Fuzzy Word Identification:
A Case Study from the Oracle Bone Inscriptions *

Zev Handel
University of Washington

In this paper, a new method of “fuzzy identification” is proposed for circumstances in which an exact match of an epigraphic written word with later attested forms is not possible (for example, because the word has been lost from the language). Based on our increasingly sophisticated understanding of early Chinese morphological patterns and word families, it is sometimes possible to achieve an approximate understanding of pronunciation and meaning in the absence of a precise identification.

As an illustration of this approach, I consider the oracle-bone graph 甲 as it appears in a famous eclipse inscription. This graph has been identified as 斬 zhuó and 剅 dōu (among others). I argue that any such identification is overly precise. A fuzzy identification, as a member of the word family based on root *tok with meaning ‘cut, chop’, is a more accurate reflection of the state of our knowledge and provides greater insight into the possible pronunciations and range of meaning and function of the word.

Keywords: oracle bone inscriptions, fuzzy identification, word families, eclipse

1. Introduction

The identification of graphs and words is a central concern of philologists who interpret ancient Chinese manuscript and epigraphic texts, including the early

---

* Earlier versions of this paper were presented in Berkeley on October 1, 2011 at a symposium in honor of David Keightley and in Boston on March 17, 2012 at the American Oriental Society meeting. I am grateful for suggestions received from participants at both venues, especially to Liu Xueshun for providing me with helpful advice and references to relevant scholarship. I owe special thanks to Ken Takashima for a careful reading of a draft of this article. His detailed feedback has saved me from a number of embarrassing errors and provided many valuable suggestions that have been incorporated into the final version. Because we disagree on some points, it is important for me to note that I am fully responsible for all viewpoints and errors herein. This article is dedicated to David Keightley, who first opened up the world of oracle bones to me when I was a student at Berkeley in the 1990s.
oracle-bone and bronze inscriptions. Conventionally, such identifications have been made by linking an ancient graph to a later attested graph, and then using the medieval sound gloss given for that graph (in the form of a 反切 反切 spelling) to derive a modern Mandarin pronunciation. The combination of a graphic and phonetic identification, converging on a medieval dictionary entry with a semantic gloss that fits the epigraphic context, is considered by many scholars to be both necessary and sufficient to specify the word being written. This conventional approach is based on the implicit notion that the medieval Chinese lexicographic and phonological tradition (as codified in such works as the rhyme books Qièyùn 切韻, Jíyùn 集韻, and Guǎngyùn 廣韻) encodes the definitive summa of the Chinese lexicon, to which all earlier and later stages and varieties of Chinese can be linked. While this approach is certainly of value in generating useful hypotheses, it also has potential pitfalls, most obviously in the case of ancient words which did not survive into later varieties of Chinese. In such cases the approach will either fail to produce an identification, or will produce a false identification. Though it is axiomatic that a significant subset of words in use at one point in time will have disappeared from the language thousands of years later, I am not aware that this fact has played any significant role in shaping methodological approaches to epigraphic identification.

As Takashima (2000:370) notes, there has been surprisingly little scholarly debate about the proper methodology for the interpretation of Shāng 商 oracle-bone inscriptions [OBI].¹ A notable exception has been Takashima himself, who has critiqued many practices in the field and proposed a more rigorous methodology of interpretation which he calls the “synchronic evidential approach” (Takashima 2004). His approach entails “interpret[ing] the data or issues at hand on the basis of as much intrinsic evidence as possible without drawing conclusions from the later transmitted texts and their commentaries” (2004:4). This method may be applied at every level of OBI interpretation, from the specification of the lexical properties of individual words to the interpretation of entire passages and the evaluation of Chinese historical events.

In this paper my aim is to contribute to the ongoing development of a more rigorous methodology by arguing for a new approach to the identification of individual words. This approach is meant to be supplementary to, and distinct from, Takashima’s synchronic evidential approach. My claim is that this new approach, sensitive to our current state of knowledge about the nature of the early Chinese writing system and early Chinese morphology, can be applied in cases where it is suspected that a word did not survive to be recorded in the later textual tradition. In such cases it has the potential to yield identifications that are more reliable by virtue of being less specific. This

---

¹ See Takashima (2000:370) footnote 18 for additional relevant references.
2. The identification of $<\Phi>$ in an eclipse inscription

An inscription on the right side of Yingcang 英藏 886 verso (Li Xuéqīn et al. 1985:189) is one of several in the OBI corpus that record lunar eclipses; such inscriptions have attracted considerable attention because of their potential for precise dating. A rubbing of the bone fragment is presented in Appendix 1. The inscription, a verification, can be transcribed as follows.²

(1) 七日己未庚申月有食。

The identification of all the graphs save $<\Phi>$ (.setY) is non-controversial, making it possible to transcribe them with the equivalent modern graph forms seen above. Setting aside for the moment the proper interpretation of the graph $<\Phi>$, the sentence can be translated as in (2).³

(2) “On the seventh day jǐwèi (day 56) … $<\Phi>$ … gēngshēn (day 57), the moon was eclipsed.”

The astronomical event in this inscription has been identified as the lunar eclipse known to have occurred on August 14, 1166 BCE.⁴ We use ellipses (“…”) in the translation above not to represent omitted material, but to indicate that the syntactic structure is unclear, as it depends on the syntactic role of $<\Phi>$. This graph is found between two gānzhī 干支 dates in a number of other inscriptions as well.⁵ I will refer to this usage of the graph as ‘the gānzhī context’. The graph $<\Phi>$ does not obviously represent any known word from the later textual tradition. Numerous scholars have attempted an identification of this graph and the word it writes in order to make sense of the inscription. Proposals for graphic identification have included $<豊>$, $<壡>$, $<蝕>$, $<良>$, and $<皿>$, among others (Lóng 2008:74, Yú 1941/1950:59). Let us first review some of these scholarly identifications and the methodological approaches undergirding

² The contents of typical oracle bone divinatory inscriptions can be divided into four components: preface, charge, prognostication, and verification. See e.g. Keightley (1978).
³ A similar inscription, recording the same event, is found on Yingcang 885 verso.
⁴ Liu (2012:35). The dating of this eclipse event is not directly relevant to the topic of this paper, and the evidence for it will not be presented here.
⁵ See Qiú (1993:75) for some examples.
them.\textsuperscript{6} We will not attempt to be comprehensive, but to simply present a representative range of methods employed in the service of interpretation.

### 3. Prior scholarship

Some scholars felt unable to identify the word represented by $<\phi>$, and simply made an educated guess at its approximate meaning from their understanding of the context. For example, Chén (1956:246) argued that it must refer to some kind of night-time weather phenomenon\textsuperscript{7}, while Dubs (1951:331) hazarded that it “probably means something like ‘midnight’ or ‘continuing into’”. While in a broad sense this constitutes an interpretation of the graph, it cannot be considered a lexical identification, and is therefore not directly relevant for our purposes.

