
Article:
Tranter, N. (2013) Logography and layering: a functional cross-linguistic analysis. Written Language and Literacy, 16 (1). pp. 1-31. ISSN 1387-6732

https://doi.org/10.1075/wll.16.1.01tra

Logography and layering

A functional cross-linguistic analysis

Nicolas Tranter
University of Sheffield, United Kingdom

This paper proposes a way in which the semantographic/phonographic dichotomy recognised as fundamental in logographic (or morphosyllabic) writing systems in East Asia, the ancient Middle East, and Mesoamerica, can be systematised to transcend the very different scholarly traditions in each region in order to allow valid and more meaningful cross-linguistic comparisons. A totally functional analysis with a focus on synchronic words as they occur in texts, rather than a focus on the form of signs or their etymology, ignores such formal units as the frame or even the grapheme and recognises three main compositional levels – logogram, component, and element – and the strict application of the analysis reveals cases of a fourth level, superlogogram. The application of this approach allows characterisations of writing systems that reflect the meaningful combination of signs in context and reveals greater complexity in how words are written, such as in semantic+semantic combinations, than previous analyses have recognised. It is concluded, however, that a statistical application of the analysis is prevented, not because of differences in the writing systems, but primarily because of the fundamental typological differences of the represented languages themselves.

Keywords: Logographic writing systems, Egyptian hieroglyphs, Sumerian, Hittite, Chinese characters, Japanese, Vietnamese chữ nôm, Zhuang sawndip, Tangut, Mayan

1. Preliminaries

1.1 Languages and writing systems

This paper aims to establish the principles for a framework that would allow a detailed functional analysis of, and comparison between, ‘logographic’ writing systems. Although all such systems are accepted to be based on a dichotomy of semantographic

and phonographic signs, there has to date been little attempt to establish cross-linguistic analytical principles; such a framework is essential. This paper proposes the concept of 'layering' (Section 2), which it is argued is crucial to identifying component signs and allowing cross-linguistic comparison, and proceeds to identify four basic types of component (Section 3). The strict analysis of this framework can be very revealing, requiring us to revisit what exactly is a phonogram or a semantogram (Section 4.1) and identifying a 'superlogogram' layer (Section 4.2).

There exist or have existed a wide variety of writing systems that have been characterised variously as 'logographic', 'morphosyllabic', 'morphosyllabographic', 'logosyllabic', 'ideographic', 'morphographic', 'logophonetic', etc. The common feature of these writing systems is that a certain proportion of signs are semantographic rather than phonographic, indicating the meaning of the encoded word rather than its phonology; all such writing systems also make use of phonographic signs. The languages that have been encoded in such systems include:[1] (1.) Mesopotamia: Sumerian, Akkadian and its later Assyrian and Babylonian relatives (Labat 1976*; Borger 2004*), Hittite (Rüster & Neu 1989*; Gragg 1996), Hurrian, Elamite (Gragg 1996), Luwian (Melchert 1996), etc.; (2.) Egypt: Egyptian (Gardiner 1957; Faulkner 1999 [1962]*; Bonnamy & Sadek 2010*); (3.) Mesoamerica: Mayan (Montgomery 2002*; Macri & Looper 2003*) – or rather two Mayan language groups (Ch'olan and Yukatecan) –, the 'pre-proto-Zoquean' language of the epi-Olmec script (Justeson & Kaufman 1993), and probably Zapotec; (4.) East Asia: Chinese (or, rather, different historical and synchronic varieties of a Chinese language family), other languages such as Japanese that have adapted Chinese writing (see below), Jurchen (Jin 1984*; Kara 1996), Khitan (Kara 1996), Yi (Yunnan Sheng Lunan Yizu Zizhixian Wenshi Yanjiushi 1984*; Ding 1993; Shi 1996: 239-240, 242), and Tangut (Nishida 1964-1966*; 1980; Li 2008*). We may add to these the particular development of Chinese characters into the heavily phonographic nüshu (Chiang 1995) system used till recently by women in certain southern Chinese communities. In Europe we also find early Greek in Crete's Linear B (Bennett 1996).

There have also been proto-writing systems or non-linguistic semasiographic systems that nevertheless use similar principles, notably the rebus principle that

[1] Works with an asterisk in Section 1.1 are sign-based dictionaries, or more accurately signlists with varying degrees of lexicographic information, that I have used extensively in this research.

underlies the phonographic part of logographic writing, such as Aztec and Mixtec pictography (Boone 2000: 32, 35-38, 51; Fedorova 2009) in Mesoamerica, or Naxi so-called ‘Dongba script’ in East Asia (He & Jiang 1985: 17-126; Ramsey 1987: 266-270), and even the iconography in Mayan or Egyptian pictorial traditions display features found in writing. The fact that these do not clearly correspond to specific sentences means that they have been excluded from the present study.

