5. Peter A. Boodberg, “The Chinese Script: An Essay on Nomenclature (the First Hecaton),” *Zhongyang yanjiuyuan lishi yuyan yanjiusuo jikan* 中央研究院歷史語言研究所集刊 29.1 (1957): 118.

6. A note on the early Chinese readings used here: Early Chinese phonology is known to us only in outline; although some highly detailed reconstructions have been published, most of that detail cannot be fully confirmed or in many cases even tested. Regrettable though it is, that is the nature of the subject. But two elements are essential for

Table 3.1 *Xìn* 信 as an example of a phonogram

	Component elements		Compound graph
Modern graph	人	言	信
Meaning	person	to say; words	trust
Readings	{N-in (真)}	{NG-an (元)}	{N-in (真)}
	Phonetic element	Semantic component	

Although this example seems simple enough, *xìn* 信 may be considered a “crypto-phonogram” because in the mainstream Chinese tradition, the phonetic function of *rén* 人 was not understood until very late. Instead, *xìn* was described in the *Shuowen* as having *huiyi* 會意 ‘combined meanings’ structure — in other words, as being a purely semantic compound.

As it happens, *xìn* 信 seems to be a relatively late graph; it is not attested in oracle bones or bronze inscriptions, although there are plenty of Qin and Han examples of it. In the received corpus, it appears in the *Book of Poetry* (*Shijing* 詩經), the *Zuozhuan* 左傳, and the *Erya*, and it is the given name of two important military figures who died at the beginning of the Han; it is often used interchangeably in early records with 申 and 伸 {L-in (真)} and, to a lesser degree, with 新 and 親 {S-in (真)} and with 身 {L-in (真)}. So it is an important word. It is curious and worth noting that although modern scholars see the graph to be a phonogram, *Shuowen*, compiled at the end of the Han, considers it a semantic compound. (And it

explaining the behavior of *xiéshēng* 諧聲 contact in phonetic compound characters: the place (but not usually the manner) of articulation of the syllable-initial, and the syllable-final or “rime category,” which is typically named with an exemplary Chinese character. In order to represent this information as a “reading,” I am using a set of ten capital letters representing the main initial types together with William Baxter’s 1992 reconstruction of the Old Chinese rime categories. “Baxter’s 1992 reconstruction” refers throughout this chapter to the system published as William H. Baxter, *A Handbook of Old Chinese Phonology* (Berlin and New York: Mouton de Gruyter, 1992). Although Baxter, working with Laurent Sagart and other collaborators, has greatly revised his reconstruction, at present the 1992 version is the most complete in print, and that is the version cited here. See the section “Initial Types for *Xiéshēng* Groups” at the end of the chapter for a list of the initial types and what they correspond to in terms of fuller phonological systems.

may be significant that the reading {N-in (真)} is at variance with the readings {S-in (真)} and {L-in (真)}.)

Boodberg's theory was not merely that the compilation had missed some of the phonetic elements in its explanations but that many early Chinese graphs were polyphonous" — having more than one reading — and that sometimes an otherwise unknown reading has served as the phonetic element in a compound graph. It is because that phonetic function is claimed to have been unknown that the special term "crypto-phonogram" is needed; as explained below, it seems important that some phonetic elements have been well known through history and others unknown. The term "crypto-phonetic" is used here to mean the historically unrecognized phonetic element of a crypto-phonogram.

A well-known example of a polyphonous graph is that of *shān* 𦍋 'hair', which has a usual reading of {L-om (侵)}. Boodberg proposed that it once had a second reading with a labial initial, {P-iw (幽)}, as evidenced by the pair of crypto-phonograms shown in table 3.2.⁷ But note that *biāo* 彪 has a second reading in medieval sources, equivalent to *shān* 衫 and *shān* 衫; Boodberg's first example of 𦍋 being read {P-iw (幽)} is thus not strictly necessary.

Table 3.2 *Biāo* 彪 and *biāo* 彪 as crypto-phonograms

	Component elements		Compound graph
Modern graph	𦍋	長	彪
Meaning	hair	long	long hair
Readings	*{P-iw (幽)}	{T-ang (陽)}	{P-iw (幽)}
	Crypto-phonetic	Semantic component	

	Component elements		Compound graph
Modern graph	𦍋	虎	彪
Meaning	hair	tiger	tiger's stripes
Readings	*{P-iw (幽)}	{K-a (魚)}	{P-iw (幽)}
	Crypto-phonetic	Semantic component	

7. See Boodberg, "Some Proleptical Remarks," 329–72.

An even more celebrated case is that of the graph *xì-xī* 夕 ‘night’ (shown in table 3.3), which has an early Chinese reading of {L-Ak (鐸)} but which Boodberg proposed once had a second reading identical to 冥 ‘night’ {M-ing (耕)}.⁸ Boodberg argued that *kǒu* 口 ‘mouth’, too, had a reading presumably of {M-ing (耕)} (see table 3.4).⁹

Table 3.3 *Míng* 名 as a crypto-phonogram

	Component elements		Compound graph
Modern graph	夕 standing for what is now written 冥	口	名
Meaning	night	mouth	name
Readings	*{M-ing (耕)}	{K-o (侯)}	{M-ing (耕)}
	Crypto-phonetic	Semantic component	

Table 3.4 *Míng* 鳴 and *mìng* 命 as crypto-phonograms

	Component elements		Compound graph
Modern graph	口	鳥	鳴
Meaning	mouth	bird	to sing, sound
Readings	*{M-ing (耕)}	{T-iw (幽)}	{M-ing (耕)}
	Crypto-phonetic	Semantic component	

	Component elements		Compound graph
Modern graph	口	令	命
Meaning	mouth	command	command
Readings	*{M-ing (耕)}	{L-ing (耕)}	{M-ing (耕)}
	Crypto-phonetic	Semantic component	

Boodberg’s original statement of this theory was elaborated in two rejoinders to articles by Herrlee Creel;¹⁰ it was eventually codified with

8. Boodberg, “Some Proleptical Remarks,” 342; cf. Boltz, *The Origin and Early Development*, 63, 103–5.

9. Boodberg, “Some Proleptical Remarks,” 342.

10. Boodberg, “Some Proleptical Remarks”; Peter A. Boodberg, “‘Ideography’ or

far more substantial support and more careful argument by William Boltz.¹¹ At the outset, Boodberg seems to have been elaborating an original hypothesis, although he sounds as though he is stating established fact.¹² Of Boltz's many contributions to the hypothesis, his most important has been to introduce the theory of word families as a way of justifying otherwise irretrievable readings.

Reactions to the crypto-phonogram theory among sinologists and Asianists are mixed but rarely neutral. One active school of thought, represented by John DeFrancis (1911–2009) and J. Marshall Unger, opposes the notion of purely semantic compounds without explicit phonological content, known by the popular terms “ideogram” and “ideograph.”¹³ In his influential study of writing systems, Ignace Gelb mentions Boodberg's opposition to Creel's use of the term “ideography,” adding (without judgment) that it had remained in common use in Egyptology and cuneiform studies.¹⁴ Today, the situation in those fields remains as it was in Gelb's day, but resistance to these terms seems to be quite pronounced among some sinologists.¹⁵

Iconolatry?” *T'oung pao* 35 (1940): 266–88; Herrlee G. Creel, “On the Nature of Chinese Ideography,” *T'oung pao* 32 (1936): 85–161; Herrlee G. Creel, “On the Ideographic Element in Ancient Chinese,” *T'oung pao* 34 (1938): 265–94.

11. Boltz, *The Origin and Early Development*.

12. David Lurie has described Boodberg's papers as exhibiting a more developed sense of sinological “disciplinarity” than Creel's; David B. Lurie “Language, Writing, and Disciplinarity in the Critique of the ‘Ideographic Myth’: Some Proleptical Remarks,” *Language & Communication* 26 (2006): 250–69, 254–55. That may be a gentle way of criticizing Boodberg's austere academic tone, since Creel seems to have been rather more familiar than Boodberg with the early script as attested in bronze inscriptions; see Herrlee G. Creel, “Bronze Inscriptions of the Western Chou Dynasty as Historical Documents,” *Journal of the American Oriental Society* 56.3 (1936): 335–49. It does not appear that Boodberg took much time to look at excavated materials.

13. John DeFrancis, *The Chinese Language: Fact and Fantasy* (Honolulu: University of Hawai'i Press, 1984); J. Marshall Unger and John DeFrancis, “Logographic and Semasiographic Writing Systems: A Critique of Sampson's Classification,” in *Scripts and Literacy*, ed. I. Taylor and D. R. Olson (Dordrecht, Netherlands: Kluwer Academic Publishers, 1995), 45–58; J. Marshall Unger, *Ideogram: Chinese Characters and the Myth of Disembodied Meaning* (Honolulu: University of Hawai'i Press, 2004).

14. Ignace J. Gelb, *A Study of Writing: The Foundations of Grammatology* (Chicago: University of Chicago Press, 1952), 107; Ignace J. Gelb, *A Study of Writing* (Chicago: University of Chicago Press, 1963), 107.

15. In 2002 and 2003, I learned from conversations with the Egyptologist James P. Allen and cuneiform authorities Gene Gragg and Craig Melchert that it is generally accepted in their fields that some graphs are composites of semantic elements only. I wish to

The positing of such a “concept script” is a Neo-Platonist idea with a long history in Europe; in antiquity, it was the mystical lens through which Egyptian hieroglyphics were viewed, and in the Renaissance, it found its way into discussions of Chinese. On the one hand, some modern opponents of ideography indulge in polemics, as perhaps befits a response to mysticism intruding into objective philology. In that respect, they recall Peter Duponceau (1760–1844), whose assertion that Chinese writing represented spoken language rather than pure ideas is considered correct and inspired, even though he lacked the evidence to make a clear, accurate, and self-contained case.¹⁶







On the other hand, most native speakers as well as many sinologists and Asianists reject the categorical way in which the universality of the phonetic principle is sometimes asserted by its supporters. Articulate expression of this rejection has been made by Françoise Bottéro, who has published the most spirited of the responses to Boltz. One of her main points is that Chinese has many compound graphs, attested from an early period, in which neither element appears to be phonetic.¹⁷ As an example, she cites 武, analyzed in an early native text as 止 ‘foot; to stop’, combined with 戈 ‘halberd’, shown in table 3.5.¹⁸ Neither 止 nor 戈 has received readings with any phonetic likeness to that of the compound graph 武.¹⁹ We could make quite a list of such characters, clearly

acknowledge their help and also that of the scholar of Persian and Sogdian David A. Utz, grammatologist Peter T. Daniels, and the scholar of Canaanite Eva von Dassow.

16. Y. R. Chao commented, “With such poor language equipment, it is all the more remarkable that Du Ponceau had such a sound and penetrating view of Chinese writing”; see Yuen Ren Chao, “A Note on an Early Logographic Theory of Chinese Writing,” *Harvard Journal of Asiatic Studies* 5.2 (1940): 189–91, 190.
17. Françoise Bottéro, “Review of Boltz, *The Origin and Early Development of the Chinese Writing System*,” *Journal of the American Oriental Society* 116.3 (1996): 574–77; Françoise Bottéro, “Writing on Shell and Bone in Shang China,” in *The First Writing: Script Invention as History and Process*, ed. Stephen D. Houston (Cambridge: Cambridge University Press, 2004), 252–54.
18. Bottéro, “Review of Boltz,” 576.
19. After completing this chapter, I learned of Boltz’s 2006 paper, which proposes that 止 had a second reading {M-a (魚)}, allowing it to serve as the phonetic in wǔ 武; Boltz notes that 武 is attested in the meaning “footstep” and that the reading {M-a (魚)} is more commonly associated with the graph bù 步 {P-a (魚)} ‘footstep, pace’. William G. Boltz, “Phonographic Motivation in the Formation of Compound Chinese Characters: The Case of Wǔ 武,” in *Écriture chinoise: Données, usages et représentations*, ed. Françoise Bottéro and Redouane Djamouri (Paris: École des Hautes Études en Sciences Sociales, Centre de Recherches Linguistiques sur l’Asie

containing two elements yet with no obvious phonophore.²⁰ But even if adduced in great number, they cannot disprove the crypto-phonogram hypothesis, for a simple reason: although we need ancient readings in order to illustrate the phonetic structure of every compound graph, the hypothesis does not claim that all such readings survive or are even deducible.

Table 3.5 Is wǔ 武 a phonogram?

	Component elements		Compound graph
	OBI  HJ: 20197 Bronze  JC: 4292	OBI  HJ: 3335 Bronze  JC: 11117	OBI  HJ: 36080 Bronze  JC: 2523
Graph	止	戈	武
Meaning	foot; to stop	halberd	war; martial
Readings	{T-i (之)}	{K-aj (歌)}	{M-a (魚)}

Note: Oracle bone inscription (OBI) graphs are cited by HJ numbers, from *Jiaguwen heji*; bronze graphs are cited by JC numbers, from *Yin Zhou jinwen jicheng*.)

