BRILL

BULLETIN OF CHINESE LINGUISTICS 8 (2015) 48-60

BULLETIN
of CHINESE
LINGUISTICS
brill.com/bcl

Proposal for a Transcription of Chinese Characters in the Study of Early Chinese Language and Literature

Nathan W. Hill
SOAS, University of London
nh36@soas.ac.uk

Abstract

This paper outlines the pitfalls of the current anachronistic practice of transcribing early Chinese documents by identifying each character with a *kǎishū* 楷書 equivalent. In its place, I suggest a way of transliterating characters directly, by rendering into roman letters the phonetic and semantic information encoded by a character. (This article is in English.)

Keywords

transcription – Chinese

1 Current Practice

The standard approach to deciphering pre-Qin Chinese documents is to identify each pre-Qin character with a *kǎishū* 楷書 character, to take each *kǎishū* character as representing the morpheme that it represents in Pǔtōnghuà 普通話, and finally to read the result as if it were standard Classical Chinese. False assumptions underpin this methodology at every step; the procedure assumes a direct linear progression in both script and language from the Shāng 商 dynasty until today. W. S. Coblin has demonstrated in various publications that the linguistic side of this assumption is false (e.g. 2001, 2007: 69–103); Imre Galambos shows that the paleographic assumption is also false (2006: 146–150).

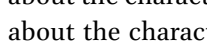
Transcription can serve three purposes: (1) to typographically present words originally written in one script system using text written in another script system,[1] (2) to draw attention to the structure of the script system in a way that is more explicit than the native script itself reveals,[2] and (3) to regularize

1 The typographical representation of Chinese words in Roman script is easily satisfied. As an index referring to a certain philosopher 'Confucius' serves perfectly well. This series of Roman letters has a very tenuous relationship with the name this philosopher would have been called in his own life, but by convention it is what he is called in English. No accuracy or fidelity is gained by calling him 'Kongzi' instead. This series of letters accurately reflects what this man is called according to one romanization system for Pǔtōnghuà, but neither the letters Kongzi nor the pronunciation they indicate would be any more meaningful in the philosopher's ears than the equally arbitrary and more conventional 'Confucius'.

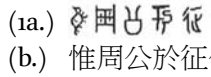
2 For example, in hand written Tibetan the letter ད, the letter ད, and the punctuation mark *tsheg* are difficult to distinguish; transliterating them respectively with a 'd', 'ḋ', and a space disambiguates them clearly. In a similar fashion, transcriptions

 | DOI 10.1163/2405478X-00801003

the idiosyncrasies of an individual text, i.e. to reflect the specimen of *langue* implied by a text rather than the *parole* that it instantiates. These three goals are often in tension and sometimes the most sensible course of action is to transcribe a text twice. For example, the narrow Hittite transcription ^{URU}KÙBBAR-ša-aš renders the cuneiform script into Roman letters and draws attention to the structure of the original script system by distinguishing respectively a sumerogram determiner, a sumerogram, and two Hittite syllabograms, but this narrow transcription fails to reflect the Hittite language encoded by this writing; for the third goal as mentioned, a broad transcription of *Ḫattušaš* is more appropriate (cf. Fortson 2010: 178).

The current practice of transcribing early Chinese texts fails to achieve all three of these goals. The string of symbols 惟周公于征伐東 (cf. Shaughnessy 1991: 68–73) is *a priori* no more intelligible than  ; both are equally non-Roman. The current practice fails to reveal the structure of the script system more explicitly than the original. Although a *kǎishū* 楷書 equivalent may help to disambiguate similar looking characters, it also risks collapsing together differently written characters or artificially imposing differences not present in the original text. Writing  as 東 or 束 reveals nothing about the character's relationship to the other characters in the orthographic system, reveals nothing about the character's probable pronunciation at the time of writing, and reveals nothing about the character's meaning.[3]

The road toward more accurate *kǎishū* renderings is never ending. Shaughnessy opts for 1c, presumably thinking that option 1b would be too interpretive, but he rejects option 1d as unnecessarily strict.

(1a.) 
(b.) 惟周公於征伐東
(c.) 隹(i.e. 惟 or 維)周公  (i.e于)征伐東
(d.) 隹(i.e. 惟 or 維)  (i.e.周)公(i.e于)征伐束(i.e.東)

Any of these three transcriptions (1b.-1d.) may be helpful for some readers at some time, but none of them can be 'correct,' because they are all anachronistic. If the purpose of such a transcription is it to capture the graphic form of the original, then version (a), a graphically standardized representation of the bronze characters themselves, is already a fully satisfactory transcription; this transcription is still an abstraction of a reproduction of the original rubbing.

Galambos points to the structure of the Chu character  (equivalent to 仁) as an example where a correct analysis in terms of *kǎishū* transcription is not possible (2006: 75–76). The Chu character is ultimately built on the radical  (equivalent to 人) but also contains in its graphic structure the characters  (equivalent to 千) and (equivalent to 身).