We may group many of these writing systems into ‘writing families’. For example, we may talk of the Sumerian writing system, the Akkadian writing system, and the Hittite writing system as separate entities, each with its own internal structure and principles, i.e. *systems*, whereas all three belong to a common writing family because Sumerian writing was adopted and adapted by the other two languages. The largest writing family is undoubtedly that based on Chinese characters, which were developed to write Chinese but were adopted and adapted, in varying extents together with the creation of non-Chinese ‘Chinese characters’, to write Korean (see Nam (2012) regarding their adaptation to Old Korean idu, hyangch’al, and kugyŏl systems; Sasse (1980) regarding made-in-Korea characters; and Hannas (1997: 65-72) regarding postwar use), Japanese (see Seeley (1991: 16-58) regarding their adaptation to Old Japanese; Reiman (1983) regarding made-in-Japan characters; and Backhouse (1984), Smith (1996) and Tranter (2008) regarding modern character use, both standard and non-standard), Vietnamese – in which it is known as chữ nôm (Takeuchi 1988*; Hannas 1997: 78-84; Nam 2001: 85-99), Zhuang – in which it is known as sawndip (Wei 1980: 97; Ramsey 1987: 242-243; Su 1989*; Wei 1989; Zhang 1990), Bai (Xu & Zhao 1984: 128-133; 1990), Kam (Liang 1980: 89), Tay-Nung (Doan 1996: 87-95), etc. The dictionaries of chữ nôm and sawndip contain vast numbers of non-Chinese characters, as well as widespread rebus use of authentically Chinese characters. Explicitly written Cantonese has also developed large numbers of Cantonese-specific characters (Bauer 1988: 260-266); for example, of the 61 characters of Bauer’s (1988: 258) representative extract of written Cantonese, five characters = eight tokens are formally Cantonese-specific, two are rebus-based phonograms, and one is a roman letter. Stretching further the terminology of historical linguistics, we also observe various writing ‘isolates’ such as the Tangut writing system.

All these writing systems have at their core a simple dichotomy of phonography and semantography. Respective texts represent a combination of phonograms and semantograms, varying between very low amounts of phonography in Tangut (at least

in writing native words) and minimal semantography in modern Korean writing (which in most texts is now written entirely in the phonographic han'gŭl alphasyllabary, without Chinese characters), or even fully phonographic adaptations of another variety's less phonographic writing, e.g. in Sumerian 'unorthographic' texts or those in the Emesal variety (Thomsen 1984: 281-284), or in the Old Japanese poetry in *Kojiki* (711/712) and *Nihon Shoki* (720/721) (Seeley 1991: 45, 48). This basic dichotomy is not at issue, and forms the basis of all presentations of such writing systems. Also not at issue is the typical existence of 'multivalence' or 'polyvalence' of individual signs within such systems (Boltz 1994: 73-126), where a given sign may have more than one possible value. Within context the multivalence is normally resolved by the users of the writing system, either by the linguistic context or by the graphic context of sign collocations.

Each of these writing systems has its own scholarly tradition. It is not uncommon in introducing one writing system to make comparisons with a writing system somewhere else in the world, yet there have been remarkably few detailed cross-linguistic comparisons even between the long-studied Sumerian, Egyptian and Chinese systems and these tend to treat each system in a separate section or chapter, e.g. Gelb (1963), Sampson (1985), Coulmas (1989), DeFrancis (1989). Boltz (1994: 24-28, 55-59, 75-90, 102-103) does dedicate considerable space to looking at Egyptian and Sumerian even though his is a study of Chinese writing and his analysis of Chinese writing is carried out within this comparative framework. Nevertheless, most of what is written, including the above works, focus either on the development and evolution of logographic writing rather than its synchronic use, or on the classification of these systems within a framework dealing with all types of writing. In addition, though Mayanists occasionally refer to other writing systems (e.g. Mora-Marín 2008), Mayan has tended not to be included in truly comparative works because breakthroughs in its decipherment have been relatively recent and are ongoing. For example, Gelb (1963) predates decipherment and, though decipherment had begun, the problems in availability of much of the research results in mere passing references in Sampson (1985) and Coulmas (1989), though DeFrancis (1989: 121-128) notably does tackle it. Similarly, many East Asian writing systems are also not included largely because of a lack of scholarly research published in western languages. For example, Tangut writing, which has been the subject of substantial Japanese and Chinese (and earlier Russian) research and whose decipherment has been helped by the fact that much of what its users left to

posterity was substantial lexicographic works and translations from other languages, is really only found in English in the outdated Grinstead (1972) or the briefest descriptions by Kychanov (1996: 228-230) or Coulmas (1996). Cooper (2004: 92-93) argues that comparison between what he calls 'logophonetic' systems "will often help us to understand a puzzling feature in one system by identifying a better-understood analogue in another, and the differences in writing systems that comparison reveals enable a better understanding of the individual systems in their own right". Though his argument specifically concerns the development of such systems in their earliest stages, it is as valid to the synchronic analysis of their structures. Such a cross-linguistic comparison is proposed in this paper.

1.2 Form and function

Because writing is a visual medium, there is understandably great interest in its form, ranging from the writing technology to the calligraphic style to the formal organisation of signs relative to each other. The latter includes direction of writing, whether signs are organised purely linearly or in groups or 'conflated' one on top of or inside another, the identification of the grapheme vs. the etic needs of epigraphic study, and the concept of the frame.