As a theory, the crypto-phonogram hypothesis is weak because it is not strictly falsifiable at all. Indeed, neither Boodberg nor Creel was offering falsifiable explanations of graphic structure; it is difficult to frame either one in predictive terms that could be disproved conclusively. Creel's method is completely passive: he simply combines the meanings of the component elements in whatever way is necessary to produce the known meaning of the compound. It is a beautiful example of tautology. By comparison, the crypto-phonogram theory is stronger in two respects. First, it can at least be rephrased as a sort of long-term pseudo-prediction: in what appear to be purely semantic compounds, at least one element will eventually prove to have a reading that makes it plausible as a phonetic element. In the shorter term, Boodberg's method could also perhaps be

Oriente, 2006), 55–73.

20. See Bottéro, "Writing on Shell and Bone," 252–54, for more recent comments. The example she gives of *qǔ* 龠/𪗇 'tooth decay' as a composite of 齒/牙 in combination with "worm" is treated in table 3.10, below.

subjected to statistical analysis, although no one has attempted it to date; statistical analysis could be used to justify or attack the significance of the number of attested crypto-phonograms. Those two facts make crypto-phonogram hypothesis at least potentially superior to Creel's method.

To summarize, neither the crypto-phonogram hypothesis nor the “ideographic” principle of pure semantic compounding is falsifiable, although in formal terms, the former is somewhat stronger. It goes without saying that this whole question is modern. Today we follow Duan Yucai's 段玉裁 (1735–1815) dictum that “[characters] sharing a phonophore are always in the same rhyming group” (*tōngshēng bì tóngbù* 同聲必同部).²¹ But phonology never seems to have been understood as the organizing principle of the script until Duan's time. It is one thing for us to posit models to explain the past and confirm their plausibility; but of course that does not mean that the models themselves were also known in antiquity.²²

Complex Pictographs and Analytical Composite Graphs





If the crypto-phonogram and ideographic approaches appear irreconcilable, they are not. Both are concerned with explaining the origins of compound characters as composites of two or more known simpler characters — as if the reader were consciously assembling whole words out of their component parts, just as the writer does. That, however, is a clear anachronism. Neither is a true analytical composite of the kind that developed later. Chinese characters are read, not deciphered.²³

21. “The original sounds are transformed, so that even when [graphs] share a phonophore, they are scattered through different rimes and rhyming groups. [. . . The patterns] are irregular and the tradition of scholarship has often doubted them. But if we seek their origins, [graphs] with the same phonophore must have been in the same rhyming group” (自音有變轉，同一聲而分散於各部各韻 . . . 參差齊，承學多疑之，要其始，則同諧聲者必同部也); Duan Yucai 段玉裁, “Gu shiqibu xiesheng biao 古十七部諧聲表,” *Liushu yinyun biao* 六書音韻表, 2.1A, in *Shuowen jiezi zhu* 說文解字注 (Jingyun Lou cangban 經均樓藏版, repr. Shanghai: Shanghai Guji Chubanshe, 1986 [1981]), 818.

22. A parallel case is that of derivational morphology in early Chinese. The history of our slow awakening to the plausibility of early morphology is narrated in Branner, “On Early Chinese Morphology”.

23. Since presenting this paper in 2003, I have become aware of the important work of Galambos (2006), discussing the history of structural variation in the early Chinese

Many compound graphs seem to have originated as what may be called “complex pictographs”: they contain two or more recognizable elements interacting in some way. Through the process of normalization, at some point those elements came to be written as discrete characters, and in time the whole characters were analyzed as compounds. Complex pictographs in their original forms are considered here as indivisible units and not necessarily as the juxtaposition of independent characters into which the whole is supposed to be divided analytically.


Take as an example the graph depicting an arrow in a bow. Oracle bone inscription (OBI) forms such as  (HJ: 29355) and  (HJ: 28392) are common and quite typical; a minority of similar forms (e.g.,  JC: 2559) and a majority with a hand added next to the line depicting the bowstring  (JC: 2784) are seen in bronze. This graph is well established as the antecedent of modern *shè* 射 ‘to shoot’ {L-ak (鐸)}. The original is not simply one pictograph juxtaposed to another — *shǐ* 矢 ‘arrow’ {L-ij (脂)} plus *gōng* 弓 {KW-ing (蒸)} — but actually a single unit in which the relative positions of the elements are part of the depiction. It is a whole depiction. The original OBI form could perhaps have been normalized as 𠄎 but is instead 射, which is structurally opaque, seeming to consist of *shēn* 身 ‘body’ and *cùn* 寸 ‘inch’ (in addition, there is a received variant 𠄎). *Shè* 射 in its original form was a complex pictograph, but it has been normalized as a false compound graph; the original pictograph is now represented as what appears to be 身, historically a different graph and word.²⁴

script. His discussion of “the myth of an ideal structure” in the period leading up to *Shuowen* is especially interesting. See Imre Galambos, *Orthography of Early Chinese Writing: Evidence from Newly Excavated Manuscripts* (Budapest: Department of East Asian Studies, Eötvös Loránd University, 2006), esp. 66–77.

24. The received tradition retains another graph *shǐ* 𠄎 and its variant 矧, which *Shuowen* describes as “To get the point of what is being said, [as quickly as] the firing of an arrow” (取詞之所之如矢也) and glossed elsewhere in received scholia as “the gums”; its most common use is as a loangraph for a grammar particle meaning “no less; how much more so.” See Ding Fubao, *Shuowen jiezi gulin* 說文解字詁林 (Shanghai: Yixue Shuju, 1930), no. 3297. The sense of “drawing back” implied by the components 弓 and 引 is better suggested by the usage “to bare the gums”: “[During mourning], laughter/smiling does not reach the point of baring the gums and anger does not reach the point of revilement” (笑不至矧, 怒不至詈); see *Shisanjing zhushu* (Qu li 曲禮 2 of *Liji* 禮記, 2:1243–44. Assuming 引 is phonetic in 矧, this graph would be read

The process of normalization was not at first an act of standardization, because it neither was planned nor involved consistent application of principles. It continued over a very long time; the medieval rime books of the fifth to seventh centuries must be considered an important part of the process at a late stage, since they promulgated not only standard *kǎishū* 楷書 (square script) forms but also standard readings and equivalences between characters and, beyond that, a model of phonology vastly more detailed than the explicit phonology of the script.²⁵ It is evident that as time passed, the principle of phonetic compounding was accepted and phonograms made up the great majority of standard characters. Many phonograms were produced by adding semantic determinatives to simpler graphs, but the marketplace of normalization sometimes also allowed competing phonograms and unrelated pictographs (simple or complex) for the same word to coexist in the rime books (table 3.6).

In the received tradition, there are few clear examples of characters normalized as composites of two whole written words until rather late. The earliest examples are scribal ligatures. These are seen in Warring States–Han bamboo manuscripts, usually with what is traditionally called a “doubling mark” (*chóngwén hào* 重文號) under the two characters, to show that they are being written in the space of one.²⁶ In Western Zhou bronze inscriptions, this mark usually seems to mean that the character is to be read twice, but by the time of the bamboo manuscripts, what is apparently the same sign is being used to mark ligatures (granted, not strictly “doubling” in the original sense). A ligature in Chinese usually consists of two characters fused into the space of one and sharing some graphic element; this hybrid character has the mark under it, indicating that two words are meant. Usually a ligature replaces a recognized compound, such as *jūnzǐ* 君子 ‘person of breeding’; the bottom part of 君 is made to

{L-in (真)}, which is equivalent to 身. So perhaps the normalization of  as 身 is not entirely arbitrary, although even with the reading {L-in (真)}, 身 is not a plausible crypto-phonetic in 射 {L-ak (鐸)}.

25. “Rime book” is the translation of the Chinese term *yùncū* 韻書: these are not books of rhyme-words for poetry but philological dictionaries organized by “rime” (syllable-final category). Hence the special spelling “rime,” distinct from “rhyme.”

26. Matthias Richter has given considerable thought to distinguishing the functions of the different marks found in the bamboo manuscripts; see his “Textual Identity and the Role of Literacy in the Transmission of Early Chinese Literature,” chapter 6 in this volume.

Table 3.6 Redundant pictographs and phonograms in the medieval phonological sources

Meaning	Pictograph (simple or compound)	Phonogram	Reading
ceremonial covering for the knees	市	韍	P-ot (月)
to discern, make out	采	辨	P-an (元)
lovely	美	嫩	M-ij (微)
quiver (for arrows)	葡	箛	P-ik (職)
large tripod for cooking	鬲	鑿	L-eik (錫)
insects	蚰	蛄	K-un (文)
flame	焱	焰	L-am (談)? L-om (侵)?
to look to either side	眴	眴	KW?- L?- + -a (魚)? KW?- L?- + -o (侯)?
to demote, devalue	𠄎	貶	P-am (談)? P-em (侵)?

double as the top part of 子, as illustrated in the Guodian forms shown in figure 3.1.



Figure 3.1 A normal ligature from the Guodian manuscripts. (From *Jianbo shufa xuan bianji zu*, *Guodian Chu mu zhujian: Tang Yu zhi dao*, 16, and *ibid.*, *Xing zi ming chu*, 20, 56).

With regard to the difference between compound and composite characters, ligatures are significant in that they are clearly made of two word-characters, not just two elements. Perhaps some of them represented actual contractions in speech or reading. For instance, a number of these

ligatures in the bamboo manuscripts stand for multiples of ten, three of which survived as graphs into the rime book tradition and as oral contractions into modern times, as shown in table 3.7.²⁷

Table 3.7 Numeric ligatures representing oral contractions, a type of composite graph

Ligature in the rime books	Contraction of	Baxter's 1992 reconstructions of uncontracted numbers	Presumed reading of ligature in Baxter's 1992 reconstruction	Modern Mandarin character readings	Modern Taiwanese
廿	二十 'twenty'	*njjs + *gjip	*njip	niàn	jī cháp > jiáp
卅	三十 'thirty'	*sum + *gjip	*sup	sà	sa ⁿ cháp > siap
卌	四十 'forty'	*s(p)jij/ts + *gjip	*sjip	xì	*sì cháp > *siap

27. In Taiwanese, the expected reading for “forty” would be homophonous with “thirty,” which perhaps explains why it is not attested. Note that the Taiwanese pronunciations are true contractions of the current colloquial names of the numerals, not reflexes of the early Chinese or medieval readings. Taiwanese forms are from William Campbell, *A Dictionary of the Amoy Vernacular* (Tainan, Taiwan: Ho Tai Hong Print, 1913), 276, 618. The Mandarin dictionary readings *sà* 卅 and *xì* 卌 match the medieval readings and are never used in speech. The Mandarin reading *niàn* 廿 is both a dictionary reading and phonologically irregular, and there is an interesting problem connected with it. *Niàn* is homophonous with *niàn* 念, and the same homophony is attested in a significant number of Wu dialects, in which 廿 represents a common colloquial word for “twenty” (personal communication, Richard VanNess Simmons). That reading, though incompatible with the *Guangyun* 廣韻, is attested as far back as the eleventh century: Gu Yanwu 顧炎武 (1613–1682) writes, “On the back of the stele it says, ‘. . . written on the 念-fifth day of the tenth month of 1091’; the practice of writing 廿 as 念 is first attested here” (碑陰 . . . 有曰, . . . 元祐辛未陽月念五日題, 以廿為念, 始見於此); see Gu Yanwu 顧炎武, *Jinshi wenzi ji* 金石文字記, in *Zhihai congshu* 指海叢書 (*Congshu jicheng chubian* 叢書集成初編 edition, vols. 1517–18) (Shanghai: Shangwu Yinshuguan, 1935–40; repr., Beijing: Zhonghua Shuju, 1991), 3.173–14. It is unclear why the standard Chinese and Wu pronunciation of *niàn* 廿 has a nasal coda, with *-m changing to -n in standard Chinese and to vowel-nasalization in conservative Wu; Shanghai has [ɲiɛ₁₃]; see Xu Baohua 許寶華 and Tao Huan 陶寰, *Shanghai fangyan cidian* 上海方言詞典 (Nanjing: Jiangsu Jiaoyu Chubanshe, 1997), 14, 96. Is the *-m coda perhaps an early attempt to preserve the original *-p ending? Or was there another word for “ten” in Chinese, now lost to memory?

Contractions represented as ligatures are a very clear case of composite graphs that are intended to be read analytically. We may assume that *niàn* 廿 and *sà* 卅, at least, were being read as single syllables in the second century B.C.E., because in the metrically regular stele inscriptions of the first Qin emperor, they occupy the space of a single character and syllable; two of these inscriptions survived in physical form late enough to be attested in partial rubbings, so we can be reasonably confident that their received texts in the *Records of the Grand Scribe* (Shiji 史記) are not middle period fabrications.²⁸

And the bamboo manuscripts contain another sort of ligature that is even clearer — this kind seems to be a graphic pun. A recognized character appears with the “doubling” mark under it. There is no reason to think that these graphs represent oral contractions or rhyming binoms or the traditional repetition of a whole word. Contextually, we grasp that two characters are meant: the character we see, and also a second character whose form is entirely contained within the strokes of the recognized character. A few examples are shown in table 3.8.