1. (人)
2. (人 + 十 = 千)
3. (千 + 厶 = 身)
4. (身 + 心 = 仁)

---

of Greek into Roman letters well achieves the goal of revealing structure inherent but inexplicit in the original. The short vowels ε and ο do not look like their long counterparts η and ω, but the transcription of the former as 'e' and 'o' and the latter as 'ē' and 'ō' makes their systematic relationship explicit with the use of a macron to represent vowel length. Similarly writing 'kh' and 'ph' makes the relationship of χ and φ to κ and π much clearer than the forms of the symbols themselves.

3 The character 東 no more directly indicates the meaning 'east' than itself.

It is impossible to construct a *kǎishū* transcription that correctly captures these relationships. As Galambos points out: "if I were to transcribe  as either  or  , I would lose the structural information inherent in the original character and obscure the connection between the components and the entire character" (2006: 75–76). If a researcher wants to capture the structure of the original character, reproducing as is, offers the most accurate representation.

Current scholarly practice treats early Chinese texts not as documents produced in the language and orthography of their own day, but instead as a mischievous code which the ancients perversely used to represent the Chinese language of our own time. In so doing, the current practice abandons any attempt to represent the underlying *langue* recorded in early Chinese documents. If one were to approach Old English with an analogous technique, the following lines from the end of Beowulf (2a.) might be represented as (2b.); the morphemes and the phonemes are those of modern English, but the words and word formation are those of Old English. The modern English edition is neither a faithful reflection of the original nor a useful guide to understanding its meaning.

(2a.) cwǣdon þæt hē wǣre wyruldcyninga
manna mildust ond monðwǣrust
lēodum līðost ond lofgeornost
(2b.) (They) quoth that he were (of) world-king(s)
Man mildest and monthwerest
(among) lede lithest and lofe-yearnest.

The traditional approach to transcribing early Chinese characters is intellectually incoherent and is not executable in a rigorous fashion. This approach is an impediment to research on the history of the Chinese language. Sinologists are increasingly cognizant of the drawbacks of this traditional approach (Takashima 2000, Richter 2003, Xing 2005), but their proposed solutions fall short of abandoning the search for *kǎishū* equivalents altogether.

2 Proposed Transcription

If "bronze inscriptions are nothing more and certainly nothing less than transcriptions of the language of their time" (Shaughnessy 1991: 63), then a transcription of a bronze inscription should attempt to reflect such information regarding pronunciation present in the text itself, as it was available to the original reader. Such a transcription would rigorously separate the phonetic and semantic information that the script presents, and present the non-sinologist reader with all of this information in an easily comprehensible series of Roman letters. The methodological and intellectual pitfalls of transcribing pre-Qin characters into *kǎishū* characters merits the abandonment of this practice by all those who engage in the study of early Chinese language and literature. A Roman transcription of Chinese characters will both sharpen the scholarly practice of experts and render their findings more useable and more meaningful to those in associated disciplines, such as early history and historical linguistics—those who are ignorant of *kǎishū* characters.

For clarity of presentation I first exhibit my proposed transcription system using *kǎishū* characters and then present some examples of its employment for pre-Qin characters. A well-designed transcription system can be used for characters in any period, since the phonetic and semantic information in the script of any two periods present itself to its intended audience in distinct

ways.[4] Consequently, it would be a mistake to transliterate oracle bone inscriptions in the same way as bronze inscriptions. A good transcription system will not impose an artificial uniformity on Chinese of all periods, but will instead sharpen the understanding of differences among periods and texts. Nonetheless, the principles behind a Chinese transcription system should be the same for all periods.

The vast majority of Chinese characters from all periods are phonetic compounds (*xíngshēngzì* 形聲字). The phonetic element of a character does not directly indicate the pronunciation of the morpheme which that character represents, but rather indicates the pronunciation that is common to all of the characters within the same *xiéshēng* series; thus, the presence of the phonetic element 刀 in the characters 召 and 超 does nothing to indicate that the first has a voiced initial (drj-) and the second a voiceless aspirate initial (trh-). Instead, the element 刀 indicates that that the initial of the readings that characters 召 and 超 indicate is some kind of dental, and that these readings have the rime –aw. A transcription of the phonetic determiner of a character should aim to represent the phonetic information that is implied by the character's presence in a particular *xiéshēng* series. In general this will mean that voicing, morphological prefixes, segments giving rise to tones (i.e. final -ʔ and -s), and the A/B distinctions are ignored.

By way of example, consider the *xiéshēng* series built on the character 刀. Each character is followed by its Middle Chinese and Old Chinese reading in the system of Baxter & Sagart (2014).

刀 taw < *C.tˁaw
召 drjewH < *[d]raw-s
超 trhjew < *tʰr[a]w
昭 tsyew < *taw
炤 tsyewH < *taw-s
照 tsyewH < *taw-s
沼 tsyewX < *tawʔ
詔 tsyewH < *taw-s

The basic character from which all other members of a *xiéshēng* series are built can be written in capital letters, thus 刀 TAW. The semantic components shall be written in superscript with the first few letters of a Latin word that indicates the appropriate meaning (cf. Table 1). The following conventions allow for the physical relationship of the phonetic and the semantic components to be reflected in the transcription.