Frame is a useful term espoused by DeFrancis (1989). It refers to the division of a text into smaller visually identifiable blocks, which roughly parallel the use of word-division in modern languages written in the Roman script in that they constitute a visual break in the writing that may facilitate reading. DeFrancis (1989: 115) defines the frame as "the smallest segment of writing conventionally receiving special status, such as being surrounded by white space and listed in dictionaries", for example the written word in Roman alphabet writing or the individual Chinese character within its idealised square, and he treats frame as the same as the 'lexeme'. Here, however, I separate frame from lexeme, and use frame slightly differently to refer only to the idealised square or oblong that is used to organise the signs of certain writing systems. I also extend it beyond DeFrancis's application to East Asian scripts to any idealised space that units of writing fill.

The frame is not universal; for instance, later forms of Sumerian-derived writing, such as the Neo-Assyrian style, have no identifiable frames. However, Chinese (and

East Asian systems generally), Egyptian, Mayan, and older Sumerian and Akkadian writing all make use of frames.

In standard written Mandarin Chinese (hereafter 'Chinese'), for example, each character (as conventionally defined) occupies the area of an imagined square shape. The components that make up a single character occupy and divide up this shape, which in print is always the same size (and in handwriting is as much the same size as is possible), regardless of how simple or complex the character is. Examples of Chinese in this paper show different degrees of complexity, but all occupy the same idealised square. The Chinese language typically consists of monosyllabic morphemes. The relationship of writing to language can be characterised roughly as 1 character = 1 morpheme = 1 syllable.[2] The organisational principle of the frame in Chinese is so strong that all other East Asian writing systems that have been influenced by Chinese – whether they adapt Chinese characters or are new systems – make strict use of an idealised square. The phonological scripts that developed in Japan (hiragana and katakana, collectively kana) and in Korea (han'gŭl) strictly continue to use the same idealised square that Chinese characters use. In the case of han'gŭl, this involves combining the individual letters of what is essentially an alphabet into blocks, each corresponding to a syllable and occupying an imagined square space (Taylor & Taylor 1995: 216-220). Khitan lesser script earlier also combined its phonograms (syllabograms plus apparently consonant signs) into blocks (Chingertai et al. 1985: 147-149; Liu & Yu 1990: 247-248; Kara 1996: 230), although in this case the blocks corresponded to words rather than syllables. Mayan writing also uses such idealised spaces, typically filled with one 'main' sign and often with one or more flatter 'affix' signs attached, the overall effect of a frame being characterised as 'pebble-like'. Egyptian writing of all periods tended to group signs not just linearly but in small stacks, allowing tall-thin signs to be written side-by-side in vertical writing, for long-thin signs to be written on top of each other in horizontal writing, and for small signs to be fitted wherever they would fit. It is notable that despite such stacking, Egyptian normally does not allow a sign at the end of one content word and a sign at the start of the next to be

[2] The characterisation of Chinese as an isolating language with monosyllabic morphemes is broadly true, though there are a few polysyllabic native morphemes, polysyllabic (and phonographically written) loanwords, and a subsyllabic morpheme *-r* (Norman 1988: 154). Similar statements can be made of other isolating languages of the region, such as Vietnamese (Thompson 1965: 106).

stacked in this way; rather, each word has its own idealised oblong which the various signs fill appropriately, without overlap between oblongs. The only apparent exceptions in orthodox writing are (pro)clitics, including *n* 'of; to', which are typically written within the neighbouring word's frame. The role of the frame is more haphazard in earlier Sumerian and Akkadian writing in which the clay medium is scored into boxes known as 'cases' or 'registers'. These have an organisational role because the signs that make up a word are not normally split between such cases (and in older texts may not even be read linearly within a case), but the linguistic level that each 'case' corresponds to varies widely.

However important the frame is in the organisation of signs in many writing systems, it is not universal and, crucially, it is a unit of the formal analysis of writing. To understand writing systems and to compare them, we need to conduct a functional analysis.

There have been a range of problems hindering meaningful cross-linguistic and cross-regional comparison. Firstly, even in the East Asian tradition, there are few scholars who admit to be able to cope with all the four major languages and writing systems (Chinese, Korean, Japanese, Vietnamese; for example, the cover blurb of Hannas (1997) presents the author's ability in the four as unique). Even fewer scholars can be expected to have an ability in Sumerian, Hittite, Egyptian, Mayan, and Tangut. As Houston, Baines & Cooper (2003: 431) rightly point out, specifically in connection with the comparative study of the terminal phases of writing systems, "No one person can command all relevant information", adding that "Interpretations change rapidly, and publication typically lags behind the outer envelope of knowledge, especially for the evolving understanding of New World thinking". One can add to this that for certain writing systems, such as Tangut as mentioned above, most research is not written in European languages. It is with trepidation that I propose a framework for such a cross-linguistic approach.