It is as though today we were to write “竝々” to represent “竝立,” “夫々” to represent “大夫,” and so on.²⁹ What is significant about these ligatures is the sense of play that seems to have been responsible for their creation. Play is surely a sign of considerable comfort with the rules of writing; it may be said to be a sign of the maturity of a writing system.

In the received tradition, there is no such clear evidence of play with the structure of characters until much later. Though there are a few dubious early cases of “portmanteau” graphs — whole characters combined playfully as whole words (e.g., *liè* 劣 and *fěi* 腓/ *pò* 咄 and a few others) — this type of structure is not discussed in *Shuowen*, and records of definite cases do not appear until well into the medieval period. Table 3.9 shows some representative examples. Portmanteaux constitute a graphic game of a kind that is still enjoyed today in the Chinese-literate world.³⁰

28. For texts, see the critical translations in Martin Kern, *The Stele Inscriptions of Ch'in Shih-huang: Text and Ritual in Early Chinese Imperial Representation* (New Haven: American Oriental Society, 2000). *Niàn* 廿 appears on page 12, text 1 line 10; page 18, text 2 line 4; page 25, text 3 line 1; page 35, text 4 line 1; page 38, text 5 line 1. *Sà* 卅 appears at page 44, text 7 line 4.

29. 々 is one of the “ditto marks” in handwritten Chinese, normally representing a repetition of the previous character.

30. For detailed discussion of these and other examples, see David Prager Branner,

Table 3.8 Guodian “punning” ligatures, a type of composite graph.

	Ligature	Combines these two graphs		Meaning
Manuscript graph				to stand side by side
Modern normalization		並(竝)	立	
Manuscript graph				grandee (“great man”)
Modern normalization		大	夫	
Manuscript graph				to be in dire straits (“one’s self is used up”)
Modern normalization		身	窮	
Manuscript graph				land
Modern normalization		土	地	

Note: Examples from the Guodian manuscripts. See Zhang Shouzhong, Sun Xiaocang, and Hao Jianwen, *Guodian Chu jian wenzibian* (Beijing: Wenwu Chubanshe, 2000).

Table 3.9 Medieval portmanteau graphs, a type of composite graph

Character	Presumably to be interpreted analytically as	Representing this word and conventional graph	When attested
諝	<i>qiǎoyán</i> 巧言 ‘facile words’	<i>biàn</i> 辯 ‘to argue’	6th century
懋	<i>bǎiniàn</i> 百念 ‘one hundred concerns’	<i>yōu</i> 憂 ‘worry’	6th century
墨	<i>míngkōng</i> 明空 ‘to brighten the void’	<i>zhào</i> 照 ‘to shine on’	7th century
逃	<i>wàizǒu</i> 外走 ‘run away’	<i>táo</i> 逃 ‘to flee’	8th century

“Portmanteau Characters in Chinese,” to appear in the *Journal of the American Oriental Society* 131.1 (2012): 73-82.

What is significant is how late these true composite graphs are attested in Chinese history, probably not much before the middle of the sixth century. Even counting the earlier manuscript ligatures, the first clear examples of analytically composite graphs — two whole graphs being combined in such a way that their meanings are seen to combine, *as words* — come very late in the history of written Chinese. Both ligatures and portmanteaux, in any case, are varieties of explicitly analytic character structure, since the components represent whole words rather than meanings independent of their vocalized form.³¹






Now compare those analytic composite graphs with this chapter's so-called complex pictographs in table 3.10.³² These graphs may be explained as either crypto-phonograms or semantic compounds; indeed, they may not be complex pictographs at all and may have been created as compounds to begin with. But it is not necessary to analyze them in pieces as is done with ligatures and portmanteau words; they can and should be recognized as units. That is to say, no matter whether they can be analyzed phonogrammatically or not, they are fundamentally logographs: whole symbols for whole words. Asserting that Chinese characters are logographs is a matter not of philosophical viewpoint but of reproducible measurement: a fluent reader's brain recognizes them as whole words or whole morphemes.

Long after Boodberg's time, it was demonstrated empirically that there is a "word superiority effect" in the reading of Chinese characters and Japanese *kanji*, just as in the reading of alphabetic and even patently phonetic scripts — meaning that it is measurably more efficient for the brain to recognize whole and meaningful written words than to piece those words together analytically out of their components. Fluent readers of Chinese tend to recognize whole characters rather than analyzing them by

31. It may be objected that a portmanteau character involves elements that can actually be vocalized as words while a putative ideograph or pure semantic compound involves non-vocalized elements, and that these are two different things. But the experimental evidence for logography means that all elements with meaning must function essentially as though vocalized. Though there is no explicit support for this particular image from the Chinese linguistic world, one may recall Ezra Pound's ideographic explanation of *xin* 信 'sincere' as "man standing by his word"; Ezra Pound, Appendix to *The Chinese Written Character as a Medium for Poetry*, by Ernest Francisco Fenollosa (San Francisco: City Lights Books, 1936), 41.

32. For the example of 毓, see Boltz, *The Origin and Early Development*, 110–13; for 鯨, see *ibid.*, 123–25.

Table 3.10 “Complex pictographs,” perhaps not composite

Ancient graphs	As complex pictograph, apparently represents	Modern graph seems to be composite of	Reading and gloss of recognized word	Crypto-phonetic explanations?
OBI  HJ: 27145 Bronze  JC: 5396	parturition: woman giving birth to child	毓(育): 每 + 充; not obvious as modern graphs	L-uk (覺): to give birth, rear	充 = inverted 孛 ‘child’, i.e., 子: related to 孝 {L-uk (覺)}
Bronze  JC: 2200	fish on a line	鯨: 魚 ‘fish’ + 系 ‘thread’	KW-in (真): fish, big fish	系 = 玄: related to 玄 {KW-in (真)}
OBI  HJ: 13662	rot in the teeth	龋/馮: 牙 or 齒 ‘tooth’ + 禹 ‘bug, beast’ (<i>Shuōwén</i>)	KW-a (魚): rotten tooth	(<i>Shuōwén</i> says 禹 is phonetic)
Bronze  JC: 4315	rice bent over with full ear	粳: 禾 ‘rice’ + 粿 ‘fine markings’ (<i>Shuōwén</i>)	L-iwk (覺): fine-looking rice	(<i>Shuōwén</i> says 粿 is phonetic, but it is unattested before <i>Shuōwén</i>)

decomposition as do non-fluent native and foreign learners.³³ Even multicharacter words seem to be processed as whole words rather than character by character.³⁴ It has been suggested that the brain must have an

33. Hiroyuki Kaiho and Hirofumi Saito, “Measuring Various Aspects of Kanji (Chinese Characters) and Its Psychological Implications,” *Quantitative Linguistics* 39 (1989): 151–63; Fook Kee Chua, “Visual Perception of the Chinese Character: Configural or Separable Processing?” *Psychologia* 42.4 (1999): 209–21; Xiangzhi Meng, Hua Shu, and Xiaolin Zhou, “Children’s Chinese Characters Structure Awareness in Character Output,” *Psychological Science (China)* 23.3 (2000): 260–64.

34. Naoko Sakuma, Itoh Motonobu, Sasanuma Sumiko, “Recognition Units of Kanji Words: Priming Effects on Kanji Recognition,” *Shinrigaku kenkyū* 心理学研究 (The Japanese journal of psychology) 60.1 (1989): 1–8; Ignatius G. Mattingly and Yi Xu, “Word Superiority in Chinese,” *Haskins Laboratories Status Report on Speech Research* 113 (1993): 145–51.

“orthographic processing system” for indivisible character components, but the exact positioning and complexity of those components seem to be essential to that system; components are not recombined freely before the whole character is recognized.³⁵

Still more significantly, a long series of experiments since the 1980s has shown that recognizing a Chinese character or Japanese kanji has an immediate phonological implication in the brain.³⁶ Whether or not it has an explicit phonogram structure, every character has a phonological value for the brain as soon as it is recognized. For fluent readers, there is no lag between recognition of a character — due to graphic, analytic, or other processes — and assignment of its phonological value.

It may, of course, be argued that all these experiments have been conducted on people who are already familiar with explicitly phonetic writing, in the form of *zhùyīn fúhào* 注音符號, pinyin, or *kana* 仮名, and that these findings do not explain decisively how people read characters in the days before phonetic scripts were widespread. But for the foreseeable future, experiments such as these will be the only measurements that can be brought to the study of this problem. The neural basis of modern

35. Marcus Taft and Xiaoping Zhu, “Submorphemic Processing in Reading Chinese,” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23.3 (1997): 761–75; Guosheng Ding, Danling Peng, and Marcus Taft, “The Nature of the Mental Representation of Radicals in Chinese: A Priming Study,” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30.2 (2004): 530–39. It is regrettable that both phonetic and semantic components are usually called “radicals” by experimental researchers, confusing *bùshǒu* 部首 ‘dictionary section-heads’ with *bùjiàn* 部件 ‘component parts’ or ‘elements’. Some authors do, however, explain which sense they mean. It would be better if experimenters followed normal sinological usage.

36. A major review of the literature may be found in Li-hai Tan and Charles A. Perfetti, “Phonological Codes as Early Sources of Constraint in Chinese Word Identification: A Review of Current Discoveries and Theoretical Accounts,” *Reading and Writing* 10.3-5 (1998): 165–200. See also Lien-chong Mou and Nancy S. Anderson, “Graphemic and Phonemic Codings of Chinese Characters in Short-term Retention,” *Bulletin of the Psychonomic Society* 17.6 (1981): 255–58; Connie Suk-Han Ho and Peter Bryant, “Learning to Read Chinese beyond the Logographic Phase,” *Reading Research Quarterly* 32.3 (1997): 276–89; Connie Suk-Han Ho, “The Importance of Phonological Awareness and Verbal Short-term Memory to Children’s Success in Learning to Read Chinese,” *Psychologia* 40.4 (1997): 211–19; Chih-Wei Hue and James R. Erickson, “Short-term Memory for Chinese Characters and Radicals,” *Memory & Cognition* 16.3 (1988): 196–205; Sumiko Sasanuma, Sakuma Naoko, and Kitano Kunitaka, “Reading Kanji without Semantics: Evidence from a Longitudinal Study of Dementia,” *Cognitive Neuropsychology* (United Kingdom) 9.6 (1992): 465–86.

reading was not necessarily “prefigured” in the historical evolution of the script, but it is axiomatic that writing systems evolve through a process of optimization, which includes cognitive optimization as well as societal pressures and surely imponderables now unknown to us.

Thus there is no reason to assume that a compound character lacking an obvious phonetic is meant to be analyzed semantically. The brain recognizes it today as a whole graph, and in the absence of other evidence, it is reasonable to assume that this has always been the case. Such a graph may be characterized as a complex but unitary pictograph.

The Value of Phoneticism

If all characters, regardless of their structure, are read as logographs, then what is the significance of phonogram structure, whether patent or concealed? Experimental research has been inconclusive to date. In the 1980s, a number of studies suggested that the presence of a phonophore facilitates recognition of low-frequency but not high-frequency characters.³⁷ More recent studies suggest that reading high-frequency characters, too, is facilitated by phonophores,³⁸ but others suggest facilitation by non-phonetic components as well.³⁹ It seems that the most important point of disagreement is the mechanism by which “recognition” is facilitated. It is experimentally clear, however, that the presence of a phonetic speeds recognition.

In the absence of a settled experimental explanation, let us approach phoneticism by returning to another point made by Bottéro, involving variant graphs. She cites *shī* 豕 ‘pig’, which has a likely reading of {L?-ij (脂)} but which in *zhú* 逐 ‘to chase’ she suggests might have the reading

37. See the review in Li-hai Tan et al., “Phonological Codes,” 191–94.







38. Alexander Pollatsek et al., “The Role of Phonological Codes in Integrating Information across Saccadic Eye Movements in Chinese Character Identification,” *Journal of Experimental Psychology: Human Perception and Performance* 26/2 (2000): 607–33; Jie-Li Tsai et al., “Use of Phonological Codes for Chinese Characters: Evidence from Processing of Parafoveal Preview When Reading Sentences,” *Brain and Language* 91/2 (2004): 235–44.

