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 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 ཤ, the letter ཉ, and the punctuation mark tsheg are difficult to distinguish; transliterating them respectively with a ‘d’, ‘n’, and a space disambiguates them clearly. In a similar fashion, transcriptions
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 *uruKUBBAR-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 唯周公于征伐東 (cf. Shaughnessy 1991: 68–73) is *a priori* no more intelligible than 宍唯公于征伐東; both are equally non-Roman. The current practice fails to reveal the structure of the script system more explicitly than the original. Although a *kāishū* 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 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 *kāishū* 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.

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 仁 (equivalent to 仁) as an example where a correct analysis in terms of *kāishū* transcription is not possible (2006: 75–76). The Chu character 身 is ultimately built on the radical 人 (equivalent to 人) but also contains in its graphic structure the characters 十 (equivalent to 千) and 心 (equivalent to 心).

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

---

3 The character 東 no more directly indicates the meaning ‘east’ than 身 itself.
It is impossible to construct a kāishū transcription that correctly captures these relationships. As Galambos points out: “if I were to transcribe 言 as either 言 or 言，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 言 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 wære wyruldcyninga
    manna mildust ond monðwǣrst
    lēodum hōost 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 kāishū 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 kāishū 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 kāishū characters.

For clarity of presentation I first exhibit my proposed transcription system using kāishū 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. 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 (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.tˤaw
召 drjewH < *[d]raw-s
超 trhjew < *thʔ[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 taw⁵⁰ˢ 超 cur-taw⁵⁰ˢ, 昭 sol-taw⁵⁰ˢ, 臥 ign-taw⁵⁰ˢ, 沼 aqua-taw⁵⁰ˢ, 詔 dic-taw⁵⁰ˢ.

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 kāishū by readers of Chinese today is very different than at the time of the script’s codification.
### Verbs and Adjectives

<table>
<thead>
<tr>
<th>Chinese</th>
<th>Pinyin</th>
<th>Chinese</th>
<th>Pinyin</th>
<th>Chinese</th>
<th>Pinyin</th>
<th>Chinese</th>
<th>Pinyin</th>
<th>Chinese</th>
<th>Pinyin</th>
</tr>
</thead>
<tbody>
<tr>
<td>行</td>
<td>eo</td>
<td>支</td>
<td>er</td>
<td>走</td>
<td>ro</td>
<td>欠</td>
<td>dehao</td>
<td>广</td>
<td>ulo</td>
</tr>
<tr>
<td>言</td>
<td>dic(o)</td>
<td>見</td>
<td>vid(eo)</td>
<td>篷</td>
<td>sig(no)</td>
<td>广</td>
<td>magn(us)</td>
<td>白</td>
<td>alb(us)</td>
</tr>
<tr>
<td>广</td>
<td>lax(us)</td>
<td>食</td>
<td>ed(o)</td>
<td>大</td>
<td>magn(us)</td>
<td>广</td>
<td>inf(irmus)</td>
<td>黑</td>
<td>nig(er)</td>
</tr>
<tr>
<td>姓</td>
<td>obs(isto)</td>
<td>白</td>
<td>alb(us)</td>
<td>止</td>
<td>ces(so)</td>
<td>高</td>
<td>alt(us)</td>
<td>立</td>
<td>sto</td>
</tr>
</tbody>
</table>

### Natural World

<table>
<thead>
<tr>
<th>Chinese</th>
<th>Pinyin</th>
<th>Chinese</th>
<th>Pinyin</th>
<th>Chinese</th>
<th>Pinyin</th>
<th>Chinese</th>
<th>Pinyin</th>
<th>Chinese</th>
<th>Pinyin</th>
</tr>
</thead>
<tbody>
<tr>
<td>日</td>
<td>sol</td>
<td>土</td>
<td>ter(ra)</td>
<td>木</td>
<td>arb(or)</td>
<td>竹</td>
<td>bamb(us)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>月</td>
<td>luna</td>
<td>禾</td>
<td>gran(um)</td>
<td>帚</td>
<td>herb(a)</td>
<td>羽</td>
<td>plum(a)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>火</td>
<td>ign(is)</td>
<td>火</td>
<td>aq(uu)</td>
<td>雨</td>
<td>pluv(ia)</td>
<td>貝</td>
<td>conch(a)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>山</td>
<td>mon(s)</td>
<td>金</td>
<td>met(allum)</td>
<td>郜</td>
<td>tum(ulus)</td>
<td>玉</td>
<td>gem(ma)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>角</td>
<td>corn(u)</td>
<td>穴</td>
<td>cav(us)</td>
<td>示</td>
<td>spir(itus)</td>
<td>石</td>
<td>lap(is)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Human World