Several scholars (e.g. Chang 1980:325, Chou 1964-65:245) have followed Yè (1929, as cited in Yú & Yáo 1996:2798) in identifying the graph as writing the word $y\check{\textit{in}}$, a type of sacrifice. This is based on a graphic interpretation: the two vertical lines atop $<\phi>$ appear to resemble steam rising, presumably from a sacrificial vessel; and the graph as a whole resembles the modern graph $<\text{垔}>$, which is the phonetic component of $\check{\textit{in}}$. Interpreting the iconic value of a graph, and using it to speculate about the meaning of the word written with that graph, is common practice among some epigraphers. It is based on the fact that many Chinese graphic shapes were pictographic in origin, and that their appearance was motivated by a connection in the mind of the graph’s creator between the meaning of the word being written and a particular visual representation. However, because graphic forms are stylized and judgments about the plausibility of pictograph-meaning relationships are subjective and idiosyncratic, this kind of identification is highly speculative. It can generate hypotheses, but cannot be considered sufficient for lexical identification, as it fails to constrain multiple hypotheses by use of phonetic or syntactic evidence. In this particular instance, the pitfalls of such an approach are apparent: the identification falls apart upon further examination. $<\text{垔}>$ originally writes $y\check{\textit{in}}$ ‘dam up, build up’ (modern graph $\check{\textit{in}}$), which is not consistent with the interpretation of the graph as depicting rising steam. Furthermore the bronze graph $<\text{垔}>$ from the Warring-States-era Yīn gē $\check{\textit{戈}}$, which is ancestral to $<\text{垔}>$, does not resemble the OBI graph nor does its top part resemble rising steam.

\textsuperscript{6} I am grateful to Liu Xueshun, whose presentation at the October 1, 2011 symposium in honor of David Keightley comprehensively summarized the published scholarship related to this issue. I have relied on Liu’s presentation as a guide to that scholarship.

\textsuperscript{7} Specifically, that it refers either to a clear sky with visible stars, or to a cloudy sky without visible stars.
Fân (1986:317-8) notes that the graph $<\nspl\>\hbox{101}$, in addition to its occurrence in eclipse inscriptions, is also seen in several inscriptions preceding the names of animals, in syntactic and semantic contexts equivalent to those in which known verbs of animal sacrifice are found. He concludes that the graph writes the name of a sacrifice, and that it should be understood this same way in the eclipse inscriptions as well. This hypothesis is problematic because in the sacrificial context the word written is clearly a verb, but Fân proposes that it functions as a noun in the gānzhī context. Moreover, because not only the syntactic but also the semantic context is different, there is no guarantee that the graph writes the same word in each case, as Fân supposes it must. Again, we cannot consider this to be a proper lexical identification, because neither pronunciation nor syntactic function is part of the proposed identification.

Qiú (1993) argues (following Lián 1989) that the graph is identical to that normally writing the word mǐn 輿 ‘carry out a blood sacrifice’, i.e. $<\n spl\>\hbox{101}$.

This argument is based on graphic and contextual considerations. If the arched lines near the top of $<\n spl\>$ were extended to join together, forming the top half of a circle, and the outer vertical prongs were moved toward the center, the graph would become quite similar to $<\n spl\>\hbox{101}$. Qiú further argues that $<\n spl\>$ is found in the same two contexts as $<\n spl\>$: before words designating animals and between gānzhī dates. This fact is consistent with the hypothesis that the two graphs are equivalent and write the same word. Finally, Qiú argues that the word written by this graph in the gānzhī context is not mǐn, but xiàng 廝 =逢 ‘to face towards’, whose ancient pronunciation is similar to that of mǐn 輿. The phrasing therefore presumably indicates the time when one day is ‘facing’, or passing into, the next. This argumentation is contorted at best; it relies on a series of cascading speculations, each of varying degrees of plausibility.

All of the examples above are imperfect attempts at lexical identification. Based on guesswork, speculation, and/or subjective interpretation of graphic appearance, they make a connection to a specific attested word in the later lexicographic tradition (yīn 輿 =禋, yì 輿 =曀, xiàng 輿 =逢). But they fail to present compelling phonetic or syntactic evidence that supports the identification or serves to eliminate alternative hypotheses. They all assume implicitly and unquestioningly that the word being written must be one

8 Qiú’s claim, following Lián Shàomíng, that the words mǐn 輿 and xuè 輿 ‘blood’ were written identically ($qíshí shì yīgè zì “其實是一個字”) (1993:76) in the OBI might be challenged. The graph $<\n spl\>$ is normally interpreted as writing xuè, and $<\n spl\>$ is the graph normally interpreted as writing mǐn.

9 As Qiú (1993:87) points out, both are Old Chinese yáng 陽 rhyme group words (reconstructed with rhyme *-aŋ); one is reconstructed with initial *m- and the other with *ŋ-. Alternation between these two initials within a single phonetic series, or within a word family, is common. Based on this, Qiú believes that the graph normally writing mǐn could be employed to write the near-homophonous word xiàng.
that survived to be recorded in later texts. They all present modern graphs with modern Mandarin pronunciations as if this in and of itself constitutes a complete identification; and they fail to make a sophisticated syntactic accounting that includes specification of a part of speech and the number and types of arguments.

As noted in the introduction, the methods underlying these identifications reflect the conventional view that holds sway in the field of Chinese epigraphy. This view depends on the implicit and unquestioned notion that the medieval Chinese lexicographic and phonological tradition (as codified in such works as the rhyme books Qièyùn 切韻, Jiùn 集韻, and Guǎngyùn 廣韻) contains all the lexical information necessary for an understanding of earlier stages of the language. Acceptance of this notion leads naturally to the process of making identifications by linking ancient graphs to later attested graphic forms and then using their sound glosses to generate modern Mandarin pronunciations. The combination of a graphic and phonetic identification, converging on an entry in the medieval lexicographic tradition, is considered both necessary and sufficient to specify the word being written.

We need not dwell further on the proposals summarized above; instead, let us turn to some more sophisticated attempts at identification.

Keightley (1978, 1982) identifies <虤> as graphically ancestral to <犠>, the phonetic component of the phonetic series (aka xiéshēng 諧聲 series) of characters that Karlgren 1957 [GSR] labels 1235. Adopting Karlgren’s reconstruction, he gives the pronunciation of the graph as Old Chinese *d’ūg (implying modern reading tōu), and argues that it here writes the word later conventionally written with GSR 1235b <犠>, i.e. *tík (modern reading zhuó) meaning ‘chop, cleave’. This is the meaning the word is assumed to have in the oracle bones when it precedes animal names, referring to the chopping up of meat in a ritual ceremony. In the gānzhī context, however, Keightley (1982:550) argues that its meaning is a semantic extension of ‘cleave’, so that the gānzhī expression refers “to the no-man’s-land period in the dark hours of night when one kan-chih [gānzhī] day cleaved to another and dreams, toothache attacks, childbirths, and lunar eclipses were all recorded as occurring…” Keightley therefore translates inscription (1) as in (3) below:

---

10 Consider in this regard the two identifications yīn 禿 and yì 曝 mentioned earlier. These may simply look like modern forms, but in fact they are not found in modern Mandarin or Modern Standard Written Chinese. Their Mandarin pronunciations are based on medieval sound glosses and learned reading traditions. In Guǎngyún, yīn 禿 has the fānqiè spelling yú zhēn 於真 (homophonous with yīn 因); yì 曝 has the fānqiè spelling yú jì 於計.
(3) “On the seventh day (after the original divination and prognostication of harm), in the t'ou [tóu] period when chi-wei [jìwēi] (day 56) cleaved to keng-shen [gèngshēn] (day 57), the moon was eclipsed.”