Secondly, a significant hindrance to the scholar who does try to broaden their study of writing systems is the fact that the various traditions have very different approaches to transcription/transliteration. On the one hand, Sumerian, Akkadian, and Hittite are mostly transliterated in graphical-phonological systems, that reproduce each sign as its unique transcription, with homophones distinguished by subscript numbers and no morphological analysis. On the other hand, Egyptian and East Asian languages are largely transcribed in phonological-morphological systems, in which the phonology is

transcribed and there is morphological analysis. A third type of system, a purely graphical one, removes phonology from the transliteration and takes the form of numbers, each uniquely corresponding to a specific sign. Egyptian transliteration uses A-Z + a number derived from Gardiner (1927) at a more epigraphic stage of analysis. Because decipherment is still ongoing, Mayan is widely transliterated by 'T-numbers' based on Thompson (1962).[3]

Thirdly, individual scholarly traditions may have different analytical frameworks. In particular, in the case of the Chinese writing family we also have the burden of the six-way *liushu* categorisation of all Chinese characters dating from the early second-century *Shuowen Jiezi* dictionary (Wieger 1927: 10-11; DeFrancis 1989: 96-98; Coulmas 2003: 50-53). Widely used still in East Asia, the *liushu* are problematic because they are etymological rather than synchronic, the 'huiyi' (semantic+semantic) category is only justifiable synchronically because all such characters appear etymologically to be really semantic+phonetic combinations, two of the categories (the rebus 'jiajie' and the problematic 'zhuanzhu') constitute adaptations of characters in one of the other four, and scholars differ on even how to interpret the very small 'zhuanzhu' category.

Fourthly, the concept of what constitutes a grapheme in a writing system differs between scholarly traditions. 'Grapheme' is used to refer to fundamentally and widely different levels of analysis, ranging from the smallest discrete sign in the writing system (e.g. in Egyptian or Sumerian) to the complete writing of a phonological word or discrete morpheme (e.g. in Chinese or Tangut), to the purely formal unit of the frame (e.g. in Japanese). It is noticeable that in the case of words written with a combination of signs, the intermediate level between the smallest discrete sign and the complete writing of a phonological word is not a level that is treated as a grapheme in any

[3] In the current paper, linguistic data are presented in phonological-morphological transcription and italicised, except when explicitly citing conventional transcriptions. This means that the graphical-phonological approach in, for example, Mesopotamian studies is not followed. In Section 2.2 *buranun* 'Euphrates' corresponds to Sumerological ^{íd}buranun or ^{íd}BAR$_6$.KIB.NUN (where ^{íd} represents the unpronounced semantic determinative and capitals cite individual signs that together correspond to a single word) and *šiuni-š* corresponds to DINGIR.*LIM.iš* in Hittite studies. Mesopotamianist use of ´, ` and subscript numbers to distinguish homophones is, however, retained. In terms of phonology, transcriptions all follow the commonest conventions. Middle Chinese follows the Baxter (1992) system.

scholarly tradition.[4] Daniels (1991) in connection with writing systems generally argues that ‘grapheme’ is a term with no coherent definition and in case does not deserve the -eme suffix. I find ‘grapheme’ in any case not to be a useful term in the analysis of logographic writing systems, and do not use it.

2. Layers and layering

2.1 Reclaiming ‘logogram’

As shown above, terminology is problematic, in part because of differing regional scholarly traditions. Moreover, there is no fully accepted term that can characterise such writing systems in contrast with the ‘phonography’ of alphabetical, abjad, and syllabary systems (see the seven terms in the opening sentence of 1.1), even less an accepted noun that can be used to refer to one or more signs that encode a word into writing.

Some terms are inappropriate because they focus on visual form or etymology (‘pictographic’), or they wrongly imply that the writing only encodes meanings or ideas rather than items of spoken language (‘ideographic’). The term ‘logographic’ at least recognises that the writing encodes words, but because it has been used to contrast with ‘phonographic’ it can imply a lack of a phonographic component within ‘logographic’ writing, but also, as DeFrancis (1989: 115; 223) argues, it does not reflect the differences between the Sumerian and Chinese systems in terms of what is a grapheme and the role of the frame. However, I consider both to be formal features and my interest is in the functional structure of the system rather than in the definition of the grapheme.

As noted above, all such systems have both semantic and phonetic elements. So DeFrancis (1989: 115-116) champions the term ‘morphosyllabic’, which suggests both morphemes and phonological units and reinforces the fact that it encodes language. The term has become fairly widespread in recent literature, but there is still a lack of a

[4] DeFrancis (1989: 115) argues that, though Chinese characters are predominantly semantic+phonetic combinations (the *liushu* ‘xingsheng’ category), the character itself is not the grapheme but just the frame and that the phonetic half of such characters is strictly the grapheme. However, he regards the semantic half as a secondary unit as indispensable as the frame is and thus does not treat it explicitly as a grapheme.

corresponding noun; ‘morphosyllabogram’ is unwieldy and it is notable that DeFrancis himself avoids such a noun, preferring to use words like ‘character’, etc. Other authors continue to use ‘logographic’, such as Coulmas (2003: 41, 59-60) with qualifications because it implies ‘inaccurately’ that the word is the prominent unit of writing. Moreover, ‘logogram’ has remained a standard term among Mayanists, though this is used in contrast with ‘syllabogram’ (or ‘phonogram’).