39. Xiaolin Zhou et al., “Is There Phonologically Mediated Access to Lexical Semantics in Reading Chinese?” in *Reading Chinese Script: A Cognitive Analysis*, ed. Jian Wang et al. (Mahwah, NJ: Lawrence Erlbaum Associates, 1999), 135–71; Xiaolin Zhou et al., “The Relative Time Course of Semantic and Phonological Activation in Reading Chinese,” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26/5 (2000): 1245–65; Ding, “The Nature of the Mental Representation,” 530–39.


later written *zhū* 豬 ‘pig’ {T-a (魚)}.⁴⁰ She notes that in early inscriptions, the same word is written with what is plainly an antlered deer in place of the pig and observes that, under the circumstances, an element that is replaced in variant forms of the compound graph cannot be claimed as phonophore.



There are two answers to this. The first is that *lù* 鹿 ‘deer’ is more plausible than either *shǐ* 豕 or *zhū* 豬 as the phonetic element of *zhú* 逐, since both initial and rime are much closer, as shown in table 3.11. The reader who has entered into the spirit of crypto-phonogrammism may

Table 3.11 Is *zhú* 逐 a phonogram?




	Component elements		Compound graph
Ancient graph	OBI  HJ: 20197	OBI  HJ: 28334	OBI  HJ: 10654
Modern graph	止	鹿	逐
Meaning	foot; to stop	deer	to chase
Readings	{T-i (之)}	{L-uk (屋)}	{L-iwk (覺)}
	Semantic component	Crypto-phonetic?	
	Component elements		Compound graph
Ancient graph	OBI  HJ: 20197	OBI  HJ: 10307	OBI  HJ: 35263
Modern graph	止	豕	逐
Meaning	foot; to stop	pig	to chase
Readings	{T-i (之)}	{L?-ij (脂)}, based on received <i>shǐ</i> or *{T-a (魚)}, standing for the word <i>zhū</i> now written 豬	{L-iwk (覺)}
	Semantic component	Crypto-phonetic?	

40. The status of the early Chinese initial of *shǐ* 豕 is uncertain, hence I have marked it with a question mark. The likeness of the Mandarin readings *zhū* and *zhú* is wholly fortuitous.

wonder, first, whether 逐 is not merely a variant of the true phonogram  rather than the other way around and, second, whether *zhǐ* 止 does not have a second reading that could be phonetic in *zhú* 逐.⁴¹

But turning Bottéro's objection around offers a second answer, an answer to what she is really asking: If  and 逐 represent the same word, how can the part that varies be the phonophore? The answer is another question: How do we actually know that  and 逐 represent the same word in the first place?

A historical phonologist relies wholly on specialists for the identification and normalization of ancient graphs. The process of identification consists of a series of guesses that are concealed by the final equation of an ancient graph X with a specific modern word Y, and that process has been carried out for the most part by people who are traditionally trained and believe in the validity of semantic compounds. Crypto-phoneticism was first proposed only seventy years ago in the West and remains very poorly known among Chinese scholars. Sun Haibo 孫海波 (1909–1972) and the other specialists who compiled and revised *Jiaguwen bian* 甲骨文編 surely never considered that 鹿 might be phonetic in 逐, just as the *Shuowen* identifies the phonogram *xìn* 信 as a semantic compound.

The main pieces of information that equated  and  were surely graphic parallelism as well as usage in context. The equation itself is not in doubt. But without the modern graph 逐 and its association with the word *zhú* to equate with , and without the context of OBI sentences in which

41. *Zhǐ* 止 is widely seen in early graphs, in places where it has been replaced (as here, by *chuò* 𠄎/𠄏) or omitted in the process of normalization. Too little is known about the phonological value of *chuò* {apparently T? N? L?-ak (鐸)} to hazard a guess as to whether it might be the phonetic in *zhú* 逐, although on the basis of the apparent rime group, that seems unlikely.

it appears to be used as a verb describing an action followed by the capture of an animal, how would we know that “foot” was combined with “pig” to mean “to chase” in the first place? Why not “to track or follow (prey),” “to approach (prey) in stealth,” “to harry (prey),” “to run (said of prey),” “to walk (a stud boar)” instead?

Outside of contextual usage, there is no obvious heuristic for choosing which word is meant by a semantic compound graph. To decode it analytically inevitably risks arbitrariness on one side and circular reasoning on the other.

In contrast, explicitly phonetic writing makes readers much more certain about the word they are reading. It lets them identify the word from the sound, assuming they actually know the word (surely a necessary precondition of literacy). Systematically speaking, a phonetic compound represents a specific word more parsimoniously than does a semantic compound, and the larger the number of explicit phonograms in the inventory of graphs, the more efficiently that inventory can be learned.

Another advantage is that unfamiliar characters can be assigned a reading instantly; whether correctly or not is secondary to the issue of speed. Anecdotally, it is common to encounter phonograms pronounced not according to their dictionary definitions but using the isolation reading of one of the phonetic elements. The examples in table 3.12 use the names of edible plants.⁴² This has been a very common process in Chinese: the

Table 3.12 Phonograms popularly assigned the modern reading of the historical phonophore

Standard name in dictionaries	Name commonly heard	Characters	Meaning	Graphic substitution implied in the name commonly heard
<i>gǒuqǐzi</i>	<i>gǒujǐzi</i>	枸杞子	wolfberry fruit	<i>qǐ</i> 杞 read as <i>jǐ</i> 己
<i>qiánmá</i>	<i>xúnmá</i>	蕁麻	nettle	<i>qián</i> 蕁 read as <i>xún</i> 尋
<i>jìcài</i>	<i>qí cài</i>	薺菜	shepherd's purse	<i>jì</i> 薺 read as <i>qí</i> 齊
<i>piělan</i> (in Běijīng dialect, the traditional name is <i>piěla</i>)	<i>pīlan</i>	苳藍	kohlrabi	<i>piě</i> 苳 read as <i>pī</i> 丕

42. These examples were collected in ongoing interviews with literate native speakers of accented Mandarin in New York City, beginning in May, 2004.

imposition on speech of the phonology embodied in the script, without a direct phonetic intermediary.

In sum, the significance of phonetic components in the script is the economy they bring about. Any such economy streamlines the process of reading.

Continuity of Literacy

So the systematic value of phoneticism is not in doubt. What then are the implications of the crypto-phonogram hypothesis for literacy?

A key idea of the hypothesis is that phonetic information originally used in the construction of some characters has been lost. Evidently, at some point in the past, such information was not recoded by the normalization of older graphs as new phonograms. This suggests that literacy (in the sense chiefly of reading rather than writing) was not a continuous tradition in Early China, quite a surprising idea, considering the efficiencies that phoneticism bestows on learning and reading a script. The notion that Chinese written and spoken languages are of separate origin has been alive in the West since the time of Étienne Fourmont (1683–1745) and should not be dismissed out of hand, but there are other explanations — for instance, that some of the spoken words or readings necessary to interpret characters phonogrammatically became extinct early in the process of communication.

If modern views are correct about the genetic relationship between Chinese and Tibeto-Burman, then there must have been a significant shift in the typology of spoken Chinese, perhaps through creolization and presumably due to language contact brought about by colonization and migration.⁴³ But literacy must have been a continuous tradition on the whole. The primary reason is the apparent integrity of the *xiěshēng* series, the gross inventory of phonetic elements in the set of all phonograms.

As an example, take the phonophore 𠄎. It is not identified in *Shuowen* and does not seem to have been distinguished anciently from what is now written 片: the elements 𠄎 and 𠄎 are apparently interchangeable in OBI and bronze script. Xu Kai 徐鍇, in the tenth century, equates it with *chuáng* 牀 ‘bed, frame for supporting something’ and also apparently

43. David Prager Branner, *Problems in Comparative Chinese Dialectology: The Classification of Min and Hakka*, Trends in Linguistics series, no. 123 (Berlin: Mouton de Gruyter, 2000), 160–66.

with *chuáng* 疒 ‘sickness; sick abed’. The ancient graph identified as its antecedent is a complex pictograph that appears to depict a person on a bed in the presence of particles or drops. Zheng Qiao 鄭樵, in the twelfth century, considers it homophonous with *qiáng* 牆 ‘wall’ and glosses it ‘to kill; also, a split piece of wood’ (殳也、亦判木也).⁴⁴ Its early Chinese reading, based on either modern word, is {S-ang (陽)}. Table 3.13 shows 𠄎 appearing as a phonetic element in a number of ancient graphs that, if identified correctly, remain known today but share no obvious etymological content.

Being able to identify correspondence sets like this one — a list of phonograms sharing a phonetic element and with attested medieval readings that allow for reconstruction of an early Chinese reading — is a statement about the linguistic identity of the graphs but not necessarily about the cognatehood of the words they represent.⁴⁵ Continuity between the ancient graphs and the medieval readings of their normalized forms implies continuity between the language connected with the ancient script and the medieval language that by definition is Chinese as we now know it. Even if the *xiéshēng* series should turn out to have been borrowed whole from somewhere else, the fact that it appears to correspond regularly to later phonology means that there is continuity from at least the time of the borrowing until the era of the later phonology.

It is difficult to test assumptions about such continuity, because continuity is routinely taken for granted in identifying characters. Consider the principle of the *chūwén* 初文 or “protoform” (as Mattos and Norman translate it). An ancient graph X is identified as representing not the morpheme with which it is associated in later records, but some other (and semantically more basic) morpheme that is usually also attested today, written with a different graph Y. The morpheme with which X is usually associated in later records is assumed to be a loan usage. Phonetic loaning is thought to be the same process that underlies the *xiéshēng* series, but in the case of protoforms, both the sound and meaning of a modern word (usually in rime-book phonology) is involved.

44. Zheng Qiao 鄭樵, *Tongzhi* 通志 (“Liushu lüe 六書略”). Wang Shumin 王樹民 (Beijing: Zhonghua Shuju, 1995), 1.254.

45. On the nature and meaning of correspondence sets, see Branner, *Problems in Comparative Chinese Dialectology*, 24–37.



Table 3.13 Words and graphs in the *chuáng* 𠄎 *xiéshēng* series

Ancient graph	Modern normalization	Gloss	Medieval reading	Baxter 1992 reconstruction
OBI 𠄎 HJ: 13753	𠄎 = 牀	bed, frame for supporting something	{dzrang ₃ }	*dzrjang
	疒	sickness, sick abed		
OBI 𠄎 HJ: 35301	戕	to kill brutally	{tsang ₁ }	*tsang
OBI 𠄎 HJ: 5652	妝	to adorn	{tsrang ₃ }	*tsrjang
Bronze 壯 JC: 2840.2	壯	robust	{tsrangH ₃ }	*tsrjangs
Bronze 𠄎 JC: 154	𠄎 (=醬)	meat sauce	{tsangH ₃ }	*tsjangs
Bronze 𠄎 JC: 100	臧	good	{tsang ₁ }	*tsang
Bronze 𠄎 JC: 10175	牆	wall	{dzang ₃ }	*dzjang
Bronze 𠄎 JC: 2588	𠄎 (=莊)	impressive	{tsrang ₃ }	*tsrjang
Bronze 𠄎 JC: 118	斨	axe-head	{tshang ₃ }	*tshjang

Note: For an explanation of the transcription of medieval phonology, please see the tables in David Prager Branner, “Appendix II: Comparative Transcriptions of Rime Table Phonology,” in *The Chinese Rime-tables: Linguistic Philosophy and Historical-comparative Phonology* (Amsterdam: John Benjamins, 2006), 265–302.

A classic application of the protoform principle is to explain the original graph *yǒng* 永 as meaning not “eternally” but “to swim” (table 3.14).

Table 3.14 *Yǒng* 永 as protoform of *yǒng* 泳 ‘to swim’

	Ancient graphs	Normalized modern graphs	Gloss	Reading
Graphs	OBI  HJ: 25759	永	eternally	{KW-ang (陽)}
	Bronze  JC: 2829	派 (=派)	tributary	{P-e (支)}
Possible protoform of		泳	to swim	{KW-ang (陽)}






In other words, *yǒng* 永 in the modern sense “eternally” is a phonetic loan, and the graph  was originally created to write *yǒng* ‘to swim’, now written 泳. This  appears to be a complex pictograph, consisting of the figure of a person surrounded by lines suggesting water. Table 3.15 shows another example.







Table 3.15 *Yě* 也 / *tā* 它 as protoform of *shé* 蛇 ‘snake’

	Ancient graphs	Normalized modern graphs	Gloss	Reading
Graphs	OBI  HJ: 22296	也	[particle]	{L-aj (歌)}
	Bronze  JC: 10243	它	other	{L-aj (歌)}
Possible protoform of		蛇	snake	{L-aj (歌)}

Yě 也 (particle) and *tā* 它 ‘other’ as read today are phonetic loans and (according to this theory, for there are others)  were originally created as pictographs to write *shé* ‘snake’, now 蛇. In table 3.16, two modern words are possible sources of the basic graph.



The modern particle *qiě* and an obsolete word *jū* ‘good’ are loangraphs, and the form 且 is a pictograph depicting some sort of board, perhaps a chopping block (modern *zǔ* 俎) or ancestral tablet (modern *zǔ* 祖, now ‘ancestor’), or perhaps they were the same word. (Competing explanations, for instance, that the original graph represents a word for “penis,” are not supported by protoform evidence.)