A.B means A is to the left of B.
AxB means A contains B
A:B means A is on top of B

Following the conventions outlined so far, the *xiéshēng* series built on 刀 may be transliterated as follows:

刀 TAW, 召 taw$^{:os}$ 超 $^{cur.}$taw$^{:os}$, 昭 $^{sol.}$taw$^{:os}$, 炤 $^{ign.}$taw$^{:os}$, 沼 $^{aqu.}$taw$^{:os}$, 詔 $^{dic.}$taw$^{:os}$.

4 To the extent that readers still read the phonetic and semantic components of characters as independent units, such a system would also be useful for Sinitic languages spoken today. However, altogether phonetics clearly aids in the process of memorizing characters, the reading experience of *kǎishū* by readers of Chinese today is very different than at the time of the script's codification.

TABLE 1 *A selection of semantic determiners*

Verbs and Adjectives							
行	eo	攴	fer(io)	走	cur(ro)	欠	dehab(eo)
言	dic(o)	見	vid(eo)	卩	sig(no)	辵	amb(ulo)
广	lax(us)	食	ed(o)	大	magn(us)	疒	inf(irmus)
舛	obs(isto)	白	alb(us)	黑	nig(er)	止	ces(so)
高	alt(us)	立	sto				
Natural World							
日	sol	土	ter(ra)	木	arb(or)	竹	bamb(us)
月	luna	禾	gran(um)	艸	herb(a)	羽	plum(a)
火	ign(is)	水	aq(ua)	雨	pluv(ia)	貝	conch(a)
山	mon(s)	金	met(allum)	阜	tum(ulus)	玉	gem(ma)
角	corn(u)	穴	cav(us)	示	spir(itus)	石	lap(is)
Human World							
人	hom(o)	女	fem(ina)	瓦	teg(ula)	戶	ost(ium)
門	por(ta)	邑	urb(s)	网	nas(sa)	糸	ser(icum)
宀	tect(um)	攵	vex(illum)	皿	vas(um)	厂	caut(es)
舟	nav(is)	肉	carn(is)	口	cla(usum)	刀	cul(ter)
巾	lint(eum)	車	vehi(culum)	田	ag(er)	敕	iuss(us)
酉	vin(um)						
Body parts							
手	man(us)	口	os	首	cap(ut)	目	ocu(lus)
耳	aur(is)	足	pes	心	cor	頁	fac(ies)
髟	cri(nis)	齒	den(s)				
Animals							
犬	can(is)	魚	pis(cis)	牛	bos	羊	ov(is)
虫	serp(ens)	馬	equ(us)	鹿	cerv(us)	鳥	av(is)
單	cic(ada)						
Exclamations							
兮	io						

Some might propose that the series is better analysed as built on 召, in which 召 itself will be transliterated as TAW and the semantic determiner 'os' can be omitted from the transcription of the remaining characters. The disadvantage of such a tack is that it splits off 刀 as unrelated to this series. In such cases a particular character may either be understood to have more than one semantic determiner, or it may be understood to be formed to a new phonetic: the character 照 receives the transcription $^{sol.}$taw$^{:os:ign}$, if 刀 is phonetic, $^{sol.}$taw$^{:ign}$, if 召 is phonetic, or taw$^{:ign}$, if 昭 is phonetic.

Latin has several advantages over other languages for the transcription of semantic determiners. The use of *pīnyīn* (e.g. 召 taw$^{:k\check{o}u}$) risks implying that Pǔtōnghuà enjoys some special epistemological relationship with earlier forms of Chinese, whereas counteracting this misconception is one of the urgent reasons to develop a Roman transcription system. The use of *kǎishū* characters to represent semantic determiners (e.g. 召 taw$^{:口}$) abandons the enterprise of Roman transcription altogether. Using *pīnyīn* and *kǎishū* characters to transcribe semantic determiners would tempt the reader to believe that the determiner indicates a word, rather than serving a disambiguating function. The use of English (e.g. 召 taw$^{:mouth}$) would look ridiculous in publications not written in English and risks lulling the Anglophone reader into believing that the transcription of a semantic determiner directly points to an idea rather than serving as a label for a graphic element. For these reasons, Latin is the better choice. In order to save space and avoid questions proper to Latin grammar (such as grammatical gender), I propose to use abbreviated versions of Latin words that omit grammatical endings. I generally seek out a three letter name for the semantic determiner, but in some cases, in order for the Latin short form to look more familiar on the basis of words in European languages or in order to maintain distinctions among the Latin names, at times I employ four letter, or even longer, versions. I cannot emphasize enough that the particularities of the system proposed here are of very little consequence. If other writers prefer to use entire inflected Latin words, or prefer to use English or other languages in the representation of the semantic determiners, so be it. If other writers began to transcribe pre-Qin characters in a way that differentiated semantic and phonetic components in any way, this would be a huge methodological step forward—regardless of the details of any particular system.