I would argue that ‘morphosyllabic’ can be as misleading as ‘logographic’, as it suggests the phonograms are all syllabic and correspond to morphemes, as in the typologically isolating languages of most of East Asia that have predominantly monosyllabic morphemes. Japanese and Korean, and all non-East Asian languages that have used such writing systems, are in contrast polysyllabic and inflecting/agglutinating and neither a morpheme nor a syllable necessarily corresponds to a unit of writing. Because of terminological problems, I propose to reclaim ‘logogram’, defined functionally as one or more written signs that represent together a phonological word. Where I need to contrast the functions of signs within a logogram, I use ‘phonogram’ vs. ‘semantogram’; the Mayanist use of ‘logogram’ corresponds to my semantogram.

2.2 Layering

The approach outlined here holds that the analysis of the writing systems under consideration must recognise the (potential) existence of more than one layer of composition, and posits three fundamental units corresponding to different layers: ‘logogram’, ‘component’, and ‘element’. As the term ‘grapheme’ tends to be defined in formal rather than functional terms and corresponds to very different linguistic levels in different traditions, it is deliberately avoided hereafter.

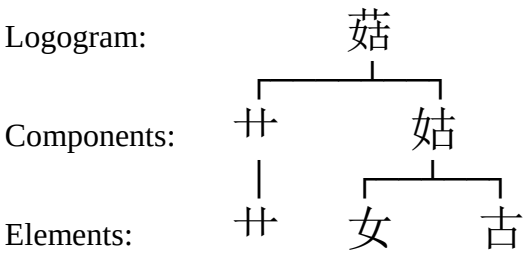
The basic assumption is that logograms are created from components and components are in turn created from elements. The element is strictly defined as the smallest unit of writing that carries any value within the writing system as a whole that relates directly to the value of a component of a logogram, but is not a component in itself. The component, on the other hand, is defined as the smallest unit of writing the value of which (in terms of semantic or phonetic information) relates directly to the value of the logogram. The logogram, in turn, is defined as the smallest complete unit of writing that corresponds to a unit of meaning in the spoken language, typically a word

in a very loose sense. Consider, for example, the Chinese *gū* ‘paternal aunt; one’s husband’s female relative’ and its representation in writing 姑. The character is made up of two components, one 女 semantic (WOMAN), the other 古 phonetic (*gu*). Both components also happen to occur as independent logograms (女 *nǚ* ‘woman’, 古 *gǔ* ‘old (vs. new)’), from which their component values are derived.

Our analysis, however, does not require components to occur as independent logograms or even to be derived originally from logograms (which would be to impose a diachronic constraint to our analysis), such as in the case of the PLANT component below or the Tangut components INSECT or WATER illustrated later. All that is required is that the value that we assign to them occurs synchronically as a component in more than one logogram. In Chinese character script, the WOMAN component is attested in a large number of characters representing morphemes involving women or activities involving women. It is from its recurrence in these characters that we identify it synchronically as WOMAN rather than from its occurrence as an independent logogram writing ‘woman’. On the other hand, an independent character used only in one case as a component of another would also constitute the minimum requirement if a semantic connection is clear.

The character used to write *gū* ‘mushroom’, 菇, on the other hand, adds an additional semantic component PLANT 艹 to 姑. 菇 now consists of three elements, but, though the topmost element can be identified as being clearly semantic (PLANT) – in this case a frequent component that has no independent existence as an independent character – and the lower right element 古 can be interpreted as contributing phonetic value *gu*, the lower left element 女 appears to contribute nothing: neither its meaning (‘woman’) nor reading (*nǚ*) as an independent logogram has any relation to ‘mushroom’ or to *gū*. Although it will be argued later that valueless components do have an important synchronic role in such writing systems, the bottom left element in 菇 is better interpreted as not being a component in its own right. This accords well with a fundamental characterisation of Chinese and Chinese-derived writing systems as

following a ‘two-component maximum principle’. A layered analysis of 菇 is that it consists of just two components <PLANT + *gu*>, with the *gu* value being provided not by 古 but by 姑, which we may represent diagrammatically as:[5]



In other words, the 女 WOMAN component and the 古 *gu* component of 姑 ‘aunt’ do not contribute directly to the value of 菇 ‘mushroom’, and therefore are not components of the latter. In strict analysis, we shall henceforth clearly differentiate between element and component, the former having a direct relationship only to the component, the latter a direct relationship only to the logogram. The hierarchical relationship between the three layers may be characterised as: logogram = one or more components; component = one or more elements. The difference between element and component is crucial for two reasons: the role of layering in the creation of logograms is more meaningful if we make the distinction; and it allows meaningful comparison between writing systems with different types of phonographic representation. In the case of Chinese writing, the strict ‘two-component maximum principle’ applies (with the possible exception of a small number of logograms created from the tripling of a

[5] This paper uses the following conventions. <…> encloses a Logogram and components are linked by + if sequential (not necessarily linear) and × if enclosed, conflated, or superimposed. The +/× distinction is actually formal rather than functional, but is applied here solely for clarity in what is a wide range of writing systems. Though elements are excluded from the analysis, where it is useful to identify which elements constitute a single component they are enclosed in […]. Phonetic components are italicised, with vowels that are consistently to be ignored in *[…]*. Semantic components are in inverted commas, unless they function clearly as semantic determinatives, in which case they are in small capitals. A slash is used to separate multiple values, either where there is more than one possible analysis or where a component is simultaneously phonetic and semantic (**Ψ**: see Section 3).