Table 3.16 *Qiě-jū* 且 as protoform of *zǔ* 祖 ‘ancestor’ and *zǔ* 俎 ‘butcher block’

	Ancient graphs	Normalized modern graphs	Gloss	Reading
Graphs	OBI  HJ: 1714	且	[particle]	{S-a (魚)}
	Bronze  JC: 2818		good/[particle]	{S-a (魚)}
Possible protoform of	OBI  HJ: 22999	祖	ancestor	{S-a (魚)}
	Bronze  JC: 271			
Possible protoform of	OBI  HJ: 18604	俎	butcher block	{S-a (魚)}
	Bronze  JC: 9727			

Another example is shown in table 3.17.

Table 3.17 *Yì* 亦 as protoform of *yè~yì* 腋 ‘armpit’



	Ancient graphs	Normalized modern graphs	Gloss	Reading
Graphs	OBI  HJ: 36511	亦	[particle]	{K-ak (鐸)}
	Bronze  JC: 2833			
Possible protoform of		腋	armpit	{K-ak (鐸)}

Yì 亦 is seen as a loangraph when it represents the grammar particle; its original meaning would be the word *yè~yì* ‘armpit’, now written 腋.⁴⁶ Such examples typically involve the same sound-bearing element in the two characters; since they are part of the same phonetic series, no revision of the phonological interpretation is needed.

Table 3.18 illustrates another protoform relationship.

46. The Mandarin pronunciation *yì* is consistent with medieval records; *yè* appears to be a case of adopting the reading of the phonetic component, mentioned on p. 107.



Table 3.18 *Yōu* 攸 as protoform of *dí* 滌 ‘to rinse’

	Ancient graphs	Normalized modern graphs	Gloss	Reading
Graphs	OBI  HJ: 39515 Bronze  JC: 4344	攸	smoothly and swiftly (of movement); [particle]	{L-iw (幽)}
Possible protoform of		滌	to wash, rinse	{L-iwk (覺)}

The ancient forms of *yōu* 攸 seem to be complex pictographs depicting a hand applying an object to a person; the bronze includes dots, which are found in a great many graphs and often are interpreted as particles or drops of liquid. This suggests that the protoform of *yōu* 攸 is likely the word *dí* ‘to wash or rinse’, now written 滌. Phonologically, it is a good match; whether this equivalence is supported in actual usage is a separate question.

Table 3.19 shows another relationship that has not yet been published.

Table 3.19 *Wéi* 為 as protoform of *huī* 擣 or *huī* 麾 ‘to lead’

	Ancient graphs	Normalized modern graphs	Gloss	Reading
Graphs	OBI  HJ: 40482 Bronze  JC: 226	為	to take as, for the purpose of, because of	{M-aj (歌)}
Possible protoform of		擣	to lead, direct	{M-aj (歌)}
		麾	to lead, direct	{M-aj (歌)}

Ancient forms of present-day *wéi*~*wèi* 為/爲 appear to be complex pictographs depicting a hand near the head of an elephant, protoforms of *huī* 擣 and *huī* 麾, both ‘to lead or direct’. They are homophones in medieval phonology as in Mandarin; Baxter’s reconstructions differ slightly, 擣 **hw(r)jaj*, 麾 **hm(r)jaj*, because of their different graphic



structures, but phonologically, both relationships appear to be regular.⁴⁷


In all six examples above, the ability to propose protoform relationships depends on assuming continuity between the language of antiquity and the medieval compendia of character readings. So, naturally, protoform relationships cannot also be used as evidence of such continuity, which would be circular reasoning. A premise can be falsified by showing that its logical implications are false, but it cannot be shown to be true using those implications. The premise may indeed be true, but if the evidence is circular, the truth of the premise cannot actually be known, even if we believe it is true.

Taking this a step further, let us admit that the protoform principle is inherently circular, because we can propose the relationship only if the modern word or a word family exists. And because of that circularity, protoforms are sometimes proposed in ways that actually seem to undermine the whole conception of phonological continuity between the period of ancient characters and medieval phonology.

Table 3.20 shows an example of undermined continuity; it is well attested but phonologically less straightforward than the examples in tables 3.14–3.19.

Table 3.20 *Zì* 自 as protoform of *bí* 鼻 ‘nose’






	Ancient graphs	Normalized modern graphs	Gloss	Reading
Graphs	 HJ: 3449  JC: 3618	自	oneself	{S-it (質)}
Possible protoform of		鼻	nose	{P-it (質)}

47. In a competing, crypto-phonetic explanation, Jerry Norman has observed that the Vietnamese word for “elephant,” *voi*, could conceivably be phonetic in  (personal communication, 1991). He has also suggested that Vietnamese words for “mouth” *miệng* and “mouthful” *miếng* could account for the crypto-phonetic reading of 口 in 名 and 鳴; see Boltz, *The Origin and Early Development*, 63 n. 14. Vietnamese forms are from Nguyễn Đình-Hoà, *Essential English-Vietnamese Dictionary* (Boston: Tuttle Publishing, 1983).

This relationship is often interpreted as evidence of polyphony.⁴⁸ 自, now read *zì* {S-it (質)}, is presumably the protoform of the modern word *bí* {P-it (質)} ‘nose’, now written 鼻, which includes a phonetic. But the fact that the rimes are the same suggests to some that a phonetic, *bì* 畀, loan was intended. If that is so, this relationship violates the basic principle we now impute to all phonetic loans: that of consonantal homorganicity, first detailed in modern linguistic terms by Li Fang Kuei.⁴⁹

The protoform relationship is evidence of the linguistic continuity that is routinely taken for granted in identifying characters. In practice, however, many protoforms seem to be suggested on graphic grounds alone and make no sense phonologically. Table 3.21 shows the first of three widely seen examples.

Table 3.21 *Xiǎo* 小 and *shǎo~shào* 少 as protoforms of *shā* 沙 ‘sand’

	Ancient graphs	Normalized modern graphs	Gloss	Reading
Graphs	OBI  HJ: 40314 Bronze  JC: 2678	小	small	{S-ew (宵)} or {L?-ew (宵)}
	OBI  HJ: 20960 Bronze  JC: 2782	少	few	{L?-ew (宵)}
Possible protoform of	Bronze  JC: 4321	沙	sand	{L? P? K?-aj (歌)}

By no stretch of the usual phonological correspondences can a word in rime {-ew (宵)} serve as phonophore for one in rime {-aj (歌)}. Either there is another reading missing somewhere or the word *shā* ‘sand’ has nothing to do with the ancient graphs for words meaning “small” and “few.”⁵⁰



Table 3.22 shows another example.

48. See Qiu, Chinese Writing, 9–10.

49. Boodberg, however, reconstructed a cluster initial “BS-”; Boodberg “Some Proleptical Remarks,” 340–42.

50. Somewhat less improbable is the idea that 水 ‘water’ {L?-uj (微)} is phonetic in 沙 ‘sand’ {L? P? K?-aj (歌)}.





Table 3.22 *Bù~pǐ~fǒu* 不~否 as protoforms of *fū* 榦 ‘drum-stand, calyx’

	Ancient graphs	Normalized modern graphs	Gloss	Reading
Graphs	OBI  HJ: 14762 Bronze  JC: 2810	不~否	[particles]	{P-i (之)}
Protoform of		榦	legged drum-stand, calyx of a flower	{P-o (侯)}

Xiéshēng contact between the early Chinese rime groups {-i (之)} and {-o (侯)} is known, although less common than {-i (之)} contact with {-i (之)}. If we say that the ancient graph for 不 actually depicts the object named by the word *fū* ‘drum-stand with legs’, we are saying we can accept protoform relationships that are not close homophones.

Table 3.23 presents one more such case.

Table 3.23 *Fán* 凡 as protoform of *pán* 盤 ‘platter’

	Ancient graphs	Normalized modern graphs	Gloss	Reading
Graphs	OBI  HJ: 33146 Bronze  JC: 2838	凡	[particle]	{P-om (侵)}
Protoform of	OBI  HJ: 2205 Bronze  JC: 10137	盤	platter	{P-on? (元)} {P-an? (元)}

This, too, is a widely seen explanation but is phonologically irregular. We are asked to accept that a word with a reading {P-on (元)} has been borrowed to write one read {P-om (侵)}. Under the principle of homorganicity, that would be considered a long shot, especially because a number of words that are read {P-i/um (侵)} and written with 凡 as the phonetic (such as 風, 芑, and 鳳) are found in medieval sources to have

merged with words that are read {P-ing (蒸)} (e.g., 馮, 夢) and {P-ong (東)} (e.g., 豐). Contact between {-on (元)} and {-om (侵)} is considered rare.

The best kind of evidence for continuity in the tradition of early Chinese literacy is that which connects the early script to mainstream Chinese words in later periods. The best example is the extant *xiéshēng* series. The practice of identifying ancient graphs as the protoforms of later words also assumes such continuity, although it flirts with circular logic and is sometimes carried out in ways that violate the usual rules of sound correspondence between early and medieval Chinese phonology, which by implication denies the continuity of the *xiéshēng* series.

Or could there be something wrong with those rules of sound correspondence?

Functional Homophony in an Underspecified Phonological Representation

Let us admit that the finely detailed early Chinese reconstructions of the Karlgren model must be regarded as fictions.⁵¹ The Karlgren model takes the precise rhyming distinctions of the late fifth and sixth centuries, when for a brief century or two it was the fashion to observe such distinctions minutely, and imposes that precision on the early period, whose chief form of phonological expression was the phonophore. Surely no reconstruction should be more precise than its sources are accurate. But what notion of phonology underlies the *xiéshēng* series and loangraphs? Simply to say that Chinese phonophores are logographs is too weak; that would be the situation for native Japanese words (*kun* 訓-readings) written with the phonophores of the 𠄎 series shown on page 109 (readings from Morohashi) (see table 3.24).⁵²

51. Surely the starting point for all such discussions should be Jerry Norman and W. South Coblin, "A New Approach to Chinese Historical Linguistics," *Journal of the American Oriental Society* 115.4 (1995): 576–84. But even that paper fails to address the idea that traditional Chinese expressions of phonology (in both the rime books and the *xiéshēng* series) are themselves formal representations, to whatever extent they were intended to be prescriptive.

52. Morohashi Tetsuji 諸橋轍次, *Dai Kan-Wa jiten* 大漢和辭典 (Tokyo: Taishūkan Shoten, 1955–60).

Table 3.24 *Kun* 訓-readings (native Japanese words) associated with graphs in the *chuáng* 𠄎 *xiéshēng* series

Character	Medieval Chinese reading	Common Japanese <i>on</i> -readings	Variety of attested Japanese <i>kun</i> -readings	Meaning
𠄎	{dzrang ₃ }	—	<i>nedai</i>	sleeping platform
牀	{dzrang ₃ }	<i>sō, shō</i>	<i>toko</i>	bed
			<i>yuka</i>	floor
			<i>nedai</i>	sleeping platform
			<i>koshikake</i>	seat
			<i>sunoko</i>	indoor platform
壯	{tsrangH ₃ }	<i>sō</i>	<i>sakan</i>	vigorous
			<i>isamashī</i>	courageous
醬	{tsangH ₃ }	<i>shō</i>	<i>hishio</i>	salted meat
臧	{tsang ₁ }	<i>zō</i>	<i>yoi</i>	good
			<i>atsui</i>	thick
牆	{dzang ₃ }	<i>shō</i>	<i>kaki</i>	fence
妝	{tsrang ₃ }	<i>sō</i>	<i>yosō</i>	style of dress
莊	{tsrang ₃ }	<i>sō, shō</i>	<i>ogosoka</i>	solemn
			<i>tsutsushimu</i>	discreet
戕	{tsang ₁ }	<i>zō</i>	<i>koroso</i>	to kill
			<i>sokonau</i>	to harm

In contrast, the Chinese script is attempting a degree of phonological representation not possible in Japanese. We can call the Japanese case (i.e., *kun* readings) pure logography because it represents whole words without considering their sounds at all, but the Chinese script is more properly “defective,” in the technical, linguistic sense of the word: it attempts to represent phonology but does so without full precision.

What notion of phonology underlies the *xiéshēng* series and loangraphs? In order to answer that question, let us consider the implications of saying that 凡 is the protoform of 盤 {P-on? -an? (元)} and was later borrowed to write an unrelated particle, read {P-om (侵)}. How common are such irregular relationships? Is it possible that early users of the script did not consider *-om* and *-on* to be different sounds?