We note some inconsistency on Keightley’s part, in that he recognizes that the word for ‘cleave’ has modern pronunciation zhuó, but he nevertheless uses the transcription tóu in his translation, apparently because he considers the graphs <𠁁> and <𠁂>, and therefore their conventionally associated readings tóu and zhuó, to be interchangeable. We also note that the plausibility of Keightley’s claim relies in part on the dual meanings of English cleave, ‘to cut apart’ and ‘to adhere together’. By using ‘cleave’ as a gloss in the two contexts, Keightley is able to pivot the Chinese word from one meaning to another without making it seem like a significant semantic shift.

Keightley’s proposal for the graphic identification is anticipated (and presumably inspired) by that of Yú (1950:59-65), who also interpreted <𠁂> as graphically ancestral to <𠁁>, noting that the bronze graph forms for hú 壺 ‘pot, kettle’ are composed with a lower part that is nearly identical to <𠁁> and an upper part that appears to be a representation of a lid. Among the words that <𠁁> represented in later usage was dòu (conventionally written <𨪐>), ‘a kind of wine vessel’. <𠁁> is therefore presumably a pictograph, identical in form to the graphs representing a kettle, but lacking the lid.11

Keightley’s identification at first glance may seem similar to those we have criticized earlier as inadequate. However, it differs in some key respects. First, the graphic connection is not based on speculation about either iconic value or variant forms. The graphic relationship upon which the identification is premised is a precise one; this precision is further correlated with, but does not depend on, the presumed iconic value of the pictographs involved. Second, a direct connection to pronunciation is made through the phonetic elements of characters found in the same phonetic series. Third, a metaphorical extension is presumed to account for the same graphic form occurring in two distinct lexical and syntactic contexts (Keightley does not specify whether this extension is accompanied by a change in part of speech or other lexical features, although this is implied by his analysis).

Takashima’s (1979, 2010) interpretation is similar, and is credited to both Yú and Paul Serruys. Takashima argues (2010:31-33) that “the graph 䈞 should be transcribed

---

11 The graphic similarity is the same cited by Jao (1959:86), but Jao’s conclusion is different. The graph <簠> is attested in Shuōwén Jìězì 說文解字, glossed “酒器也。从金，象器形。或省作簠。” “Wine vessel. Composed of <金> ‘metal’ and <簠>, which resembles the shape of the vessel. Also written in abbreviated form as <簠>”. Note that the graph with metal radical does not appear at GSR 1235. Karlgren glosses the phonetic alone, 1235a, as ‘wine vessel’, following Shuōwén.
as "dou"，thus asserting the direct graphic connection between the OBI form and the modern graph, and here writes the word "zhuo" ‘to cleave, cut’ (which was written with <斤> alone before it acquired the determinative <刀>).¹² Like Keightley, he argues for a metaphorical extension (“one can ‘cut or cleave’ a sacrificial victim, just as one can ‘cut or cleave’ the flow of time for practical demarcation and designation”), but he explicitly states that the word in both the animal-name and gānzhī context is the same, with the same lexical properties. So “the formula NP1 + 斲 + NP2 meant literally “(when) NP1 ‘cleaved/cut’ NP2”.

Finally, Liu (2012), after summarizing the previous scholarship on the issue, concurs generally with Keightley and Takashima, but further proposes that "dou" [dōu?] (also meaning ‘cleave’) is preferable to "zhuo" as an identification. It is unclear whether Liu is arguing that all three of these graphs are different ways of writing the same word (despite the different conventional pronunciations), or if his preference for

---

¹² Many academic papers provide pīnyīn transcriptions without tone marks. This can be problematic when the issue at hand is the identification of epigraphic forms, since the tone mark may be crucial in tracing the given Mandarin pronunciation back to a medieval gloss. Many obscure graphs have multiple fānqìe glosses in various medieval dictionaries, which sometimes correspond to distinct modern Mandarin pronunciations differing only by tone. In this case, Takashima’s "dou" presumably represents dòu, although this is not absolutely certain. The Guǎngyùn pronunciations for <酅> and <斤>, labeled as variants of each other (see previous note), is tú kǒu 徒口, which corresponds to modern Mandarin dòu. This also matches the pronunciation found in GSR.

¹³ A similar line of argument was advanced by Jao (1959:86, 139), but with different results. Jao concluded that <酋> wrote two unrelated words having similar pronunciations: yi 壹 = 燚 ‘cloudy’ in the gānzhī context and yi 壹 = 成 = 𭜟 ‘to bury’ in the sacrificial context. His argument is convoluted. He notes that <酋> is identical to the bottom part of bronze graph forms for hú 壶 ‘pot, kettle’, and therefore that it is the original form of the graph yi 壹. (Shuōwén 説文 says that 壹 is derived from 壶, with 吉 as phonetic; this is apparent from the seal script form.) This allows Jao to argue that <酋> might write any word that is phonetically similar to yi 壹. Since the gānzhī context suggests a celestial phenomenon, <酋> in this context should be interpreted as writing the word yi 燚 ‘cloudy’. Takashima dismisses Jao’s (1957:139) argument that this graph writes the word yi 壹 = 成 = 𭜟 ‘to bury’ as “tortuous”, but he does not explicitly acknowledge that Jao makes this claim only for the sacrificial context: “[Jao’s interpretation of ‘to bury’] does not fit the temporal context” (2010:32). Jao’s conclusion can only be evaluated if viewed as part of his broader argument that <酋> is a graphic form of yi 壹 and therefore could write either of the words conventionally written as yi 燚 ‘cloudy’ or yi 燚 ‘to bury’, depending on context. Nevertheless, Jao’s argument has a fundamental flaw, which is the absence from the OBI graph of the phonetic element 吉 remarked on in Shuōwén. There is thus no direct phonetic connection between the graphic form <酋> = <酋> and the words that Jao believes it represented.
the former (*dou* **謬/謬**) over the latter (*zhuó* **虀**) is based on an assumption that these are significantly different words.

### 4. Evaluation of prior scholarship

Keightley’s and Takashima’s identifications are the most linguistically sophisticated, and are supported most clearly by the evidence. But despite the overall soundness of their argumentation, they too proceed from the assumption that the OBI word in question can be precisely identified with a word attested in the later textual and lexicographic tradition.