single component), and the role of layering in the derivation of such as 菇 is broadly taken for granted by East Asianists. Nevertheless, studies classifying the phonetic components of Chinese characters tend to overlook layering. Wieger (1927: 427), for example, treats 古 as his ‘phonetic’ number 132 subsuming both 姑 and 菇 under it. This appears to be more for convenience, because two more prolific ‘phonetics’ based on 古 do have separate entries as numbers 368 and 450. DeFrancis (1984: 89-115) deals at length with the question of identifying Chinese characters’ ‘phonetics’ and to what extent they may constitute a very imperfect ‘syllabary’. It is clear that those earlier authors who have not ignored the phonetic component have nevertheless underestimated the number of such ‘phonetics’ in Chinese.

In other writing systems, however, no such two-component restriction applies. Egyptian writing, for instance, often uses more than two components. The tendency in the analysis of these writing systems is to assume that each element has equal status. However, layering occurs not only in Chinese but also in Sumerian, Tangut, and Mayan.[6] Chinese 認 *rèn* ‘to recognise’, Sumerian  *buranun* ‘Euphrates’ (L579.484.381.228.087; B839.756.596.378.143), Tangut  *śiwe*^{1} ‘honey’ (Li2416; N181-099), and Mayan  *chumib’* ‘sitting place’ (T644b[585]) are each made up of two components: 言 + 忍 <COMMUNICATION + *rěn*> = Chinese 認 *rèn* ‘to recognise’,  +  <RIVER + *buranun*> = Sumerian  *buranun* ‘Euphrates’,  +  <‘sweet (liquid)’ + INSECT> = Tangut  *śiwe*^{1} ‘honey’,

[6] Mayan symbols are identified where appropriate as T-numbers with a prefixed ‘T’ as is conventional, based on Thompson (1962). Extending this principle, cuneiform symbols are occasionally identified as L- and B-numbers based on Labat (1976) and Borger (2004), and Tangut by Li- and N-numbers based on Li (2008) and Nishida (1964-1966). ^{1} and ^{2} after Tangut transcriptions indicate the two tonemes of the language.

and × <*chum*/‘to sit’ × *b’[i]*> = Mayan *chumib’* ‘sitting place’. One of the components in each example above, namely Chinese 忍, Sumerian 𒀀𒇉, Tangut, and Mayan not only occurs as a logogram in its own right but is in turn also made up of two components: 刃 + 心 <*rèn* + FEELING> = Chinese 忍 *rěn* ‘to endure’, 𒀀 + 𒇉 <‘water’ + ‘spring’> = Sumerian 𒀀𒇉 *íd* ‘river’, + <WATER + ‘sweet’> = Tangut *thjwi*[1] ‘sweet (liquid)’, × <*chum*/‘to sit’ × *m[u]*> = Mayan *chum* ‘to sit’.[7]

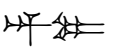
Layering also occurs in Egyptian, as the writing of *kmt* ‘Egypt’ illustrates. The two normal spellings are 𓆎𓅓𓏏𓊖 and 𓆎𓅓𓏏𓊖𓏏𓏤. The former combines the three-element phonetic component *kmt* 𓆎𓅓𓏏 with a single sign with the value SETTLEMENT 𓊖, used in names of, and words relating to, settlements: <*kmt* + SETTLEMENT> = *kmt* ‘Egypt’. A linear analysis of 𓆎𓅓𓏏𓊖𓏏𓏤 by elements (i.e. by individual hieroglyphs), however, leaves two extra signs unaccounted for, with the consonant letter 𓏏 *t* apparently written twice even though it is pronounced only once, and the addition of a short vertical line 𓏤 that has neither semantic nor phonetic value but normally only occurs in logograms that begin with a semantic component. A layering approach treats 𓆎𓅓𓏏𓊖𓏏𓏤 as a combination of 𓆎𓅓𓏏 with 𓊖𓏏𓏤, the normal spelling in the writing system of the word *niwt* ‘town’ (which as a logogram in its own right is analysed

[7] Any study of Sumerian, Akkadian, and Hittite forms has to decide on a ‘font’ that is language- and period-neutral because of the vast changes in the writing system over the long period of use. Following the recommendation of an anonymous Mesopotamianist reviewer, I use an Old Babylonian font, except for Sumerian *g̃anun* ‘granary’ in Section 4.1 where an Ur-III font is used. Mayan forms are also presented stylised.

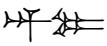
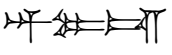
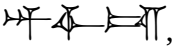
<‘settlement’ + *t* + **X**>, a spelling that does conform to the principles of Egyptian writing in its use of the vertical line). In this case,  functions as a single component in  , with the same value SETTLEMENT as  does in  .

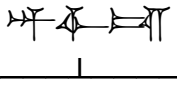
Though logographic writing systems encode language, a system may allow a single word to be represented in a variety of ways. For example, the same Mayan word ‘to sit’ is written in two ways above,  and  , the latter resulting from conflation with an optional phonetic complement *m[u]*, while the Egyptian word *kmt* ‘Egypt’ is written in two ways, with different ways of expressing the semantic component.