Generally speaking, labial and dental sounds are very well segregated by the phonological sources of the medieval period, but it is possible that this is the result of normalization. In the *xiéshēng* series, there are two important examples of the conflation of {-p} and {-t} (table 3.25).⁵³

Table 3.25 P/t contact in the larger *rù* 入 *xiéshēng* series

*-t			*-p			
Graph	Meaning	Baxter 1992 reconstruction	Graph	Meaning	Baxter 1992 reconstruction	Rime group
內	within	*nuts	入	to enter	*njup	物部
世	generation	*hljats	葉	leaf	*ljap	月部

Li Fang Kuei believed that 內 and 入 represented cognate words and resolved the apparent problem by proposing that the *-t* coda had assimilated from an earlier *-p*, in contact with *-s* representing proto-*qùshēng*: 內 *nuts < **nups. Baxter resolved the problem of the second pair of words the same way: 世 *hljats < 葉 **hljaps.

In fact, these are not two isolated pairs of cognates; there are sets of common words in both phonetic series — that is, a larger phonological relationship encompassing several different words. Table 3.26 shows the 內/入 series.

For 訥, Baxter has proposed *nut (perhaps < **nuts < **nups). For the sense “to go/send in,” the graphs 內 and 納 did not begin to be distinguished until the period of standardization. If we believe that the *xiéshēng* series does not distinguish labial and dental codas, we are saying that {N-ut} and {N-up} actually represent one and not two syllables within the precision of the representation.

Similarly, for the 世/葉 series, we are saying that {L-at} and {L-ap} represent one and not two distinct syllables within the limit of tolerance (table 3.27).

53. Baxter’s full 1992 reconstructions are used in the discussion that follows, because the traditional early Chinese rime groups do not distinguish {-p} and {-t} codas in the rimes concerned. Since the rime groups themselves do not distinguish these two codas, they cannot be used to definitively construct reading-tokens (of the kind used elsewhere in this chapter) with labial-stop finals in the system of Baxter 1992.

Table 3.26 Other examples of p/t contact in the larger *rù* 入 *xiéshēng* series

*-t			*-p		
Graph	Meaning	Baxter 1992 reconstruction	Graph	Meaning	Baxter 1992 reconstruction
內	within	*nuts	入	to enter	*njup
訥	halting of speech	*nut	納	to go in, send in	*nup
訥	halting of speech	*nrjot	輓	inward-facing halter for outer horses on a team	*nup
焮	to burn (熱)	*njot			

Table 3.27 P/t contact in the larger *shì* 世 *xiéshēng* series

*-t			*-p		
Graph	Meaning	Baxter 1992 reconstruction	Graph	Meaning	Baxter 1992 reconstruction
世	generation	*hljats	葉/葉	leaf	*ljap
泄	to spread, leak	*sljat	蝶	butterfly	*lep
(Same)	garrulous	*ljats	鞞	archer's thumb-ring	*hljap

There is also a common word that seems to confuse *-m* and *-n*: 矰 ‘spear handle’, read {K-in (真)}; it has an ancient variant 矰, with 今 {K-im (侵)} suggesting *-m* coda. This example seems weak — it is isolated and depends on a variant graph.

There are, however, several cases of initials **m-/*n-* being conflated within a phonetic series in common words such as those shown in table 3.28.


To summarize, if we suppose that *xiéshēng* series were only a rough way of representing relationships in sound, then we can find a number of common words in which labial and dental codas and also nasal initials do not appear to be distinguished. It is not that there was no distinction between them in speech or reading aloud but that the script treats them as functionally homophonous — sufficiently congruent for the purpose of written representation.

Table 3.28 Examples of m/n contact in the *xiéshēng* series

*n-			*m-		
Graph	Meaning	Baxter 1992 reconstruction	Graph	Meaning	Baxter 1992 reconstruction
柔	pliable	*nju	矛	spear	*m(r)ju
耳	ear	*nji?	弭	tip of a bow	*mje?
爾	many (flowers)	*njaj?; njij?	彌	to end; distant	*mjej?

Perhaps even what we think of as polyphony may originally have been a kind of lax but functional homophony, within the limits of tolerance of the early users of this script. Here are two more pairs of words whose early graphs are considered “cognate” by many — that is, originally a single polyphonous form, later differentiated — and whose reconstructed pronunciations might be considered crudely alike if labial and dental nasals are allowed to be congruent for the purposes of the writing system. The first involves initials (table 3.29).

Table 3.29 A dubious example of m/n contact in the *xiéshēng* series

	*n-			*m-		
Ancient graphs	Modern graph	Meaning	Baxter 1992 reconstruction	Modern graph	Meaning	Baxter 1992 reconstruction
	女	woman	*n(r)ja?	母	mother	*m(r)o/i?

The second involves codas (table 3.30).

Granted, this last example is extreme, but Chinese paleography has already accustomed itself to this degree of phonological laxness. After all, consider how few distinctions there are among the early Chinese rime categories in their traditional twentieth-century paradigm, as used by Li Fang Kuei (table 3.31).

Table 3.30 A dubious example of m/n contact in the *xiéshēng* series

Ancient graphs	*-n			*-m		
	Modern graph	Meaning	Baxter 1992 reconstruction	Modern graph	Meaning	Baxter 1992 reconstruction
言	言	to speak	*ngjan	音	to make music	*ʔ(r)jim

Table 3.31 The Old Chinese rime reconstructions of Li Fang Kuei (1971)

Number	“Open” coda	Li 1971 reconstruction	Stop coda	Li 1971 reconstruction	Nasal coda	Li 1971 reconstruction
1	之部	*-əg	職部	*-ək	蒸部	*-əng
2	幽部	*-əgw	覺部	*-əkʷ	冬部	*-əngʷ
3	宵部	*-agw	藥部	*-akw	—	
4	侯部	*-ug	屋部	*-uk	東部	*-ung
5	魚部	*-ag	鐸部	*-ak	陽部	*-ang
6	支部	*-eg	錫部	*-ek	耕部	*-eng
7	歌部	*-ar	月部	*-at	元部	*-an
		*-ad				
8	脂部	*-id	質部	*-it	真部	*-in
9	微部	*-əd	物部	*-ət	文部	*-ən
10	—		緝部	*-əp	侵部	*-əm
11	—		葉部	*-ap	談部	*-am

Li uses a total of thirty-one rime categories. (Tonal distinctions are generally ignored in *xiéshēng* contact.) It is usual for words with open, stop, and nasal codas to be allowed to be in contact (*duìzhuǎn* 對轉 ‘contact between corresponding codas’), bringing the effective number of discrete rime categories down to eleven (the numbered rows of table 3.31), without even mentioning the tendency of allowing contact between separate rimes (*xiéyùn* 叶韻, such as between 之部 and 幽部).

Among initials, the system used in this chapter reduces the fifty-five simple and complex initials of Baxter’s 1992 system to ten basic types,

representing each of them using only its core consonant. It appears that this principle is a reasonably accurate model of *xiéshēng* initial-contact.

A system of ten initials and eleven rimes, without tones or medials, can distinguish only 110 syllables, a clear case of phonological underspecification. Quite possibly even 110 distinct syllables are more than the *xiéshēng* system is meant to represent, but suppose there were meant to be even twice or three times that many. (Since there are plenty of examples of loangraph contact between different *xiéshēng* elements after standardization, it seems unlikely that each element was intended to serve as a distinct unit of phonology.) Compare an inventory of two hundred or three hundred rough syllable types with the fine distinctions made in Karlgren's Chinois archaïque and its successor reconstructions, built on the model of the *Qièyùn* 切韻. How much more is lost by allowing 月部 to be merged with 葉部, and 物部 with 緝部? To look at early Chinese phonology this way goes against the philosophy of all reconstructive work since Karlgren's time and of the late fifth-century philosophy of euphony that led to the *Qièyùn*. But it follows the practice of the most eminent modern paleographers of Chinese and may represent the actual phonological system of the *xiéshēng* series.

If *xiéshēng* relationships are based on very crude distinctions, plausible crypto-phonetic explanations should be easy to find, leaving aside the more difficult question of whether they are historically valid.

Literacy and Diglossia

What does such a crude system of phonological representation suggest about early Chinese literacy?

The answer depends on how we imagine the system coming into being. William Boltz has proposed that after a period of rapid advance around the third century B.C.E., true phoneticism was arrested by conflict with what he calls "the inherited orthodox world-view," which favored the semantic explicitness of phonograms.⁵⁴ Imre Galambos considers that preference for phonograms to have been motivated by the adoption of the script by colonized peoples who would have used different spoken languages.⁵⁵

54. Boltz, *The Origin and Early Development*, 168–77.

55. Galambos, *Orthography of Early Chinese Writing*, 146–50. An even more extreme statement of this idea is contained in Jerry Norman's suggestion about Vietnamese words being at the bottom of some "Chinese" cryptophonograms, mentioned in note 47.

These views are compatible in that both see the dramatic growth of phonograms as part of a new process of normalization brought on by pressure for political unity around the time of Qin and early Han.

Assuming that to be so, any phonological relationships we can extract from the *xiéshēng* series must effectively date from around the time of the normalization, regardless of their prior lineage. Normalization was a watershed in the history of phonology because it introduced a systematic set of phonophores. Earlier graphic forms (such as most of those cited in this chapter) should therefore be kept strictly out of the study of *xiéshēng* phonology, though they remain of great interest for what may be called the prehistory of the phonology. As to the question of the continuity of literacy suggested by the integrity of correspondences between *xiéshēng* and later phonology, it, too, can only be claimed to go back to Qin–Han normalization. To ask (remembering Fourmont) whether the script was native or borrowed before that time is meaningless, because with normalization, it has become in effect a fresh system. And the crypto-phonogram hypothesis, which by definition asserts that vital phonological information has been omitted from normalized phonograms and lost to the heritage of literacy, also becomes a matter only of prehistory. That would explain the lack of continuity discussed above and also the presence of a residue of phonetically opaque non-phonograms within an increasingly phonogrammatic system.

What then is the relationship of the written phonological system to the rest of the language and to literacy? Of course, *xiéshēng* was cruder than the diverse distinctions of any real language — any hope of realistic phonetic representation was lost. But what was gained? It seems that diglossia was reduced, and that seems to have opened the door to a flowering of narrative and exegesis, sure evidence of self-conscious literacy.

Diglossia, the “caste system” of language, is suggested by the views of Boltz and Galambos quoted above, although neither author addresses it directly. In various forms, the basic idea is popularly expressed in the English saying “Never use a shilling word when a sixpenny one will do,” the idea being that some words are of higher “price” or “value” than others, or that socially they are more elevated.⁵⁶ More technically, the canonical

56. Apparently first attested in George Johnson, “Place-names,” in *The Educational Review Supplementary Readings*, ed. George Hay (Saint John, N[ew] B[runswick], Canada: Barnes & Co., 1900), 1898.4: 87–94, 88.

modern definition is that of Charles A. Ferguson (1921–1998):⁵⁷

Diglossia is a relatively stable language situation in which, in addition to the primary dialects of the language (which may include a standard or regional standards), there is a very divergent, highly codified (often grammatically more complex) superposed variety, the vehicle of a large and respected body of written literature, either of an earlier period or in another speech community, which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any section of the community for ordinary conversation.

There is textual evidence for pre-Qin diglossia in the form of a passage in the *Analects* in which the term *yǎyán* 雅言 ‘elegant language’ appears, surely referring to high-register language. It can be punctuated two different ways:

子所雅言，詩書執禮，皆雅言也。⁵⁸ (as traditionally punctuated)

Where the Master spoke elegantly: the *Shi* and *Shu* and the practice of ritual; in each of these he used elegant language.

子所雅言詩書，執禮皆雅言也。

Where the Master used [or What the Master considered] elegant language was the *Shi* and *Shu*; in carrying out ritual, he always spoke elegantly.

Some recent readers have conceived “elegant language” to mean a putative standard language, but it more likely refers to a high diglossic register, thought fitting for use with the most elevated texts and in ritual.⁵⁹

Why should *yǎyán* not be “standard language”? Actually, there are terms for something like standard language from the second half of the

57. Charles A. Ferguson, “Diglossia,” *Word* 15 (1959): 325–40, esp. 335. “Diglossia” is first attested in Karl Krumbacher, *Das Problem der neugriechischen Schriftsprache* (Munich: Königliche Bayerische Akademie, 1902); I am grateful to the late Alan Kaye (1944–2007) for supplying this reference (personal communication, 2004). For a review of developments in the application and philosophy of diglossia since the Ferguson article, see Harold F. Schiffman, “Diglossia as a Sociolinguistic Situation,” in *The Handbook of Sociolinguistics*, ed. Florian Coulmas (London: Basil Blackwell, 1997).

58. *Lunyu* 論語 (“Shu er” 述而), *Shisanjing zhushu*, 7:2482.

59. David Prager Branner, “‘Red Cliffs’ in Taiwanese *Hànbùn*,” *CHINOPEL Papers* 24 (2002): 67–100, esp. 91–95.

Han. The principal one is *tōngyǔ* 通語 ‘common language’, which appears in the *Fangyan* 方言. The *Fangyan* lists these words, arranged topically, and then sometimes cites a “common language” equivalent for them, as in the following examples:

頷、頤，頷也，南楚謂之頷。秦晉謂之頤。頤，其通語也。⁶⁰

Hàn 頷 and *yí* 頤 mean *hàn* 頷 ‘lower half of the face’. In Southern Chu it is called *hàn* 頷. In Qin and Jin it is called *hàn* 頷. *Yí* 頤 is the common term for these words.