There is nothing inherently wrong with this approach as one way to generate hypotheses about character identifications. After all, it is certainly true that many words in varieties of Chinese spoken in the oracle-bone and bronze periods survived into later stages of Chinese, and were recorded in various dictionaries. It is almost certainly also true that some of those words that disappeared from the mainstream language were preserved in dialects or early texts, and thus made their way by different means into comprehensive dictionaries. Searching for such connections is a necessary and important component of philological scholarship. But it is not sufficient to the entirety of the task, and the reason for this is not difficult to grasp: there must have been many ancient words that yielded no surviving direct descendants in later varieties of Chinese. All languages show lexical attrition over time. In the case of the OBI, the millennium that elapsed before the first systematic character dictionary, *Shuōwén Jiězì* 說文解字, was produced ensures that the number of lost words would have been significant.14

When attempting to identify OBI words that fit into this category, the conventional approach will either fail to produce an identification or will produce a false. Now, one might speculate that in the case of a word attested in the oracle-bone inscriptions that does not survive into later times and is not recorded in later dictionaries, there is no hope of lexical identification. But this is not necessarily the case; I suggest that for some words a “fuzzy identification” can be made: a middle ground between the claim of precise identification with a later attested word and the claim that nothing meaningful can be said. I will take the identification of the word written with *<Φ>* in the eclipse inscriptions as a case study for this approach.

If we wish to improve on the conventional approach, it is first necessary to define as precisely as we can what it means to identify a word given a target graph in an

---

14 By way of analogy, consider the large amount of words in the Old English text *Beowulf* that have not survived into the English of today. An example that will be discussed below is *ūrig* ‘moist’, seen in the compound *ūrig-feðera* ‘dewey-feathered’.
epigraphic source such as the OBI. We can characterize the conventional approach as having two parts, as follows:

a. Locate a kāishū 楷書 (‘modern form’) graph in the received tradition with a semantic gloss that fits the OBI context, ideally but not necessarily having a graphic or phonetic connection with the target graph;

b. Use the Middle Chinese sound gloss for the modern graph to derive a Mandarin pronunciation.

The combination of the kāishū graph and the Mandarin pronunciation, with the semantic gloss, are then presented as a complete identification.

A more sophisticated identification would separate graphic and lexical identifications (while explaining the link between them), and would expand on the features necessary for complete lexical identification, viz.: 15

a. Graphic identification: To which graphs or graph components of later times can we diachronically link the OBI graph?

b. Lexical identification: What spoken word of the OBI language is represented by the OBI graph?

i. meaning

ii. lexical features (part of speech, valency, etc.)

iii. pronunciation

Ideally, the pronunciation component should be the actual pronunciation of the word, i.e. as good an Old Chinese reconstruction as is feasible given the current state of our knowledge. It is of course acceptable to “roll” such a pronunciation forward into a modern Mandarin pronunciation for convenience of reference, but we must not lose sight of the fact that the modern pronunciation is nothing more than a convenient fiction.

We should admit the possibility that a partial identification, in which only some of the four criteria above can be specified, may be the best we can do, and that such an

15 Among the few scholar who adopt a rigorous approach to identification similar to what is outlined here is, again, Takashima (2000:371-372). In his 2009 article on the OBI graphs ji 祭 and sì 祀, Takashima examines in the syntactic properties of the word each graph writes, evaluates the presumed set of word-family members in order to identify the morphological variant being written and its archaic pronunciation, and supplements this with appropriate palaeographic analyses. The result reveals lexical features of the two words as used in the OBI that are quite different from what is found in later Chinese textual usage. But Takashima’s methodology still presumes that the phonological and morphological OBI forms he identifies are reflected in directly-descended later-attested Middle Chinese forms.
identification is not without value. Though incomplete, it is more reliable than a complete identification lacking an evidential foundation. One problem with many of the proposed identifications of 小黑 that were reviewed above is that they are over-specified on the basis of limited evidence. A guess is made about the semantics of the word based on context (e.g. the word refers to a weather-related phenomenon); a similarity to a later graphic form is apprehended; and then a very precise identification somehow follows.

5. The role of writing, word families, and morphology

We have long recognized that the traditional view of the Chinese writing system, the view that underlies the entire lexicographic tradition, is historically not applicable to the epigraphic and manuscript texts of the 漢 and earlier. The lexicographic tradition attempts to standardize a one-graph-to-one-morpheme relationship, fixing the form of the graph that “properly” represents each morpheme. But as numerous studies have shown, Chinese writing in practice was much more fluid. Words were frequently written with graphs that approximated their sounds, and those graphs may or may not have had semantic determiners added to help disambiguate their referent morphemes. Even within a single text, the specific determiner (or its absence) might vary each time the same morpheme was written.

This fact is well known to epigraphers, which is why identifications involving an attested graphic form and its conventional graphic equivalent, like “yǐn 廠=禋”, “yì 壹=壹”, and “zhuó 𠁁=斲”, can be made without a need for elaborate explanation. But graphic variability is not the only factor that must be considered when establishing a link between written forms and spoken words. The existence of word families in Chinese—sets of etymologically related words that are similar in pronunciation and related in meaning—has long been recognized (Karlgren 1933). A simple example of such a word family in Modern Standard Mandarin is:

(4a) cháng 長 ‘to be long’
(4b) zhǎng 長 ‘to grow (intr.)’
(4c) zhāng 張 ‘to stretch (tr.)’
(4d) zhàng 帳 ‘curtain’
etc.

Indeed, one might well argue that it is not applicable to unofficial or informal Chinese writing in any era, given e.g. what is found in medieval Dūnhuán 敦煌 manuscripts or modern Chinese handwritten shop signage. For the purposes of this paper, however, we need only be concerned with the writing found on non-received texts of the pre-Hàn period.

See e.g. Boltz (1994).
In this, Chinese is no different from other languages; consider for example these members of an English word family:

(5a) water
(5b) wet
(5c) wash
(5d) winter
(5e) otter
(5f) whiskey
(5g) inundate
(5h) vodka
(5i) hydrant

etc.