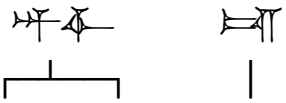
Because of issues of decipherment, especially in Mesoamerica, it is easier to observe layering in writing systems that still exist or are entirely deciphered, and, in the case of Mesopotamia, in the way that Sumerian writing was adapted to other languages, showing a reinterpretation of signs and layers in Akkadian, and then again in Hittite. As the three languages are genetically unrelated and signs were borrowed to write semantically similar but etymologically totally unrelated words, phonographic components in Sumerian no longer function this way in Akkadian, Akkadian in turn adds further phonographic components, and these in turn no longer function this way in Hittite. Consider, for example, the totally unrelated words for ‘god’ in Sumerian (*diĝir*), Akkadian (*il*-) and Hittite (*šiuni*-/*šiwani*-).

 writes *diĝir* ‘god’ in Sumerian. Akkadian can also use  , but as well as grammatical number the language has case (nominative/accusative/genitive) and state (absolute/construct) inflections, so these may be expressed by phonetic complements, i.e. *il-um* ‘god’-NOM written as  <‘god’ + *lum*> and *il-im* ‘god’-GEN as  <‘god’ + *lim*>.[8] Hittite in turn can write its own word for ‘god’ with the Akkadian spellings, but the Akkadian *lum* and *lim* components no longer have any

[8] The standard morphological convention of using a hyphen is introduced here, together with the following morphological tags: NOM = nominative, GEN = genitive, DAT = dative, IMP = imperative. I apologise to Mesopotamianists for the clash between different disciplines’ conventions, given that a hyphen in Mesopotamianist convention can indicate a glottal stop in Akkadian.

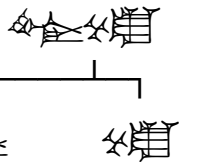
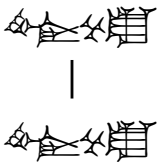
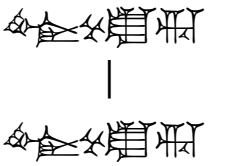
phonographic value in writing the Hittite word; both  and  are used indiscriminately to write, for example, the Hittite nominative *šiuni-š* ‘god’-NOM. Hittite in turn typically adds its own phonographic component, writing *šiuni-š* as  or  with the addition of the phonogram  *iš* to  or  . If we take Hittite  , a layering approach gives us the following, with only two components <‘god’ + *iš*>.

Logogram: 

Components: 

Elements: 

A more extreme example is the word for ‘horse’:

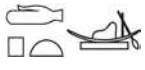
	Sumerian *anše kur-a* ‘horse’	Akkadian *sisû* ‘horse’(-NOM)	Hittite ?*r-i* ‘horse’-DAT
Logograms:			
Components:			

In Sumerian, there are two independent words *anše* ‘donkey’ and *kur-a* ‘of the hills’, which would be analysed as two logograms.  *anše* cannot be broken into further components, but  *kur-a* can be analysed as <‘hill’ + *ra*>. When the word is borrowed to write Akkadian *sisû*, which is a single simple word, the original phonogram

 *ra* no longer contributes any value to what for Akkadian must now be treated as a single logogram. It is this single logogram that is borrowed to write the Hittite word of as yet unidentified pronunciation (though complements show that its root ended in *r*). Hittite can add further phonographic components to write, e.g., the dative case *-i*. Therefore  ?*r-i* above (L86; B142) is analysable as just two components <‘horse’ + *ri*> (contrast this with its conventional non-layered transcription ANŠE.KUR.RA.*ri*). These ‘god’ and ‘horse’ examples illustrate how a spelling sequence borrowed between languages can result in different structural interpretations presented here in the form of layering, but the ‘horse’ examples also show how a sequence of more than one logogram in one language, where each logogram represents a separate word, can be reinterpreted as a single logogram, representing a single word, in a borrower language. This is one area in which functional structure can transcend the formal structure of the frame in East Asian languages.

Consider a Japanese example. In Chinese, the meaning ‘tastiness’ (or more specifically ‘a delicacy’) is expressed by the combination of two lexical morphemes 美味 *měi wèi*, literally ‘beautiful taste’. In Japanese, *oishi-i* ‘to be tasty’, with the monomorphemic root *oishi-*, can be written as 美味しい, combining the two Chinese characters 美味 plus a phonetic complement (or ‘okurigana’ as it is known in Japanese studies) しい *shii*: <‘tastiness’ + *shii*>. The form given has the additional nonpast inflection *-i*. Despite the important formal structural principle of the frame in both Chinese and Japanese writing, the Japanese word consists functionally of a single logogram despite occupying multiple frames.