甌、甌、甌、甌、甌、甌、甌、甌、甌、甌。靈桂之郊謂之甌，其小者謂之甌。周魏之間謂之甌。秦之舊都謂之甌。淮汝之間謂之甌。江湘之間謂之甌。自關而西、晉之舊都河汾之間、其大者謂之甌，其中者謂之甌。自關而東，趙魏之郊謂之甌，或謂之甌。東齊海岱之間謂之甌。甌，其通語也。⁶¹

Gāng 甌, *dǎn* 甌, *wǔ* 甌, *yóu* 甌, *zhèng* 甌, *chóng* 甌, *zhui* 甌, *wèng* 甌, *bùlǒu* 甌, and *yì* 甌 mean *yīng* 甌 ‘pitcher’. In the outskirts of Ling and Gui it is called *gāng* 甌, and the small variety is called *dǎn* 甌. Around Zhou Wei it is called *wǔ* 甌. In the old capital of Qin it is called *zhèng* 甌. Between the Huai and Ru Rivers it is called *yóu* 甌. Between the Yangzi and Xiang Rivers it is called *chóng* 甌. West of the Pass, in the old capital of Jin and between the Yellow and Fen Rivers, the large variety is called *zhui* 甌 and the medium-size variety *bùlǒu* 甌. East of the Pass, in the outskirts of Zhao Wei, it is called *wèng* 甌. In the eastern part of Qi and the Bohai-Dai region it is called *yì* 甌. *Yīng* 甌 is the common term for this thing.

Tōngyǔ is very clearly being used in contrast to regional language and must mean some kind of *koinē*. (It could not truly have been a “standard” unless promulgated as such by some authority.) In contrast, teaching “the *Shi* and *Shu* and the practice of ritual” are exactly the sorts of formal contexts in which by definition we would expect to encounter high diglossic language. So it seems that the notion of diglossia was already in existence before the normalization of the script.

There is also linguistic evidence for the early existence of diglossia in

60. Standard Chinese readings following Zhou Zumo 周祖謨, *Fangyan jiaojian* 方言校箋 (Beijing: Kexue Chubanshe, 1950 [1956]; repr., 1993), 10:35, 66.

61. Zhou Zumo, *Fangyan jiaojian* 5.10, 33–34.

the typological conflict between ancient writing and highly conservative modern dialects.⁶² Characters in the era before normalization would have had a set of readings not necessarily even close to spoken language — similar to the diversity of Japanese *on*- and *kun*-readings or some of the examples from the *Fangyan*. Standardization of the script with phonograms eventually made it much easier for the script to influence spoken language.

We must not think of the received Chinese phonological tradition as descended directly from the high-register inventory of readings, which are abstract and not whole syllables. Rather, the laxity of the writing system's phonological structure enabled the development of the received tradition. Then where did the Han exegetic tradition and its elaborate medieval systematization come from? Presumably, they were the reconstruction of an imagined earlier reading tradition — perhaps influenced by a reaction against or even a hyper-correct Xianbei 鮮卑 vision of such an imagined tradition — but not the faithful preservation of that tradition. The spread of a standard writing system with built-in phonology implies the spread of that phonology, as the basis of a high diglossic register shared by all. It is paradoxical that the representation of phonology may be essentially artificial and even strange to most speakers, but if it is shared by all, it will eventually influence all spoken languages. Written forms, especially those that embody explicit phonological relationships, tend to trickle down into lower-register language. Though not the ancestor of the spoken languages, over the long run the high-register form has the ability to encroach constantly on the low because it embodies a stable set of phonological relationships (though not absolute phonetic detail) that can constantly affect spoken language. And so, whatever the situation in earlier times, diglossia must ultimately have narrowed as a result of the normalization of written Chinese and the increase of phonograms in the Qin-Han.

How do script and speech interact? No one disputes that the Chinese script masks both the phonological detail and the variation of spoken language. A logographic script tends to conceal detail and variation by ignoring or lexicalizing it. Evidently, there is no means of observing or reconstituting colloquial language of the Han, but unexpected variation persists even now in China (although it is rarely remarked upon), as shown in one or two examples from modern times.

62. David Prager Branner, "Common Chinese and Early Chinese Morphology," *Journal of the American Oriental Society*, 122.4 (2002): 706–21, 717–19.

Consider a major regional language that is virtually never committed to writing in the normal course of things. In spoken Taiwanese, linguistically a variety of the large and far-flung Minnan dialect group, a number of common morphemes belong to upper-register tone-categories and have aspiration where it is not expected on historical or comparative grounds. The examples in table 3.32 are in the Church romanization system widely used in Taiwan.⁶³

The uncanonical aspiration in these words may be a relic of some systematic sound change, or some lost phonological or morphological feature, or possibly a reduction of homophone pressure by morphemes moving to less densely occupied phonological spaces. Those are precisely the sorts of things one could also propose to explain the diversity of word families in early Chinese or the irregularity of correspondence sets in Sino-Tibetan, explanations that would apply to lower-register language masked by a logographic script.

There are many more such examples among the character readings in Campbell's dictionary, but table 3.32 illustrates ten that survive in speech, as documented in the descriptive lexicons of Russell Sprinkle and Bernard L. M. Embree. All of these are attested in the speech of older people now living, and some (such as *khɔ̃*, *khia*, and *thióng*) are still the usual ways to say these things, even for young speakers. Others have been replaced in common usage by the expected forms — *piau-chún* replacing *phiau-chún*, *kui-tō* replacing *khui-tō*, and so on — which more closely match Mandarin phonology (though not phonetics). It is as if a group of Mandarin speakers said *qūliú* for *jūliú* 拘留, *chǎnkai* for *zhǎnkai* 展開, and so forth. If those people came to feel pressure to adjust their pronunciation to match everyone else's, *qūliú* and *chǎnkai* would soon disappear. Phonological

63. Campbell, *A Dictionary of the Amoy Vernacular*; Russell Sprinkle et al., *Amoy-English Dictionary* (Taichung: Maryknoll Fathers, 1976); Bernard L. M. Embree et al., *A Dictionary of Southern Min* (Taipei: Taipei Language Institute, 1984). I have chosen aspirated stops in upper-register tones because they seem to be special to Southern Min. Distinctive patterns of aspiration in lower-register tones and in syllables with affricate sibilant initials are much more widely distributed in Min and Hakka; for discussion and references to Jerry Norman's seminal works, see Branner, *Problems in Comparative Chinese Dialectology*, 63–67, 109–13. For correspondence sets illustrating the regularity of the rime of *khia* 奇, see David Prager Branner, "A Gutyan Jongbao Dialect Notebook," *The Yuen Ren Society Treasury of Chinese Dialect Data I* (March 1995): 243–338. All examples given in this table happen to be in the *píng* and *shǎng* tones; *rù*-tone examples are also known, but for some reason *qùshēng* examples are vanishingly rare. I have intentionally supplied examples with initials from the three main places of articulation.

representation in the Chinese script is too rough to distinguish aspirated and unaspirated initials, and there would be nothing but oral tradition to preserve them.

Table 3.32 Rogue aspiration in upper-register Taiwanese words

Medieval reading	Taiwanese			
	Expected reading	Attested character reading	Example in a living word	Sources
標 {pau _{3y} }	* <i>piau</i>	<i>phiau</i>	<i>phiau-chún</i> 標準 'standard'	Campbell 1913:578; Sprinkle et al. 1976:617; Embree et al. 1984:215
奔 {pen ₁ }	* <i>pun</i>	<i>phun</i>	<i>phun-cháu</i> 奔走 'to hasten, hurry'	Campbell 1913:578; Sprinkle et al. 1976:628; Embree et al. 1984:218
波 {pe ₁ }	* <i>po</i>	<i>pho</i>	<i>pho-lōng</i> 波浪 'wave'	Campbell 1913:580; Sprinkle et al. 1976:620; Embree et al. 1984:216
展 {tranQ _{3b} }	* <i>tián</i>	<i>thián</i>	<i>thián=khui</i> 展開 'to unroll, open up, develop'	Campbell 1913:1067; Sprinkle et al. 1976:829; Embree et al. 1984:283; no sandhi on first syllable
冢 {trungQ _{3c} }	* <i>tióng</i>	<i>thióng</i>	<i>thióng</i> 冢 'tomb, grave'	Campbell 1913:753; Sprinkle et al. 1976:833; Embree et al. 1984:285
刁 {tau ₄ }	* <i>tiau</i>	<i>thiau</i>	<i>thiau-bân</i> 刁蠻 'clever but obstinate'	Campbell 1913:751; Embree et al. 1984:284, second syllable miswritten 難
奇 {ki _{3bx} }	* <i>kia</i>	<i>khia</i>	<i>khia-sò</i> 奇數 'odd number'	Campbell 1913:388; Sprinkle et al. 1976:399; Embree et al. 1984:146
拘 {kuo _{3c} }	* <i>ku</i>	<i>khu</i>	<i>khu-liú</i> 拘留 'to detain'	Campbell 1913:401; Sprinkle et al. 1976:419; Embree et al. 1984:161
軌 {kwiQ _{3cx} }	* <i>kuí</i>	<i>khuí</i>	<i>khuí-tō</i> 軌道 'track, path'	Campbell 1913:413; Sprinkle et al. 1976:421; Embree et al. 1984:162
箍 {kuo ₁ }	* <i>ko</i>	<i>khō</i>	<i>khō-tháng</i> 箍桶 'to coop a barrel'	Campbell 1913:401; Embree et al. 1984:128; Sprinkle et al. 1976:409

So much for an unwritten language. But in a spoken language that has a normal written form, it is natural for even highly colloquial expressions to appear in print. Consider Northern Chinese, the regional language on

which standard Mandarin is based, promoted as a vehicle for colloquial writing since the May Fourth movement. Colloquial Northern Chinese furnishes plenty of examples of irregular tone change in recent usage, many of which, unlike unexpected aspiration in Taiwanese, have made their way into print (table 3.33).⁶⁴

The point is that not only does the tone of one morpheme vary, but the written form varies with it. The logographic script can be made to express variation by lexicalizing it. In other words, literate speakers are interpreting competing pronunciations as having different literal meanings and are varying the way they are written accordingly, reinforcing those competing pronunciations by formalizing them in print.⁶⁵ Because Mandarin became closely linked to the written language during the twentieth century, and because during that time writers have been swept by an ideology prescribing lower-register writing, it is not surprising to see the script mirroring the vernacular differences in pronunciation and understanding.

The advantage of expanding the inventory of phonograms in the Qin-Han was not just that a normalizing written language became semantically more stable and more flexible for use by colonized people because it was divorced from phonetics. A larger inventory also made possible a much closer relationship between high- and low-register language than had existed before, by virtue of the phonology that was now

64. Not all such changes make their way into print. In mid-twentieth-century Beijing, the colloquial word for “harelip” (both the person and the condition) was *tùchún*; the standard Chinese word was *tùchún* 兔唇. For the second syllable of the colloquial word to be in the third tone violates the rules of regular correspondence, and I suggested to my informant that it was the intrusion of *chǔn* 蠢 ‘stupid’, to add pejorative sense. (Pejoration is common in regional words for physical handicaps: in Taiwanese, for instance, to be unable to speak is expressed as to be *é-káu* 啞狗 ‘mute-dog’; “deaf” is *chhàu hī lāng* 臭耳聾 ‘stinky deaf’.) She insisted that it was an irregular change of tone, without any negative sense. But 蠢 is the only character at all common enough to represent the syllable *chūn* in print, and indeed I can find no evidence at all of this word ever having been lexicalized; evidently there has been no temptation to write 兔蠢兒. For the informant’s background, please see David Prager Branner, “The Linguistic Ideas of Edward Harper Parker,” *Journal of the American Oriental Society* 119.1 (1999): 12–34.

65. All of these examples can be found in current or recent use on the Internet (October 2008) as well as in Chinese dictionaries from the 1980s onward. *Niù píqì* 拗脾氣 is the only form I have not found in a dictionary. *Shùn kǒu liù*, with 溜 read *liù*, is now occasionally seen as 順口六 “‘six’ easily out of the mouth’.