Two basic historical-linguistic processes lead to the existence of word families: derivation of new words from existing words, and dialect/language mixture (i.e. borrowing). The words in (5) are all ultimately derived from the Proto-Indo-European root *wed- ‘water, wet’. Winter, for example, is inherited via Germanic *wintruz, descended in turn from the nasalized form *we-n-d- of the PIE root, meaning ‘wet season’. Like winter, the words wet, wash, water, and otter are all directly inherited from PIE via Germanic, with various morphological processes developing them at different stages. In contrast, whiskey, inundate, vodka, and hydrant are borrowings into English from Irish, Latin, Russian, and Greek respectively of words derived from *wed- via a series of morphological developments within the histories of those languages. For example, whiskey is from Old Irish uisce ‘water’, in turn from the suffixed zero-grade form *ud-skio- of the PIE root.18

It is increasingly clear that the origin of many word families in Chinese can be attributed to morphological processes that were active during the Old Chinese period (and perhaps earlier). While these processes remain somewhat poorly understood, there is an increasingly clear consensus that they primarily involved sub-syllabic affixation on root syllables (see e.g. Baxter & Sagart 1998), and scholars are now making good progress in uncovering the semantic and phonological patterns underlying these

18 For these etymologies, see the American Heritage Dictionary [AHD], 5th Edition, p. 2063, under wed . The various grades and forms of PIE roots are described on pp. 2032-2033. There is considerable variation in PIE reconstruction. For example, the AHD reconstructs the source of English water as the “suffixed o-grade form” *wod-ōr, while the same PIE word is reconstructed as *wōdr- by Mallory & Adams (2006:125). The phonetic details of the reconstructed forms are not relevant to the argument made here.
processes, and in some cases in proposing derivational affixes with specific phonological and semantic effects. These effects can include shifts in word class (part of speech), valency, and meaning. As an example, consider the word family given in (4) as reconstructed in the new (and still somewhat in flux) Old Chinese reconstruction system of Baxter & Sagart.\(^{19}\)

\[(4a) \text{cháng 長} 'to be long' \quad \*Cə-N-تراَج > درjang
\]

\[\quad [*N- prefix derives an ‘endopassive’ intransitive verb: ‘to be in the state of having been stretched’]\]

\[(4b) \text{zhāng 長} 'to grow (intr.)' \quad *تراَج-ʔ > ترَنجX
\]

\[\quad [*-ʔ suffix derives an ‘endoactive’ verb: ‘to stretch oneself’]\]

\[(4c) \text{zhāng 張} 'to stretch (tr.)' \quad *C.تراَج > ترَنج
\]

\[\quad [the reconstruction of *C is supported by the Sino-Vietnamese pronunciation with initial gi-]\]

\[(4d) \text{zhàng 幕} 'curtain' \quad *تراَج-s > ترَنجH
\]

\[\quad [*-s suffix derives an ‘exopassive’ noun from verb: ‘that which has been stretched’]\]

This set of words implies the existence of a hypothetical root *تراَج ‘to stretch (tr.)’—a root that is not listed in Baxter & Sagart 1.00 and is apparently not attested in the Old Chinese textual corpus. While the precise phonological and semantic details of the derivations remain matters of discussion and dispute among scholars, the fact that the members of such word families are derived through some combination of morphological derivation and borrowing is not in doubt.

The relationship between the words that make up a word family and the characters that are used to write them is a complex one. Because of the necessary similarity of pronunciation of words within a single word family, it is quite common for them to be written with the same phonetic element, as is the case with the ‘stretch’ family in (4), in

---

\(^{19}\) Baxter & Sagart OC reconstruction version 1.00 (February 20, 2011). The reconstructions are presented in the format *OC > MC, where OC is Old Chinese and MC is Middle Chinese. I have modified the reconstructions slightly for clarity of presentation. A hyphen “-” indicates that Baxter & Sagart have identified a morphological affix; a period “.” indicates that no evidence for an affix has been identified. Because not all the details of the reconstruction have been published, it is not always possible to say with certainty what affix has been identified; the notes above represent my own interpretations, influenced by the analysis of Schuessler (2007).
which we see recurrence of the phonetic element <長>. This was presumably more likely to be the case if the graphs were selected before phonological changes obscured the morphological relationships by deforming the similarities in pronunciation of the root syllable across the family. But it is important to recognize that the members of a word family needn’t be written with similar graphs, and that word families that are not represented by a single phonetic series of graphs are quite common. For example, Schuessler (2007:551) argues that the words yíng 迎 ‘meet’, cí 諛 ‘dispute’, yà 訝 ‘meet’, and sù 訴 ‘go up against’ (among others) belong to the same word family, yet the graphs used to write them are unrelated to each other (see Appendix 2).

Some important conclusions follow from this understanding of graphic variability and word family relationships; these conclusions bear directly on the methodology of epigraphic identification.

The first conclusion is that, in early texts, we must expect that any of the words in a single family could be written in a variety of ways, possibly all involving the same phonetic element; these graphs will not necessarily correspond to the conventional graphs of the later lexicographic tradition. For example, we might expect to see any of the words in (4) written with the simple graph <長> (or, more properly, with this graph’s ancestral form) or with any number of graphs employing <長> as phonetic. But we might also see some of these words written with another graph associated with root syllable shape *tranj.*

The second is that, just as in general a language will experience lexical attrition over time, for any given word family it is possible that some words of the family that existed in OBI times did not survive into later stages of the language. But because other words in the word family do survive, these obsolete words can be said to have surviving ‘nieces and nephews’ that are attested at later times.

By way of analogy, there are numerous examples in English of word family members that have become obsolete. Old English ūrig ‘moist’ has no descendant in Modern English, but the word family members udder and urine (the latter borrowed from Latin āurīna), both ultimately from the same PIE root, do both survive.20 Old English wer ‘man’, which survives opaquely in the compound werewolf, is otherwise obsolete. But English still has word family members world and virtuoso, also ultimately from PIE *wē-ro- ‘man’ (American Heritage Dictionary of the English Language 2011:2065).

---

20 See American Heritage Dictionary of the English Language 2011:2043 *euo-dh-r* and p. 2065 *wē-r*; and also Mallory & Adams 2006:126, where the PIE root is reconstructed as *we/oh,r.*
6. Fuzzy identification

When the epigraphic evidence suggests an approximate pronunciation (based on graphic form) and a general semantic area (based on context), this may be sufficient for us to identify a word family that the attested word likely belongs to. At the same time, we may lack sufficient evidence to make a specific identification, especially if the word in question failed to survive into later times. This is the situation in which a fuzzy identification is possible: we may be able to specify the root syllable shape, suggest possible affixes, and propose derivational semantics. While this would not permit a precise OC reconstruction or modern Mandarin pronunciation, and would not yield a particular graph from the later lexicographic tradition, it nevertheless embodies a specific set of hypotheses and can be considered an identification, one that is academically rigorous in that it is forthcoming about the degree of precision that the evidence accords.

Indeed, such a “fuzzy identification” is preferable to either a failure to make an identification at all or to an overly-specific identification with a later attested word or graphic form.21

Let us consider how we might make a fuzzy identification of $<\Phi>$ that resolves or avoids the problems identified in earlier scholarship. Based on that scholarship, we can establish the following givens:

a. $<\Phi>$ is graphically ancestral to $<\Xi\Xi>$ (here we agree with the graphic analysis of Keightley (1982) and Takashima (2010));
b. $<\Phi>$ writes two etymologically related words: (a) a verb meaning ‘cut, cleave’ that takes the name of an animal as its object and designates a ritual activity; (b) a word that links two sequential dates, indicating the pre-dawn time of night at which they ‘cleave’ together.22

---

21 Takashima’s 2009 study of OBI ji 祭 and sì 祀, already mentioned in an earlier note, takes a similar approach in that it considers entire word families, for example suggesting a number of words that are etymologically related to ji 祭 (2009:39). However, Takashima’s study doesn’t consider the possibility that ji 祭 represents a word in this family that is not attested in the later lexicographic tradition, and only considers the two possible Old Chinese pronunciations for ji 祭 of *tsjats and *tsriats, i.e. the two that are backwards projections of attested pronunciations in Middle Chinese. (These reconstructions are in the system of Li Fang-Kuei, rather than the Baxter & Sagart system used here.)