<table>
<thead>
<tr>
<th>Chinese</th>
<th>Pinyin</th>
<th>Chinese</th>
<th>Pinyin</th>
<th>Chinese</th>
<th>Pinyin</th>
<th>Chinese</th>
<th>Pinyin</th>
<th>Chinese</th>
<th>Pinyin</th>
</tr>
</thead>
<tbody>
<tr>
<td>人</td>
<td>hom(o)</td>
<td>女</td>
<td>fem(ina)</td>
<td>瓦</td>
<td>teg(ula)</td>
<td>戶</td>
<td>ost(ium)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>門</td>
<td>por(ta)</td>
<td>邑</td>
<td>urb(s)</td>
<td>網</td>
<td>nas(sa)</td>
<td>糸</td>
<td>ser(icum)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>艿</td>
<td>tect(um)</td>
<td>曳</td>
<td>vex(illum)</td>
<td>瓶</td>
<td>vas(um)</td>
<td>厂</td>
<td>caut(es)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>舟</td>
<td>nav(is)</td>
<td>肉</td>
<td>carn(is)</td>
<td>口</td>
<td>cla(um)</td>
<td>刀</td>
<td>cul(ter)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>巾</td>
<td>lint(eum)</td>
<td>車</td>
<td>vehi(culum)</td>
<td>田</td>
<td>ag(er)</td>
<td>救</td>
<td>iuss(us)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>酉</td>
<td>vin(um)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Body parts

<table>
<thead>
<tr>
<th>Chinese</th>
<th>Pinyin</th>
<th>Chinese</th>
<th>Pinyin</th>
<th>Chinese</th>
<th>Pinyin</th>
</tr>
</thead>
<tbody>
<tr>
<td>手</td>
<td>man(us)</td>
<td>口</td>
<td>os</td>
<td>首</td>
<td>cap(ut)</td>
</tr>
<tr>
<td>耳</td>
<td>aur(is)</td>
<td>足</td>
<td>pes</td>
<td>心</td>
<td>cor</td>
</tr>
<tr>
<td>彣</td>
<td>cri(nis)</td>
<td>齒</td>
<td>den(s)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Animals

<table>
<thead>
<tr>
<th>Chinese</th>
<th>Pinyin</th>
<th>Chinese</th>
<th>Pinyin</th>
<th>Chinese</th>
<th>Pinyin</th>
<th>Chinese</th>
<th>Pinyin</th>
</tr>
</thead>
<tbody>
<tr>
<td>犬</td>
<td>can(is)</td>
<td>魚</td>
<td>pis(cis)</td>
<td>牛</td>
<td>bos</td>
<td>羊</td>
<td>ov(is)</td>
</tr>
<tr>
<td>虫</td>
<td>serp(ens)</td>
<td>馬</td>
<td>equ(us)</td>
<td>鹿</td>
<td>cerv(us)</td>
<td>鳥</td>
<td>av(is)</td>
</tr>
</tbody>
</table>

### Exclamations

<table>
<thead>
<tr>
<th>Chinese</th>
<th>Pinyin</th>
</tr>
</thead>
<tbody>
<tr>
<td>兮</td>
<td>io</td>
</tr>
</tbody>
</table>
Latin has several advantages over other languages for the transcription of semantic determiners. The use of pīnyīn (e.g. 召 taw-kǒu) 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) 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 or 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 ‘a’, and -u- rather than -w-, as the sequences ‘qu’ and ‘ku’ looking less exotic than ‘qw’ and ‘kw’. Again, if others prefer to write ‘a’ and ‘w’ or ‘w’ so be it. Karlgren includes both 蛟 and 螂 in the xiéshēng series built on the character or. According to this analysis, both characters have the semantic 虫 ‘serp’ and the phonetic determiner or quyk; they would be transliterated respectively as 蛟 serp.quyk and 螂 serp.clαxquyk. However, if one instead sees 螂 as built on 國 in order to specify a velar initial, then one may write 國 as KUYK rather than clαxquyk and transliterate 螂 as serp.kuyk rather than serp-clαxquyk. In this analysis, Karlgren’s series 0929 is subdivided into two.