The *xiéshēng* series built on the character 或 exemplifies the power of the new system to explicitly represent the analysis of individual scholars. In order to make the transcription more transparent to non-specialists, I use the letter 'y' in preference to 'ə', and -u- rather than -^{w}-, as the sequences 'qu' and 'ku' looking less exotic than 'qw' and 'kw'. Again, if others prefer to write 'ə' and '^{w}' or 'w' so be it. Karlgren includes both 蜮 and 蟈 in the *xiéshēng* series built on the character 或. According to this analysis, both characters have the semantic 虫 'serp' and the phonetic determiner 或 QUYK; they would be transliterated respectively as 蜮 ^{serp}·quyk and 蟈 $^{serp.cla^{x}}$quyk. However, if one instead sees 蟈 as built on 國 in order to specify a velar initial, then one may write 國 as KUYK rather than $^{cla^{x}}$quyk and transliterate 蟈 as ^{serp}·kuyk rather than $^{serp.cla^{x}}$quyk. In this analysis, Karlgren's series 0929 is subdivided into two.

或 QUYK, 域 ^{ter}·quyk, 棫 ^{arb}·quyk, 閾 $^{nas:}$quyk, 閾 $^{por^{x}}$quyk, 淢 ^{aqu}·quyk, 馘 $^{cav:}$quyk, 緎 ^{ser}·quyk, 惑 quyk$^{:cor}$, 蜮 ^{cap}·quyk, 聝 ^{aur}·quyk

國 KUYK, 蟈 ^{serp}·kuyk, 膕 ^{carn}·kuyk

In a similar case, Karlgren includes 囷 in the series built on 袁, but whereas 囷 and those characters built from it have a reading with the vowel -e-, the character 袁 and those characters that are derived from it have readings with the vowel -a-. In this analysis Karlgren's series 0256 can be divided into two. Schuessler also follows this approach; he places 袁 in 25–15 (2009: 268) and 囷 in 23–11 (2009: 246–247).

袁 QUAN, 園 $^{cla^{x}}$quan, 轅 ^{vehi}·quan, 遠 $^{amb^{x}}$quan

囷 QUEN, 還 $^{amb^{x}}$quen, 環 ^{gem}·quen, 擐 ^{man}·quen, 圜 $^{cla^{x}}$quen, 儇 ^{hom}·quen, 翾 quen·^{plum}

When the *xiéshēng* series does not allow us to distinguish the vowel, then the ambiguity can be represented with 'v.' For example, the *xiéshēng* series derived from 門 is transliterated as 捫 man·mvn, 聞 mvn^{x}aur, 問 mvn^{x}os, and 閩 mvn^{x}serp, because the Middle Chinese reading of some characters points to Old Chinese 'ə' (e.g. 捫 *mwon* < *mˁən), whereas the Middle Chinese reading of other characters points to Old Chinese 'u' (e.g. 聞 *mjun* < *mun). If one fears that 'v' is unlikely for non-linguists to recognize as a vowel, another option would be to write a hyphen (e.g. 捫 man·m-n) for an ambiguous vowel.

If all of the morphemes written with a *xiéshēng* series are voiced, or have a particular tone, then this fact must be reflected in the transcription of the phonetic determiner. For example, all Middle Chinese readings of characters that use 口 as a phonetic element are aspirate velars of type A syllables (口 *k^{h}ˁoʔ, 叩 *k^{h}ˁoʔ, 扣 *k^{h}ˁoʔ-s, 釦 *k^{h}ˁoʔ). In such a case, the phonetic transcription must reflect both the aspiration and the 'type A' characteristic. Type A may be represented with a doubling of the initial consonant. This series is transliterated as follows:

口 KKHO, 叩 kkho·sig, 扣 man·kkho, 釦 met·kkho.[5]

The *xiéshēng* series built on the character 隹 gives a good idea of how complicated some cases can be. The Old Chinese reconstructions, according to Baxter & Sagart, group broadly into a reading TUJ (隹 *tur, 推 *t^{h}ˁuj, 椎 *k.druj, 催 *s-t^{h}ˁuj) and a reading QUI (帷 *ɢ^{w}rij, 惟 *ɢ^{w}ij, 維 *ɢ^{w}ij, 雖 *s-q^{w}ij), but the readings of the characters 稚 *lrəj-s and 崔 *dzˁuj do not match either. The graphic representation of these characters does not allow for a subdivision of the series, as was possible with 或 QUYK and 國 KUYK or 袁 QUAN and 囂 QUEN. Schuessler hypothesizes that the character 隹 originally was used to refer to two different words meaning 'bird,' both of which were onomatopoetic in origin (2009: 37).

The conventions so far illustrated require common sense in their implementation; it is precisely the fact that the system requires judgement that makes it useful. The system is a means by which different researchers can exhibit their judgements explicitly. The series built on 莫 presents an instance where the principle of finding a phonetic common denominator to all readings represented by characters in the series is not the correct approach. A few characters lack readings with the syllable final -k (模 *mˁa, 謨 *mˁa, 膜 *mˁa). Nonetheless, to incorporate this possibility into the transcription of the entire series, by transliterating MMA(K), for example, is not the right strategy. The use of this series for open syllables is a late development, derived from the simplification of -ks clusters to -H, in terms of analogy: 暮 : (*mˁaks >) *mˁaH :: 模 : *mˁa.[6] Consequently, it is right to transliterate the whole series with the phonetic MMAK.