22 It is tempting to note a parallel here with the English word cleave, which has two seemingly antonymic meanings: ‘to cut apart’ and ‘to adhere together’. But the situation is not analogous. These two English words cleave are historically distinct, deriving from different Old English verbs that go back to unrelated PIE roots. The two words became homophonous only after the Middle English period. See the relevant AHD entries for details.

15
It is of course the second of these two words that occurs in the eclipse inscription in (1).

In my view we lack sufficient evidence to identify the word in (1) as directly ancestral to that written later as <斲> (i.e. zhuó), that written later as <剅> (i.e. dōu), or any other word in the lexicographic tradition. We have no justification for this level of precision because we have no direct evidence in the lexicographic tradition or later textual corpus for a word with the specific meaning and morphosyntactic features that <餼> in the gānzhī context appears to have.

At the same time, the graph, together with the known principles of the Chinese writing system, implies a basic syllable shape; and the usage in (1) is consistent with derived semantics related metaphorically to the notion ‘cleave, cut’, as articulated by Keightley (1982). This leads us to the plausible hypothesis that the word in (1) belongs to the same word family as zhuó 斬. That family appears to be quite extensive. Schuessler identifies over a dozen Chinese members; these are assigned to four distinct root shapes that Schuessler believes are ultimately related to each other. (See Appendix 3 for Schuessler’s complete word family.) For illustrative purposes, ten of these words are presented below, with Schuessler’s 2007 “minimal” Old Chinese reconstruction, and Baxter & Sagart’s reconstruction.23

<table>
<thead>
<tr>
<th>*tok root</th>
<th>Schuessler (2007)</th>
<th>Baxter &amp; Sagart 1.00 (2011)</th>
</tr>
</thead>
<tbody>
<tr>
<td>chù 觸</td>
<td>*tʰok ‘knock against’</td>
<td>*tʰok &gt; tsyhowk ‘knock against’</td>
</tr>
<tr>
<td>zhuó 斬</td>
<td>*trôk ‘strike’</td>
<td>*t&lt;i&gt;r&gt;ok &gt; traewk [not in B&amp;S]</td>
</tr>
<tr>
<td>zhů 斬</td>
<td>*trôk ‘chop’</td>
<td>*Ca.t&lt;ro&gt;k &gt; traewk ‘chop, cleave’</td>
</tr>
<tr>
<td>zhū 斬</td>
<td>*trok ‘cut’</td>
<td>*t&lt;ro&gt;ok &gt; trjowk [not in B&amp;S]</td>
</tr>
<tr>
<td>zhuó 啄</td>
<td>*trôk ‘peck up’</td>
<td>*mə-t&lt;i&lt;r&gt;ok &gt; traewk ‘to peck’</td>
</tr>
</tbody>
</table>

23 Middle Chinese is also given in the B&S notational system. Note that “<i>r>” indicates that medial *r is a morphological infix. Not every word occurs in Baxter & Sagart’s corpus of reconstructed words. When a word does not appear, I have supplied a conservative reconstruction based on my understanding of their system and principles, and marked this with “[not in B&S]”.

24 *mə is probably the volitional prefix in this word.
### Fuzzy Word Identification: A Case Study from the Oracle Bone Inscriptions

<table>
<thead>
<tr>
<th>*to root</th>
<th>Schuessler (2007)</th>
<th>Baxter &amp; Sagart 1.00 (2011)</th>
</tr>
</thead>
<tbody>
<tr>
<td>zhòu 嗷</td>
<td>*toh ‘beak’&lt;sup&gt;25&lt;/sup&gt;</td>
<td>*tos &gt; tsyuH [not in B&amp;S]</td>
</tr>
<tr>
<td>shū 殊</td>
<td>*do ‘cut off, kill, die’</td>
<td>*do &gt; dzyu ‘die; different’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>*tuk root</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>zhú 篆</td>
<td>*truk ‘beat, stamp earth’</td>
<td>*truk &gt; trjuwk ‘pound earth, build’</td>
</tr>
<tr>
<td>zhù 祝</td>
<td>*tuk ‘cut off’</td>
<td>*tuk &gt; tsyuwk ‘cut off, break’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>*tu root</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>dǎo 擊</td>
<td>*tùʔ ‘beat, pound’</td>
<td>*tùʔ &gt; tawX [not in B&amp;S]</td>
</tr>
</tbody>
</table>

Focusing on Baxter & Sagart’s reconstructions, note that phonological variation across the family includes initial voicing and aspiration distinctions (t vs. tʰ vs. d), presence or absence of pre-initial consonants (C, m), presence or absence of medial r, presence or absence of post-final tonogenetic consonants (ʔ, s), and alternation between Type A (notated by ʷ) and Type B. All of these distinctions are potentially morphological.

The Chinese word family is descended from a recognized Sino-Tibetan root; the word family extends across the Sino-Tibetan languages. Matisoff (2003:357, 363) identifies it this way:

PTB *r-tuk ‘cut, knock’ > WT rdug, WB taut, Lushai tuk, Garo dok / dak; compare Chinese zhuó 斬, probably the same word as that written 斬

Given that one function of infix *r in Baxter & Sagart’s system is to derive a verb with a “distributed/repeated action”, it seems likely that when <क़> appears in the OBI representing a sacrificial verb taking the name of an animal as its object, it is writing a verb like *t<क़<r>ok or *Cә.t<क़>rök, and is therefore possibly directly ancestral to later zhuó 斬 or zhuó 斬.<sup>26</sup> But we would not expect to find this infix in the word that appears in the

---

<sup>25</sup> The reconstruction *tok-s is also possible given the Middle Chinese pronunciation, although it is not commensurate with the phonetic element of the character. A reconstruction *tok-s would affiliate this word with the *tok root rather than the *to root.

<sup>26</sup> While Baxter & Sagart provide an explanation of the various functions of the affixes in their system, they do not provide a morphological analysis for individual reconstructed words. In many cases the identity of the affixes is clear, but in other cases there are multiple possible interpretations, and we must await further clarification in future publications. It is quite possible that some inconsistencies are accidental, as this is a pre-publication draft work; for
ganzhī context, which would presumably be derived from the same ‘cleft’ root, because this context is not compatible with iterated action. Considering the unattested bare root in Chinese, we could propose that the ganzhī-context word represented by <議> was *t'ok (with no corresponding attested word or graph in the lexicographic tradition), which by regular rules of sound change would develop to MC tuwk and then Modern Mandarin dū. But this too would be an overly precise hypothesis, as other root forms are possible, and other derivational affixes (perhaps not even yet recognized to exist in Old Chinese) might have been present. This possibility is further supported by noting that the verb ‘chop’ in the animal context implies the existence of an animate agent. Therefore the inanimate subject found in the ganzhī context raises the possibility that we have not just a metaphorical extension of the notion of ‘chop’, but quite possibly a differently derived form whose distinct lexical properties include a non-animate agent argument.