Table 3.33 Lexicalized but irregular tone change in Northern Chinese lexicon

Standard form	... seems to have this literal meaning	Also pronounced and written	... seems to have this literal meaning	... and actually has this idiomatic meaning
<i>niú píqì</i> 牛脾氣	the temper of an ox	<i>niù píqì</i> 拗脾氣	contrary temper	stubbornness; (as verb) to be stubborn
<i>wūyǎn jī</i> 烏眼雞	[to give someone the] “black-eyed chicken” look	<i>wǔyǎn jī</i> 五眼雞	[to give someone the] “five-eyed chicken” look	to look at angrily; (as noun) sworn enemies
<i>jiāoguǒ(r)</i> 嚼裹(兒)	chew-bundle	<i>jiāoguǒ(r)</i> 澆裹(兒), <i>jiāoguǒ(r)</i> 繳裹(兒)	plant-watering bundle, pay-bundle	living expenses
<i>shuāi liǎnzi</i> 甩臉子	to throw off one’s facial expression	<i>shuāi liǎnzi</i> 摔臉子	to throw down one’s facial expression	to show one’s displeasure without regard for how it makes one look
<i>qiào biànzi</i> 翹辮子	to have one’s braid stick up straight	<i>qiāo biànzi</i> 蹺辮子	to lift one’s braid	to die (informal)
<i>tōng lóuzi</i> 捅樓子, 捅妻子	to poke the building, to poke and create a problem	<i>tōng lòuzi</i> 捅漏子	to poke the funnel (? or ‘... a leak’?)	to start an argument or make a mess of things
<i>wōli fān</i> 窩裡反	rebellion in the nest	<i>wōli fān</i> 窩裡翻	upset in the nest	infighting
<i>dǒu wēifēng</i> 抖威風	to shake one’s awesome aura around	<i>dōu wēifēng</i> 兜威風	to move one’s awesome aura around	to throw one’s weight around
<i>xiànshí bào</i> 現世報	retribution in one’s present life	<i>xiànshí bào</i> 現時報	retribution at the present moment	short-term karmic retribution
<i>shùnkǒu liū</i> 順口溜 (-liù also attested)	[thing that] slides easily out of the mouth	<i>shùnkǒu liú</i> 順口流	[thing that] flows easily out of the mouth	doggerel or other rhymed saying
<i>dà sāshǒu</i> 大撒手	to use one’s big scattering hand	<i>dǎ sāshǒu</i> 打撒手	to do the scattering-hand thing	to give up and let the situation take care of itself
<i>méi diānsān</i> 沒掂三	to lack “three on tiptoe” (?)	<i>méi diànsān</i> 沒店三	to lack “three in the store” (?)	to lack careful forethought
<i>dǎ bǎopiào</i> 打包票	to do the contract-ticket thing	<i>dǎ bǎopiào</i> 打保票	to do the guaranteed-ticket thing	to give a guarantee
<i>yìn bǎzi</i> 印把子	seal-handle	<i>yìn bǎzi</i> 印靶子	seal-target	official seal

systematic and explicit in the script. By being written down in Chinese and having a reading associated with it, every word could actually take on a high-register form, even if it had originally been low register. With the process of normalization, every character whose phonological structure was transparent could influence the spoken language. In the long run, the script became a force for phonological unity among the spoken Han languages and for reduced diglossic distance between high and low registers.

Although the gap between Chinese writing and speech seems huge to us today, the two are probably far closer than they were before the script was standardized. Literacy in antiquity surely required a larger inventory of graphs than literacy today does, because in antiquity a far higher proportion of words were monosyllabic and had to be distinguished. So far from being an obstacle to literacy because of its low level of phonetic explicitness, Chinese writing made widespread literacy possible because it adopted a very lax representation of phonological relationships, unencumbered by phonetic detail. Viewed in the large, the history of Chinese phonology has been a long contest between the defectiveness of the *xiéshēng*-based script and the excessive distinctions of the *Qieyun* tradition. The *Qieyun* tradition is the product of an ideology espousing fine phonetic precision, and that ideology has also profoundly affected modern reconstructional thinking about both the middle and early periods of the language. But the creators of the early script do not seem to have believed in phonetic precision. In the absence of a truly standardized spoken language, underspecification has done more to make literacy possible than has the phonological tradition of the medieval period, to which honor is usually shown although it is rarely adhered to in actual usage.

In sum, the defectiveness of the early script would have aided social forces in spreading literacy, while phoneticism (even imperfect phoneticism) would have aided cognitive factors. With the beginning of true standardization of spoken Chinese in the past century and the falling away of characters that are polyphonic or character-readings for graphs with irregular phonophores, we may finally be approaching a time when a phonetic script is feasible for the majority of Chinese speakers.

Supplement: Initial Types for *Xiéshēng* Groups

Since the time of Joseph Edkins (1823–1905), the study of the early period of Chinese phonology by Westerners has tended toward reconstruction —

the recovery of speech sounds either in actual phonetics or in some more abstract form. Chinese reconstructions are usually built on a number of kinds of evidence, but for the pre-Han period, the primary one is systematic medieval phonology, which provides a clear starting point for every one of the tens of thousands of characters known in the seventh through eleventh centuries. Medieval phonology is combined with information about rhyming and *xiéshēng* 諧聲 character structure and many lesser details, to manufacture what we know as Old Chinese (OC) reconstructions. Recent unpublished work by Laurent Sagart and William Baxter has also incorporated modern ideas about lost inflectional morphology, a two-way (“A/B”) distinction in initial-type, an additional type of initial category, and other ideas. In this volume, chapter 1, by David W. Pankenier, and chapter 2, by William G. Boltz, use a reconstruction of Gassmann and Behr and similar to Sagart and Baxter’s.

This chapter, however, presents the phonology embodied in the Chinese script as something intrinsically different from the diverse oral languages of the early Chinese world. This “orthographic phonology” of Chinese may be represented using a system of rough transcription for characters. It is not a reconstruction of any sort: it is not intended to represent actual speech sounds of any time or place in history, however abstractly. Rather, it is meant to encapsulate compactly what is known about significant categories of sound, encompassing medieval phonology, rhyming, and *xiéshēng* relationships in the script. For representing the phonological freight of the Chinese script in a phonetically suggestive form, it seems much better than a reconstruction because it is less explicit in matters with which phonetic realism is not to be trusted.

In this system, only the initial and “rime” (syllable-final) categories are represented, as those are the elements having the greatest bearing on the *xiéshēng* series. Tone categories (beyond those built into the rime categories themselves), reconstructed morphology, medial semivowels, and the A/B distinction are all excluded.

Rimes are represented directly by those proposed in Baxter’s 1992 reconstruction, which remain largely unchanged in his and Sagart’s latest work. Initials are represented as ten large, catch-all categories, each built around a single core consonant that appears in Baxter’s reconstruction and is symbolized by a single Roman letter consonant. OC reconstructed initials feature a significant number of initial clusters (something almost always lacking in reconstructions of medieval phonology); the single “core” consonant of such a cluster is always the one isolated as representative.

The ten initial categories symbolize the actual units of *xiéshēng* 諧聲 contact; finer distinctions among medieval or reconstructed early Chinese initial categories are more precise than is needed for an accurate representation of *xiéshēng* contact.

Table 3.34 shows the ten initial categories and the corresponding medieval initial categories they encompass. For details of the reasoning underlying the reconstruction of various initial clusters, consult the appropriate sections of Baxter 1992, which are listed in the right-most column of the tables. Corresponding medieval categories are transcribed following Branner 2006. The notation “(合)” is used to include contrastively *hékǒu* 合口 syllables and those syllables that *bùfēn kāihé* 不分开合, for example, those with rounded vowels in the rime, in *shè* 攝 such as *tōng* 通, *liú* 流, and so on, for which *kāihé* 開合 is not contrastive.

By way of example, category {P} encompasses six distinct reconstructed OC initials: **p-*, **ph-*, **b-*, **Np-*, **b-r-*, and **sp-*. What the six have in common is that a labial stop *p* or *b* is thought to be the principal element of each in the OC materials. These six reconstructed initials correspond to medieval initial categories 幫 *p*, 滂 {*ph*}, 並 {*b*}, 明 {*m*}, 來 {*l*}, and 生 {*sr*} in certain environments; note that the medieval initials are not necessarily labial in articulation.

Category {M} encompasses five OC initials, with *m* or *w* as the principal element of each, and corresponding to six medieval initials.

Category {T} encompasses five OC initials, with *t* or *d* the principal element. Category {N} encompasses five initials, all based on *n*; category {L} encompasses eight initials, all based on *l* or *r*; and category {S} encompasses seven initials, all based on *s* or *z*.

Category {K} encompasses ten initials; category {KW}, six; category {NG}, three; and category {NGW}, two. It is well established that some sort of rounding contrast is meaningful only among initials of the velar and laryngeal types in Old Chinese.

Table 3.34 Correspondence of Baxter OC to medieval Chinese initials

Initial groups	Medieval initial category, after Branner 2006			Baxter's reconstructed "OC" (early Chinese) initial categories		
	一四等	三等		二三等 if different from regular 三等	Baxter 1992	Where cited in Baxter 1992
		back vowel or *-rj- in Baxter	front vowel in Baxter			
P	幫 p	幫 p	—	*p	6.1.1, 6.1.3.2 (*pr)	
	滂 ph	滂 ph	—	*ph	6.1.1	
	並 b	並 b	—	*b	6.1.1, 6.1.3.2 (*br)	
	明 m	明 m	—	*Np	6.2.2	
	來 l	來 l	—	*b-r	6.1.3.2	
	—	生 sr	—	*sp	6.2.3.2	
M	明 m	明 m	—	*m	6.1.1	
	心 s	心 s	—	*sm	6.2.3.1	
T	端 t	章 tsy	知 tr	*t	6.1.2.1, 2, 3, 6.1.3.1	
	透 th	昌 tshy	徹 thr	*th	6.1.2.1, 2, 3	
	定 d	禪 dzy	澄 dr	*d	6.1.2.1, 2, 3	
	泥 n	泥 n	—	*Nt	6.2.2	
	—	書 sy	—	*st	6.2.3.2	
N	泥 n	日 ny	(娘>)泥 n	*n	6.1.2.1, 2, 3	
	透 th	書 sy	徹 thr	*hn	6.1.2.1, 2, 3	
	清 tsh	心 s	清 tsh	*sn	6.2.3.1	

Table 3.34 (Continued)

L	定 d	羊 y	—	*l	6.1.3.1	
	—	邪 z	—	*zl	6.1.3.1	
	—	船 zy	—	*L	6.1.3.1	
	透 th	書 sy	—	*hl	6.1.3.1	
	心 s	心 s	—	*sl	6.2.3.1, 2 (LFK's *st-)	
	—	羊 y	—	*r	6.1.3.2	
	透 th	徹 thr	—	*hr	6.1.3.2	
—	生 sr	清 tsh	*sr	6.2.3.1 (*srj > sr)		
S	精 ts	精 ts	莊 tsr	*ts	6.1.4	
	清 tsh	清 tsh	初 tshr	*tsh	6.1.4	
	清 tsh	生 sr	—	*sr	6.1.4	
	心 s	心 s	—	*s	6.1.4	
	從 dz	從 dz	崇 dzr	*dz	6.1.4	
	—	邪 z	俟(=崇 dzr)	*z	6.1.4	
	—	邪 z	—	*zl	6.1.3.1	
K	見 k	見 k	章 tsy	—	*k	6.1.5.1, 2
	溪 kh	溪 kh	昌 tshy	—	*kh	6.1.5.1, 2; 6.1.3.2 (*kr)
	匣 gh	群 g	禪 dzy	匣 gh	*g	6.1.5.1, 2; 6.1.3.2 (*gr)
	疑 ng	疑 ng	—	—	*Nk	6.2.2
	—	生 sr	—	—	*sk	6.2.3.2
	影 [0]	影 [0]	—	—	*?	6.1.5.1
	曉 h	曉 h	書 sy	—	*x	6.1.5.1, 2
	來 l	來 l	—	—	*g-r	6.1.3.2
	—	羊 y	—	—	*j	6.1.3.3
—	書 sy	—	—	*hj	6.1.3.3	

Table 3.34 (Continued)

KW	見(合) kw	見(合) kw		—	*k ^w	6.1.6
	溪(合) khw	溪(合) khw		—	*kh ^w	6.1.6
	匣(合) ghw	匣(合) ghw		—	*g ^w	6.1.6
	—	禪 dzy		—	*sg ^w	6.2.3.2
	疑(合) ngw	疑(合) ngw		—	*Nk ^w	6.2.2 (implied)
	影(合) w	影(合) w		—	*ʔ ^w	6.1.6
	匣(合) ghw	匣(合) ghw	羊(合) yw	—	*w	6.1.6
	曉(合) hw	曉(合) hw		—	*hw	6.1.6
	心(合) sw	心(合) sw		—	*sw	6.2.3.1
NG	疑 ng	疑 ng	日 ny	—	*ng	6.1.5.1, 2
	心 s	心 s		—	*sng	6.2.3.1
	曉 h	曉 h	書 sy	—	*hng	6.1.5.1, 2
NGW	疑(合) ngw	疑(合) ngw		—	*ng ^w	6.1.6
	曉(合) hw	曉(合) hw		—	*hng ^w	6.1.6