莫 MMAK, 暮 mmak:sol, 募 mmak:cul, 墓 mmak:ter, 嫫 fem:mmak, 慕 mmak:cor, 蟆 serp·mmak, 寞 tect:mmak, 幕 mmak:lint, 漠 aq·mmak, 瘼 inf^{x}mmak, 嗼 os·mmak, 模 arb·mmak, 謨 dic·mmak, 膜 carn·mmak

5 Another type A series is built on the character 行: 行 GGAŃ, 荇 herb:ggań, 桁 arb·ggań, 珩 gem·ggań, 衡 ggań^{x}corn:magn, 蘅 herb:ggań^{x}corn:magn.

6 The character 冪 mek < *mˁek offers a further complication, implying that the series should be transliterated as MMVK rather than MMAK. This would also be a mistake. The explanation is that there is no obvious *xiéshēng* series to use for the sound [mˁek], so MMAK was pressed in to service for this purpose. The lack of a series with the phonetic MEK is also what led the character 幦 mek < *mˁek (08530) to be represented with the phonetic 辟 PEK.

The series based around 酉 conveniently exhibits disagreements among researchers. Marc Miyake (through personal communication) sees the entire series as sharing the same lateral phonetic. With such an understanding one may transliterate as follows:

酉 LU, 酒 aq·lu, 酋 lu, 蝤 serp·lu, 遒 ambˣlu, 緧 ser·lu, 猶 can·lu

In contrast, Axel Schuessler believes that 酉 indicates the pronunciation YU in some words and the pronunciation TSU in other words (2009: 177).

酉 YU, 廋 laxˣyu, 栖 arb·yu, 猶 can·yu, 猷 yu·can, 輶 vehi·yu, 蕕 herb:can·yu

酒 aq·tsu, 酋 tsu, 蝤 serp·tsu, 遒 ambˣtsu, 緧 ser·tsu, 鰍 pis·tsu, 蹤 pes·tsu

Finally, Baxter & Sagart posit thee phonetic values for 酉, namely RU, QU, and TSU.

酉 RU

猶 can·qu

酒 aq·tsu, 酋 tsu, 蝤 serp·tsu, 遒 ambˣtsu, 緧 ser·tsu, 鰍 pis·tsu, 蹤 pes·tsu

Unfortunately, for the majority of the characters which Schuessler analyzes as having the phonetic YU, Baxter & Sagart have not yet distinguished the readings QU and RU.

Now that the principles of the proposed transcriptions have been exhibited with *kǎishū* characters, some consideration can be given for phenomena confronted in earlier versions of the script. It is no surprize that the *kǎishū* script often obscures the phonetic links within a *xiéshēng* series. For example, it is not clear at inspection that 蚩 and 志 have 之 as their phonetic component; from the perspective of the *kǎishū* script, they do not. It would thus be entirely sensible in transcriptions of the *kǎishū* script to distinguish three series ty_1 (之芝), ty_2 (蚩), and ty_3 (志誌). However, if one considers the earlier forms of these characters, then all of these characters may be placed in the same series, as Karlgren does. The relationship among some of the characters is only clear with reference to earlier forms.

之() TY, 芝 herb:ty, 蚩() ty:serp , 志() ty:cor, 誌 dic·ty:cor

In another example, if one keeps in mind the oracle bone forms of the characters 丁 and 正 ( and respectively), then it is possible to see the characters built on both as part of the same *xiéshēng* series; this is the treatment that both Karlgren (1964[1957]: 220–221) and Schuessler adopt (2009: 137–138).[7]

丁() TEṅ, 頂 teṅ·fac, 汀 aq·teṅ, 町 ag·teṅ, 亭 alt:teṅ, 停 hom.alt:teṅ,
正 () teṅ:ces, 征 eo·teṅ:ces, 政 teṅ:ces.fer, 整 iuss:teṅ:ces, 証 dic·teṅ:ces, 鉦 met·teṅ:ces, 窺 cav:teṅ:ces.vid,
定 tect:teṅ:ces

7 Following the Indological tradition, I prefer to use 'ṅ' for the velar nasal, but others may prefer the 'ŋ' of the International Phonetic Alphabet or simply the 'ng' digraph so common in the orthography of European languages.

However, it is reasonable to speculate that already in the Zhou period the two characters were no longer obviously related. Consequently, to transcribe 正 as either TEṄ$_2$ or as teṅ$^{:ces}$ are both valid options.

The series built on 又 shows an alternation between the final -k and final -ʔ (e.g. 郁 *qʷək and 囿 *ɢʷək versus 友 *ɢʷəʔ, 右 *ɢʷəʔ, 賄 *qʷʰˁəʔ, etc.). In a case such as this, the transcription QUY (K) might be appropriate. However, if one hypothesizes that all *xiéshēng* series that single out the 上 *shǎng* tone also allow for velar finals, then QUYX is a superior transcription.