It is therefore preferable to propose a “fuzzy identification” for <議> in ganzhī context, along these lines:

a. <議> writes a morpheme with syllable shape based on the root *tok;
b. <議> writes a word in the ‘cut, chop’ word family, and so is etymologically related to such later attested words as zhuó 斬;
c. this word appears to have no direct descendant; its precise reconstruction (and therefore modern pronunciation) is unknown because its morphological structure is unclear.

In my view, this identification obviates the need for the kind of hair-splitting like that put forward by Liu Xueshen (2012:26), who argues that “dou 鏡/鏡 is a better choice than zhuo 斬”. I would argue that there is no meaningful difference between these two identifications; rather, the choice between them is a manifestation of exactly the kind of fuzziness that we have no basis for resolving.28

...
7. Conclusion

I have argued that we cannot expect that every word found in early epigraphic sources like the Shāng oracle-bone inscriptions and the Zhōu 周 bronze inscriptions survived into later recorded stages of Chinese. In many cases there is a continuous lexical history, and in such cases solid philological work can often uncover the connection, even if the target word was written with what, from the modern perspective, is a non-conventional graph. In these cases the conventional approach of identifying the word with a particular graph in the medieval lexicographic tradition is a reasonable approach, especially when refined by the kind of rigorous methodological constraints advocated by Takashima (2000, 2004).

In the case of words that do not survive, there are two possibilities, both of which call for “fuzzy identification”. One is that the later lexicographic tradition contains no traces of the word. We may be able to say something about the word’s ancient pronunciation based on the phonetic evidence associated with the graph that writes it, and we may be able to say something about the word’s meaning and lexical features based on the context in which it occurs. But a precise identification is not possible, at least not according to the conventional approach. Of course, it is also possible that we lack sufficient context or evidence to make any kind of identification, in which case even a fuzzy identification should not be attempted.

The second possibility is that the word in question is part of a broader word family, some members of which survived in later stages of the language, and are preserved in the lexicographic tradition. Where the conventional approach would attempt to link the word to one of the lexicographically attested words, I would argue instead for a more cautious approach, in which the phonetic and semantic range of the word family, in combination with the semantic and syntactic context of the word in question, permit us to make general hypotheses about pronunciation, meaning, and perhaps even morphological structure. In such a case, a fuzzy identification is always possible. This, I believe, is the proper approach in the case of the gānzhī-context word <酐>.

One real difficulty for the philological researcher is how to tell the difference between words that have direct lexical descendants and those that do not, and therefore

---

29 We may also be able to identify likely cognates in non-Chinese languages; these cognates may in turn help us make more precise hypotheses about the pronunciation and semantics of the Chinese word, based on typical patterns of cognate correspondence.
how to decide whether to make a precise identification or a fuzzy identification. It seems to me that when graphic, lexical, or semantic features of the word under investigation differ significantly from the features of later attested words, then one should err on the side of caution. In other words, even if a plausible connection between an OBI word and a later attested word can be established, this does not prove that the later word is a historically continuous direct descendant of the earlier word, and the possibility of an indirect connection should be entertained.

Seen in this light, fuzzy identification is more academically rigorous than a precise identification that goes beyond what the available evidence can bear. It allows the formulation of a hypothesis without positing (implicitly or explicitly) a degree of precision that cannot be justified. Properly formulated, a fuzzy identification constitutes an expression of the degree of precision that we are able to achieve given the current state of our knowledge, while leaving open a range of more specific possibilities which further research or additional evidence may allow us to whittle down in the future. Perhaps equally important, a fuzzy identification is an explicit refutation of the implicit assumption that there is nothing more to the Chinese lexicon than what is found in the medieval lexicographic tradition. This assumption has unnecessarily handcuffed epigraphic and philological investigation.  

Fuzzy identification does engender a minor problem of notation. If we are unable to precisely specify a modern pronunciation or graphic form, how can we conveniently notate our fuzzily identified word?

Perhaps one solution is to borrow a mathematical symbol such as ± “plus or minus” or ≈ “approximately equal to” as an indicator of approximation.

We might then notate our identification of the gānzhī-context word <𠁁> this way:

±<𠁁> (written with this graph or another graph that can represent the same root shape, with or without semantic determinatives)

±√TOK (pronounced on root syllable *tok, perhaps with additional elements some of which might be morphological affixes)

±‘chop/cleave/strike’ (expression of general semantic range)

or, more compactly: {±<𠁁>, √TOK, ‘chop/cleave/strike’}.

We would then supplement this notation with an explanation of further details concerning semantics and lexical features.

---

30 It has also contributed, in my view, to serious methodological problems in Chinese dialectology, including an over-reliance on běnzìxué 本字學 in dialect lexicon studies. For a broader critique of the pernicious influence of this assumption on Chinese dialect studies, see Akitani & Handel (2012).
Appendix 1: Yīngcáng 英藏 886 verso (Lǐ Xuéqǐn et al. 1985:189)
Appendix 2: Schuessler 2007:551

Table Y-1  Meet, against for yâ 御近診

<table>
<thead>
<tr>
<th>yâ</th>
<th>ʔna</th>
<th>ʔnak</th>
<th>ʔŋak</th>
<th>ʔŋəg</th>
<th>ʔŋən</th>
</tr>
</thead>
<tbody>
<tr>
<td>ʔθən</td>
<td>*ʔθən</td>
<td>wəɾə *ʔθək</td>
<td>səɾə *ʔθək</td>
<td>ʔən</td>
<td></td>
</tr>
<tr>
<td></td>
<td>WT fp face, look</td>
<td>resist, oppose</td>
<td>go up against</td>
<td>ʔən</td>
<td></td>
</tr>
<tr>
<td>ʔŋən</td>
<td>ʔŋən</td>
<td>ʔŋən</td>
<td>ʔŋən</td>
<td>ʔŋən</td>
<td>ʔŋən</td>
</tr>
<tr>
<td></td>
<td>WT fp face, look</td>
<td>WT fp head, bank</td>
<td>WT fp head, bank</td>
<td>WT fp head, bank</td>
<td>WT fp head, bank</td>
</tr>
<tr>
<td>ʔŋə</td>
<td>ʔŋə</td>
<td>ʔŋə</td>
<td>ʔŋə</td>
<td>ʔŋə</td>
<td>ʔŋə</td>
</tr>
<tr>
<td></td>
<td>WT fp face, look</td>
<td>WT fp head, bank</td>
<td>WT fp head, bank</td>
<td>WT fp head, bank</td>
<td>WT fp head, bank</td>
</tr>
<tr>
<td>ʔŋən</td>
<td>ʔŋən</td>
<td>ʔŋən</td>
<td>ʔŋən</td>
<td>ʔŋən</td>
<td>ʔŋən</td>
</tr>
<tr>
<td></td>
<td>WT fp face, look</td>
<td>WT fp head, bank</td>
<td>WT fp head, bank</td>
<td>WT fp head, bank</td>
<td>WT fp head, bank</td>
</tr>
</tbody>
</table>