或 QUYK, 或 ter.quyk, 核 arb.quyk, 殁 nas-’quyk, 関 por-’quyk, 梨 aqu-’quyk, 喫 cav-’quyk, 蛟 ser-’quyk, 或 quyk-cor, 蛟 cap-’quyk, 殁 aur-’quyk

國 KUYK, 螂 serp-kuyk, 螂 carn-kuyk

In a similar case, Karlgren includes 睘 in the series built on 袁, but whereas 睘 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 睘 in 23–11 (2009: 246–247).

袁 QUAN, 園 clαx’quen, 轛 vehi’quen, 遠 ambv-quan

瞢 QUEN, 邏 ambv-quen, 璸 gem’quen, 揆 man-quen, 園 clαx’quen, 賒 hom’quen, 翀 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 a sound cannot be transliterated, because the Middle Chinese reading of some characters points to Old Chinese ‘a’ (e.g. *mwon < *mˤən*), whereas the Middle Chinese reading of other characters points to Old Chinese ‘u’ (e.g. *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. *man.m-n-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 *kʰˤoʔ* as a phonetic element are aspirate velars of type A syllables (*kʰˤoʔ, kʰˤoʔ, kʰˤoʔ-s, kʰˤoʔ*). 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:


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 (*tur, *tʰˤuj, *k.druj, *s-tʰˤuj*) and a reading QUI (*tʰˤrij, *cʷij, *cʷij, *s-qʷij*), but the readings of the characters *lrəj-s* and *dzˤuj* do not match either. The graphic representation of these characters does not allow for a subdivision of the series, as was possible with or QUYK and 国 KUYK or 袁 QUAN and 佇 QUEN. Schuessler hypothesizes that the character *隹* originally was used to refer to two different words meaning ‘bird,’ both of which were onomatopoeic 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 (*mˤa, *mˤa, *mˤa, *mˤa*). 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: *mˤaks > *mˤaH :: *mˤa. Consequently, it is right to transliterate the whole series with the phonetic MMAK.


The character *mek* 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 [*mˤek*], 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 莫 *mek (08530) to be represented with the phonetic 辟 PEK.
The series based around 西 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, 酬 serp.lu, 道 ambx.lu, 鳩 ser.lu, 酉 can.lu

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

酉 YU, 唐 lax.yu, 栖 arb.yu, 酉 can.yu, 猪 yu.can, 鳥 vehi.yu, 鳥 herb:can.yu

酒 aq.tsu, 酉 tsu, 酬 serp.tsu, 道 ambx.tsu, 鳥 ser.tsu, 鳥 pix.tsu, 踏 pes.tsu

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

酉 RU

酉 aq.qu

酒 aq.tsu, 酉 tsu, 酬 serp.tsu, 道 ambx.tsu, 鳥 ser.tsu, 鳥 pix.tsu, 踏 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 käishū characters, some consideration can be given for phenomena confronted in earlier versions of the script. It is no surprize that the käishū 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 käishū script, they do not. It would thus be entirely sensible in transcriptions of the käishū script to distinguish three series ty₁ (之芝), ty₂ (蚩), and ty₃ (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(SSI) TY, 芝 herb:ty, 蚩(SH) ty:serp, 志(SSH) ty:cor, 詰 dic:ty:cor

In another example, if one keeps in mind the oracle bone forms of the characters 丁 and 正 (和 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(□) TEN, 頂 fac, 汀 aq:ten, 町 aq:ten, 亭 alt:teŋ, 停 hom:alt:teŋ,
正 ( ) ten:ces, 征 ten:ces, 政 ten:ces, 政 ten:ces, 証 ten:ces, 鉴 ten:ces, 蹴 ten:ces, 跣 ten:ces,

定 tect:teŋ:ces

7 Following the Indological tradition, I prefer to use ‘ŋ’ 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 TEṄ2 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 囚 *ɢʷək versus 友 *ɢʷəʔ, 右 *ɢʷəʔ, 賄 *qʷʰˤəʔ, etc.). In a case such as this, the transcription QUY (κ) 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.