> 又() QUYX, 友() quyx$^{:man}$,[8] 右() quyx$^{:os}$, 佑 $^{hom\cdot}$quyx$^{:os}$, 祐 $^{spir\cdot}$quyx$^{:os}$, 有 quyx$^{:carn}$, 郁 quyx$^{:carn.urb}$, 囿 $^{cla^x}$quyx$^{:carn}$, 賄 $^{conch\cdot}$quyx$^{:carn}$

The transcription of the characters (人), (千), (身), and (仁), used to exhibit the pitfalls of the traditional approach to pre-Qin characters, merits further attention now that an adumbration of a Romanization system is in place. The goal of the transcription is to present a Roman version of the character that distinguishes its phonetic and semantic information; the goal is not to present all possible graphic information about the characters' structure. Thus, if the original reader of understood it as 'a word pronounced like that means ' then this is what the transcription should model, taking no account of the internal structure of . However, if instead the original reader understood the character as 'a word pronounced like ' then one must instead model the various semantic additions that graphically distinguish from . The transcription appropriate to 'a word pronounced like that means ' would be niṅ$^{:cor}$ with an appropriate subscript after niṅ to reflect however many series that one believes this form of the script had that were all pronounced as niṅ. The transcription appropriate to 'a word pronounced like ' might instead be something like $^{venter\cdot mille^x}$niṅ$^{:cor}$. However, it is not clear to me that either the horizontal line indicating a 'thousand' (*mille*) that distinguishes from or the loop indicating 'belly' (*venter*) that distinguishes from were ever systematically productive as semantic determiners; as a result, I am reluctant to include them in the inventory of semantics in Table 1. The correct transcription of a character will depend on how a particular researcher understands the phonetic and semantic components of the script as having been understood at a particular time and place.

Only phonetic compounds (*xíngshēngzì* 形聲字) have been considered so far. Some consideration is necessary for those characters that provide no key to pronunciation in their graphic structure. Characters that have no *xiéshēng* series can be transliterated directly with a Latin gloss in small capital letters. For example, I propose to transliterate 競 as CONTENTIO, 命 as IUSSUS, and 威 as VIS. There is no need to separately consider *xiàngxíngzì* 象形字, *zhǐshìzì* 指事字, and *huìyì* 會意, the differences among them being essentially palaeographic. Nonetheless, if any particular scholar thought that the conventions of graphic positioning described above provided a good model of how readers understood the *huìyì* 會意 characters, for example writing ARB:ARB.ARB instead of SILVA for 森, there is little to object to in such a procedure.

In sum, the proposed transcription scheme employs the following principles: (1) the phonetic element of phonetic compound characters are in normal type, (2) semantic elements, rendered as abbreviated Latin indications of meaning with conventions to represent the graphic relationship of the semantic to the phonetic element, are set as superscripts, (3) the main character of a *xiéshēng* series is set in capitals, and (4) characters that lack a direct indication of pronunciation are paraphrased in Latin

8 Since in this character the phonetic and semantics are the same another possible transcription is $^{man:}$quyx.

and set in small capitals. These principles are conveniently illustrated by an example (Ode 256) from the *Book of Songs*.

Kǎishū representation

無競維人、
四方其訓之。
有覺德行、
四國順之。

訏謨定命、
遠猶辰告。
敬慎威儀、
維民之則。

Narrow transcription

MA CONTENTIO ser·quiy NIṄ
SLIS PAṄ KY dic·quvn TY.
quyx:carn qruk:vid eo·tyk:cor GGAṄ,
SLIS KUYK lun·cap TY.

dic·qua dic·mmak tect:teṅ$_2$ IUSSUS,
ambˣquan can·qu TYR KKUK
KEṄ cor·tin VIS hom.ov:ṅaj,
ser·qui MIN TY TSYK.

Broad transcription

ma graṅ-s ɢʷij nin,
s.lijs paṅ gə qʷʰəns tə.
ɢʷəʔ kˁruk-s tˁək gˁraṅs,
s.lij-s kʷˁək m-luns tə.

qʷʰ(r)a mˁaʔ m-tˁeṅs m-reṅs,
ɢʷanʔ ɢ(r)u s-m-dər kˁuks.
kreṅ(ʔ)-s dins quj ṅ(r)aj,
ɢʷij miṅ tə tsˁək.

Translation

Is he not strong, the (real) man!
The (states of) the four quarters take their lesson from him;
straight is his virtuous conduct,
the states of the four (quarters) obey him;
with great schemes he stabilizes his (heavenly) appointment;

with far-reaching plans he makes (seasonal =) timely announcements;
he is careful of his demeanour;
he is the pattern of the people. (Karlgren 1950: 217–218)

References

Baxter, William H. and Laurent Sagart (2014). The Baxter-Sagart reconstruction of Old Chinese (Version 1.1, 20 September 2014) Online at http://ocbaxtersagart.lsait.lsa.umich.edu/

Coblin, W. South (2001). "'Phags-pa Chinese and the Standard Reading Pronunciation of Early Míng: A Comparative Study." *Language and Linguistics* 2.2: 1–62.