Notes for Table Y-1:
Two parallel stems comprise this wf, one with, one without ST / OC medial */r* (§2.5).
Open vowel and final *r* forms are inherited from ST.
The nouns in final -n are CH innovations (§6.4.3). The form yâ ʔθən 'face' is ambiguous because it could either be ʔθən + n, or be directly related to WT *par as TB final */-r* can on occasion correspond to a MC div. II syllable with final n (§7.7.2). The forms in final -k are CH innovations (§6.1). 'Outside' → wàl ʔθə does not belong to these stems.
### Fuzzy Word Identification: A Case Study from the Oracle Bone Inscriptions


**Appendix 3: Schuessler 2007:193**

<table>
<thead>
<tr>
<th>Chinese characters</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>
<th>Chinese</th>
<th>Pinyin</th>
</tr>
</thead>
</table>
| chū 楚 | (tsjhu) | LIH t'ɕuŋ, OCM *tshra? | ‘Thorny bush / tree’ | [Shi]  
| [T] Shu Siku Chu SR t’ɕu (писыва); MGZY chu (писыва); CDC chu(писыва); ONW tɕʰo > tɕʰa  
| 题 AA: PMonic *tʃla? ‘thorn, thorny bamboo’ (added to names of thorny plants), Khmu ‘corlaʔ’, Semai ‘jalraʔ’ | [Diffloth 1984: 80]. The complex AA initial needed of course to be simplified in CH, apparently by elimination of the medial l. |
| **Table C-1 ‘Knock, push, touch’ (for → chū 触)** | **Table C-1 ‘Knock, push, touch’ (for → chū 触)** | **Table C-1 ‘Knock, push, touch’ (for → chū 触)** | **Table C-1 ‘Knock, push, touch’ (for → chū 触)** | **Table C-1 ‘Knock, push, touch’ (for → chū 触)** | **Table C-1 ‘Knock, push, touch’ (for → chū 触)** | **Table C-1 ‘Knock, push, touch’ (for → chū 触)** | **Table C-1 ‘Knock, push, touch’ (for → chū 触)** | **Table C-1 ‘Knock, push, touch’ (for → chū 触)** | **Table C-1 ‘Knock, push, touch’ (for → chū 触)** | **Table C-1 ‘Knock, push, touch’ (for → chū 触)** | **Table C-1 ‘Knock, push, touch’ (for → chū 触)** |
| **knock against** | ... | WB tuiq strike against WT *pθag-pa  
| Lush. *tξk (against) | WT *pθag strike against | WB tuiq-pa strike against | WB cʰorw thrust, push, butt, stamp KNaqj *tξjak strike against |
| **strike** | Mikir tōk to strike | dōq 腹, beat, pound | WB *pθag strike against | zhōu 噣 *tξk strike zhōu 噣 *tξk strike, stamp earth | zhōu 噣 *tξk strike zhōu 噣 *tξk strike (with rod), pound, stamp Mikir chōk bēn WT cʰorw thrust, push, butt, stamp |
| **stamp, pound** | ... | ... | ... | ... | ... | ... | ... | ... |
| **cut, hack** | zhuó 别 *tξ documentary | to cut, hack | WT *pθag strike against | zhōu 噣 *tξk cut off Lushai tukL cut, chop zhōu 噣 *tξk cut off, kill, die WT *pθag strike against | zhōu 噣 *tξk cut off, kill, die WT *pθag strike against |
| **peck, beak** | zhōu 噣 *tξ documentary | to peck | ... | zhōu 噣 *tξk cut off, kill, die WT *pθag strike against | zhōu 噣 *tξk cut off, kill, die WT *pθag strike against |

Comment on the table: The distribution of meanings over the different stems appears random, yet many stems tend (1) to have a semantic focus: 

**ST *tξk** ‘knock against’ → chū 触

**PTB *tξ** ‘beat, chop’: Mikir chōk ‘beat’, WT *pθag-pa ‘hew, chop’

**ST *tξ** ‘strike against’ → zhōu 噣

**PTB *tξag** ‘strike’ → zhōu 噣

**ST *tξ** ‘push, beat’ → dōq 腹

**ST *dō** ‘cut’ → zhōu 噣

**ST *tξk** ‘cut’ → zhōu 噣

**ST *tξ** ‘cut’ → zhōu 噣

**ST *tξ** ‘cut’ → zhōu 噣

**PTB *tξ/ja(k)** ‘strike’. Lushai chul ‘beer’ → ja(k) ‘strike, pound, stamp’

> *peck, beak* Lushai chul ‘beer’ → ja(k) ‘strike, pound, stamp’

> *peck, beak*: Lushai chul ‘beer’ → ja(k) ‘strike, pound, stamp’ SER: peck, beak’ WT meču ‘lips, beak’

23
References


Jao Tsung-I (Ráo Zōngyí) 饒宗頤. 1959. *Yǐndài Zhēnbǔ Rénwù Tōngkǎi* 殷代貞卜人物通考 [Oracle-Bone Diviners of the Yin Dynasty]. Hong Kong: Hong Kong University Press.


Yú Xǐngwú 于省吾. 1941/1950. *Shuangjianchǐ Yinqi Piánzhì Xiubian* 雙劍誼殷契駢枝續編 [Humble Comments on the Oracle-bone Inscriptions from the Double-Sword Studio, vol. 2]. Taipei: Yiwen Yinshuguan. [This is apparently a reprint of a 1941 original by Beijing Zhonghua Shuju; the Taipei edition has a different title on the cover, 殷契駢枝續編, although the full title appears inside.]

早期出土文獻裏往往會有一些字無法準確隸定為任何古漢語已存在的字詞。隨著早期漢語詞族與詞法結構研究越來越深入，這些字雖然無法精確判別，卻能知悉其大概的字義和字音，因此本文提出以「模糊辨識」的新方法為這些字下定義。以甲骨文的「𠞹」字為例，這個字出現於甲骨文兩個日期之間的時候，或被認為代表「斲」字，或被認為代表「剅」字，還有其他不同的看法，但這些結論恐怕都過度精確。利用模糊辨識的方法，可以將「𠞹」字限定為 *tok（斲）詞族，而無法進一步精確地隸定於任何後期存在的字詞之下。這做法既可以得出一個不超出目前已有證據能支持範圍的結論，又同時能為此字的字音、字義及其功能提供比較可靠的說明。

關鍵詞：甲骨文、模糊辨認、詞族、月食