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 耳 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 耳. However, if instead the original reader understood the character 右 as ‘a word pronounced like’ then one must instead model the various semantic additions that graphically distinguish 右 from 又. The transcription appropriate to ‘a word pronounced like’ that means 右 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.mille˟niṅ:cor. However, it is not clear to me that either the horizontal line indicating a ‘thousand’ (mille) that distinguishes 右 from 又 or the loop indicating ‘belly’ (venter) that distinguishes 右 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*.

**Kāishū 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**

\[
\text{MA} \ \text{CONTENTIO} \ \text{ser} \ \text{qui} \ \text{NIÑ} \\
\text{SLIS} \ \text{PAÑ} \ \text{KY} \ \text{dic} \ \text{quvn} \ \text{TY}. \\
\text{quyx} \ \text{carn} \ \text{qru} \ \text{vid} \ \text{eo} \ \text{tyk} \ \text{cor} \ \text{GGAÑ}, \\
\text{SLIS} \ \text{KUYK} \ \text{lun-cap} \ \text{TY}. \\
\]

\[
\text{dic} \ \text{qua} \ \text{dic} \ \text{mmak} \ \text{tec} \ \text{teñ} \ \text{IUSUS}, \\
\text{ambyquan} \ \text{can} \ \text{qu} \ \text{TYR} \ \text{KKUK}. \\
\text{KEÑ} \ \text{cor} \ \text{tin} \ \text{VIS} \ \text{hom} \ \text{ov} \ \text{ñaj}, \\
\text{ser} \ \text{qui} \ \text{MIN} \ \text{TY} \ \text{TSYK}. \\
\]

**Broad transcription**

ma grañ-s gʷij nin,  
slïjs pañ gə qʷhəs tə.  
gʷəʔ kʳuk-s tʰək gʷəns,  
slïj-s kʷə ak m-luns tə.  
qʷh(r)a mʰaʔ m-tʰəns m-reñs,  
gʷənʔ c(r)u s-m-dar kʰuks.  
kräñ(?)-s dins qû ñ(r)aj,  
gʷij miñ tə tsʰək.

**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


Appendix: Transcriptions of a Few Chinese Characters

bulletin of chinese linguistics 8 (2015) 48-60
鵶 kkuk, 窇 cav'kkuk

真 TIN, 焼 gran-tin, 砣 ser-tin, 鬆 crin'tin, 黒 nig-tin, 鎭 met-tin, 睜 ocu-tin, 碢 lap-tin, 偻 cor-tin, 偹 hom-tin, 鎤 pes-tin, 窱 inf'tin, 頭 tin'fac, 頭 mon'tin'fac, 羸 den-tin, 界 gem-tin, 唔 octin, 門 poe'tin, 擲 man'tin, 爛 cav'tin, 塡 ter'tin, 顴 mon'tin, 鴦 tect'tin, 民 MIN, 洮 aq.min, 磠 gem.min, 眠 ocu.min, 敗 min'fer

晈 MVN, 默 cor.mvn, 梶 man.mvn, 窊 gem.mvn, 澋 lap.mvn, 榋 ser.mvn

則 TSYK, 偽 hom.tsyk, 假 cor.tsyk, 眺 aq.tsyk, 則 lax'tsyk

方 PAṄ₁, 蝦 nav.paṅ₁, 放 paṅ₁'fer, 明 sol.paṅ₁, 楊 arb.paṅ₁, 飯 bos.paṅ₁, 壯 paṅ₁'teg, 郃 paṅ₁'urb, 妝 fem.paṅ₁, 軒 ser.paṅ₁, 蓊 herb:paṅ₁, 訪 dic.paṅ₁, 鬚 cri.paṅ₁, 敗 hom.paṅ₁, 時 ter.paṅ₁, 房 ost.paṅ₁, 防 tum.paṅ₁, 堆 tum.paṅ₁'ter, 魚 pis.paṅ₁, 雨 plur.paṅ₁, 潫 aq.paṅ₁, 彳 eo.paṅ₁, 彳 spir.paṅ₁

旁 PAṄ₂, 旁 dic.paṅ₂, 潫 aq.paṅ₂, 偽 hom.paṅ₂, 騒 equ.paṅ₂, 彳 eo.paṅ₂, 彳 arb.paṅ₂, 旁 herb:paṅ₂

---

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

10 On the basis of their modern forms one might transcribe 旁 as stœ: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 < 方 and 旁 < 方) suggest that Shuōwén analysis is not correct and that 旁 is composed of 凡 ( < 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.
早期漢語言文学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