Coblin, W. South (2007). *A handbook of 'Phags-pa Chinese*. Honolulu: University of Hawai'i Press.

Fortson, Benjamin W. (2010). *Indo-European Language and Culture*. Oxford: Wiley-Blackwell.

Galambos, Imre (2006). *Orthography of early Chinese writing: evidence from newly excavated manuscripts*. Budapest: Department of East Asian Studies, Eötvös Loránd University.

Karlgren, Bernhard (1950). *The Book of Odes*. Stockholm: Museum of Far Eastern Antiquities.

Karlgren, Bernhard (1964[1957]). *Grammata Serica Recensa*. Stockholm: Museum of Far Eastern Antiquities.

Richter, Matthias (2003). "Suggestions concerning the transcription of Chinese manuscript texts—a research note." *International Research on Bamboo and Silk Documents Newsletter* 3.1: 1–12.

Takashima, Ken'ichi (2000). "Towards a More Rigorous Methodology of Deciphering Oracle-Bone Inscriptions." *T'oung Pao* (Second Series) 86.4–5: 363–399.

Schuessler, Axel (2009). *Minimal Old Chinese and Later Han Chinese: a companion to Grammata serica recensa*. Honolulu: University of Hawai'i.

Shaughnessy, Edward L. (1991). *Sources of Western Zhou history: inscribed bronze vessels*. Berkeley: University of California Press.

Xing Wen (2005). "Towards a transparent transcription." *Asiatische Studien / Études Asiatiques* 59.1: 31–60.

Appendix: Transcriptions of a Few Chinese Characters

坐 TSOJ, 髽 $^{cri:}$tsoj, 剉 tsoj$^{\cdot cul}$

言 ṄAN, 唁 $^{os\cdot}$ṅan, 閆 $^{por^x}$ṅan, 狺 $^{can\cdot}$ṅan

兢 KOT, 厥 $^{caut^x}$kot, 蕨 $^{herb:caut^x}$kot, 蠾 $^{serp.caut^x}$kot, 蹶 $^{caut^x}$kot$^{:pes}$, 蹶 $^{pes.caut^x}$kot, 闕 $^{port^x}$kot, 撅 $^{man.caut^x}$kot, 橛 $^{arb.caut^x}$kot, 橜 $^{caut^x}$kot$^{:arb}$, 巖 $^{mon:caut^x}$kot

其 KY, 騏 $^{equ\cdot}$ky, 箕 $^{bamb:}$ky, 基 ky$^{:ter}$, 朞 ky$^{:luna}$, 期 ky$^{\cdot luna}$, 稘 $^{gran\cdot}$ky, 萁 $^{herb\cdot}$ky, 諆 $^{dic\cdot}$ky, 俱 $^{hom\cdot}$ky, 欺 ky$^{\cdot dehab}$, 棊 ky$^{:arb}$, 棋 $^{arb\cdot}$ky, 旗 $^{vex\cdot}$ky, 琪 $^{gem\cdot}$ky, 祺 $^{spir\cdot}$gi, 綦 ky$^{:ser}$, 騏 $^{equ\cdot}$ky, 麒 $^{cerv\cdot}$ky. 恭 ky$^{:cor}$, 謇 ky$^{:dic}$, 塞 $^{gem\cdot}$ky$^{:ter}$, 僛 $^{hom\cdot}$ky$^{\cdot dehab}$

我 ṄAJ, 餓 $^{ed\cdot}$ṅaj, 鵝 ṅaj$^{\cdot av}$, 蛾 $^{serp\cdot}$ṅaj, 義 $^{ov:}$ṅaj, 儀 $^{hom.ov:}$ṅaj, 議 $^{dic.ov:}$ṅaj, 蟻 $^{serp.ov:}$ṅaj, 羲 $^{ov:}$ṅaj^{xio}, 犧 $^{bos.ov:}$ṅaj^{xio}

敬 KEṄ, 驚 keṅ$^{:equ}$, 儆 $^{hom\cdot}$keṅ, 憼 keṅ$^{:cor}$, 警 keṅ$^{:dic}$, 擎 keṅ$^{:man}$, 檠 keṅ$^{:arb}$

無 MA, 舞 ma$^{:obs}$, 儛 $^{hom\cdot}$ma$^{:obs}$, 譕 $^{dic\cdot}$ma, 幠 $^{lint\cdot}$ma, 撫 $^{man\cdot}$ma

于(亏) QUA, 宇 $^{tect:}$qua, 盂 qua$^{:vas}$, 竽 $^{bamb:}$qua, 芋 $^{herb:}$qua, 迂 $^{amb^x}$qua, 雩() $^{pluv:}$qua, 吁 $^{os\cdot}$qua, 訏 $^{dic\cdot}$qua, 紆 $^{ser\cdot}$qua, 圩 $^{ter\cdot}$qua, 汙 $^{aq\cdot}$qua

辰 TYR, 晨 $^{sol:}$tyr, 蜃 tyr$^{:serp}$, 振 $^{man\cdot}$tyr, 震 $^{pluv:}$tyr, 鼏 $^{cic\cdot}$tyr, 脣 tyr$^{:carn}$

告 KKUK, 誥 $^{dic\cdot}$kkuk, 郜 kkuk$^{\cdot urb}$, 皓 $^{alb\cdot}$kkuk, 晧 $^{sol\cdot}$kkuk, 浩 $^{aq\cdot}$kkuk, 梏 $^{arb\cdot}$kkuk, 牿 $^{bos\cdot}$kkuk, 酷 $^{vin\cdot}$kkuk,

鵠 kkuk$^{\cdot av}$, 窖 $^{cav:}$kkuk
真 TIN, 稹 $^{gran\cdot}$tin, 縝 $^{ser\cdot}$tin, 鬒 $^{crin:}$tin, 黰 $^{nig\cdot}$tin, 鎮 $^{met\cdot}$tin, 瞋 $^{ocu\cdot}$tin, 磌 $^{lap\cdot}$tin, 慎 $^{cor\cdot}$tin, 傎 $^{hom\cdot}$tin, 蹎 $^{pes\cdot}$tin, 瘨 $^{inf^x}$tin, 顛 tin$^{\cdot fac}$, 巔 $^{mon:}$tin$^{\cdot fac}$, 顛 $^{den\cdot}$tin, 瑱 $^{gem\cdot}$tin, 嗔 $^{os\cdot}$tin, 闐 $^{por^x}$tin, 揕 $^{man\cdot}$tin, 窴 $^{cav:}$tin, 填 $^{ter\cdot}$tin, 嶙 $^{mon\cdot}$tin, 寘 $^{tect:}$tin,
民 MIN, 泯 $^{aq\cdot}$min, 珉 $^{gem\cdot}$min, 眠 $^{ocu\cdot}$min, 敃 min$^{\cdot fer}$
昏 MVN, 惛 $^{cor\cdot}$mvn, 搰 $^{man\cdot}$mvn, 瑻 $^{gem\cdot}$mvn, 磆 $^{lap\cdot}$mvn, 緡 $^{ser\cdot}$mvn
則 TSYK, 側 $^{hom\cdot}$tsyk, 惻 $^{cor\cdot}$tsyk, 測 $^{aq\cdot}$tsyk, 廁 $^{lax^x}$tsyk
方 PAṅ$_1$, 舫 $^{nav\cdot}$paṅ$_1$, 放 paṅ$_1$$^{\cdot fer}$, 昉 $^{sol\cdot}$paṅ$_1$, 枋 $^{arb\cdot}$paṅ$_1$, 牥 $^{bos\cdot}$paṅ$_1$, 旊[9] paṅ$_1$$^{\cdot teg}$, 邡 paṅ$_1$$^{\cdot urb}$, 妨 $^{fem\cdot}$paṅ$_1$, 紡 $^{ser\cdot}$paṅ$_1$, 芳 $^{herb:}$paṅ$_1$, 訪 $^{dic\cdot}$paṅ$_1$, 髣 $^{cri\cdot}$paṅ$_1$, 仿 $^{hom\cdot}$paṅ$_1$, 坊 $^{ter\cdot}$paṅ$_1$, 房 $^{ost:}$paṅ$_1$, 防 $^{tum\cdot}$paṅ$_1$, 堕 $^{tum\cdot}$paṅ$_1$$^{:ter}$, 魴 $^{pis\cdot}$paṅ$_1$, 雱 $^{plu:}$paṅ$_1$, 汸 $^{aq\cdot}$paṅ$_1$, 彷 $^{eo\cdot}$paṅ$_1$, 祊 $^{spir\cdot}$paṅ$_1$
旁 PAṄ$_2$,[10] 謗 $^{dic\cdot}$paṅ$_2$, 滂 $^{aq\cdot}$paṅ$_2$, 傍 $^{hom\cdot}$paṅ$_2$, 騯 $^{equ\cdot}$paṅ$_2$, 徬 $^{eo\cdot}$paṅ, 榜 $^{arb\cdot}$paṅ$_2$, 蒡 $^{herb:}$paṅ$_2$

9 The variant 旊 uses 丩 rather than 方 as a non-etymological phonetic.

10 On the basis of their modern forms one might transcribe 旁 as $^{sto:}$paṅ. According to the *Shuōwén* the character 旁 has the semantic components 二 'two' and 闕 'watch tower' and the phonetic 方 (溥也從二闕方聲), e.g. $^{duo.specula:}$paṅ. Early forms of these characters (such as方 <  and 旁 < ) suggest that *Shuōwén* analysis is not correct and that 旁 is composed of 凡 (< ) above 方, e.g. $^{velum:}$paṅ. Although there is no doubt that 旁 and 方 derive from the self same series, I treat 方 and 旁 here as two distinct series because the ultimate analysis of 旁 is not clear to me.

早期漢語言文學研究中的漢字釋讀方法新探

丘內藤
倫敦大學亞非學院

提要

本文綜述以楷書轉寫早期文獻的過時做法中的常見錯誤，並提出將漢字承載的音義信息譯解為羅馬字母的直接釋讀漢字的方法。

關鍵詞

轉寫、漢語