Another way in which layering is useful is in regard to the phonetic component. Etymologically, it is recognised that the overwhelming majority of Chinese characters are of the so-called ‘xingsheng’ type (in the next section formulated as <**sP**> or <**Ps**>), consisting of as accurate a phonetic component as the script allowed plus a ‘semantic determinative’ which broadly indicates the word’s semantic field. Egyptian script made overwhelming use of a similar principle, writing the word phonetically as accurately as

the writing system allowed, combined with one or more semantic determinatives (formulated as <**Ps**(**s**)^{n}>). Leaving aside the issue of how many semantic determinatives were used, the underlying functional structure of logograms in the two scripts is remarkably similar: approximation of pronunciation + semantic determinative. However, formally the two writing systems differed in the nature of this approximation of pronunciation. In Chinese, the phonetic component was derived from an (originally) independent character that possessed the required approximation. Because of the 1 morpheme = 1 syllable = 1 character principle underlying the mapping of Chinese language to Chinese writing system the chosen independent character approximated the entire morpheme = syllable when applied as a phonetic component. In contrast, Egyptian developed a series of phonograms that represented consonants or the consonantal frame of a word, ranging from monoliterals representing a single consonant to bi- and triliterals representing two or three consonants respectively, the latter often combined with disambiguating monoliterals (e.g. biliteral  *wr* disambiguated by monoliteral *r* in an example in Section 4.2). For example, the word *dpt* 'ship' was written  with four signs:  *d*, *p*, *t*, and  BOAT. Layering allows us to treat the *d*, *p* and *t* signs as elements that together constitute the single phonetic component *dpt*, and so the logogram in question is analysed as having two components <*dpt* + BOAT>, comparable, for example, with Chinese *jiàn* 'warship' written 艦 <BOAT + *jiàn*>. In other words, the structures of Egyptian *dpt* and Chinese *jiàn* are functionally similar two-component logograms, regardless of the formal differences of consonantal vs. syllabic spellings of the phonetic component. Relegating the individual phonetic signs of Egyptian, and also those of Mayan or cuneiform, to the element layer therefore makes Chinese and Egyptian writing comparable in a meaningful functional way.

The net result of layering is the separation of the formal and functional levels of analysis, allowing the functional analysis to deal with logograms and their components only – in other words, treating the elements, which so often are treated as 'graphemes', as functionally irrelevant – and also thereby allowing more meaningful cross-linguistic analysis.

3. Component types

The application of layering, and the recognition of the frame as just a formal unit of structure, allows us to focus on the functional structure of logograms. The two conventionally recognised values of 'signs' are semantic or phonetic, hence the frequent division of signs into semantograms and phonograms. Within our analysis, a word written with an entirely unanalysable but non-phonographic logogram is treated as a true semantogram <**S**>, in practice typically a pictogram in origin; a word written entirely phonographically is similarly treated as <**P**>. All other logograms are combinations of more than one component. In this section I shall argue that the possible combinational permutations of the two values – semantic and phonetic – allows us to establish four synchronic types of component:

S	+ semantic	– phonetic
P	– semantic	+ phonetic
X	– semantic	– phonetic
Ψ	+ semantic	+ phonetic

As the analysis is synchronic, etymology should not influence our analysis. Therefore, to count as an **S** or **P**, a component must have this value in at least one other occurrence within the writing system, either as a component in another logogram or as a whole logogram with an appropriately similar meaning or pronunciation. We also notice that some components in writing systems are 'partial' or 'approximate' (defined relatively within the principles of the individual writing system) in their value, a point which I discuss later. I shall indicate these components with lower case **s** and **p**. As illustration, consider the following Chinese logograms with the natural phenomenon component tagged WEATHER in this paper:

(1) *Chinese logograms sharing the* WEATHER *component*

a. 雹 *báo* 'hail' <**sP**> = <WEATHER + *bāo*>

b. 霖 *lín* 'heavy rain' <**sP**> = <WEATHER + *lín*>

c. 露 *lù* 'dew' <**sP**> = <WEATHER + *lù*>

d. 霜 *shuāng* ‘frost’ <**sP**> = <WEATHER + *xiāng*>

e. 雲 *yún* ‘cloud’ <**sP**> = <WEATHER + *yún*>

f. 震 *zhèn* ‘quake’ <**sP**> = <WEATHER + *chén*>

g. 電 *diàn* ‘electric(ity)’ <**sX**> = <WEATHER + ?>

h. 雷 *léi* ‘thunder’ <**sX**> = <WEATHER + ?>

i. 雪 *xuě* ‘snow’ <**sX**> = <WEATHER + ?>.

The other component of (1a.-f.) synchronically functions as a phonetic component **P** (albeit in the case of (1d.) a little borderline due to sound changes since the character was created). However, (1g.-i.) all combine WEATHER with a component that has no identifiable semantic or phonetic value: for example, the component in (1h.) as an independent character is 田 *tián* ‘ricefield’. We therefore have to treat the second component in (1g.-i.) as a valueless component **X**, which serves only to make the overall logogram unique. This is similar to the concept of the ‘jihaofu’ in Chinese studies (Chen 1999: 135), which is identified when an original component with a value has been simplified in its writing so that it is no longer formally identifiable. I would argue, however, that in a synchronic analysis the effects of sound change making a **P**-component less clear are also examples of **X**. The **X**-component is also significant when a writing system is borrowed, because a phonetic component in writing a word in the source language no longer functions as such when it writes an unrelated word in the target language (see the Sumerian *ra* ‘phonogram’ 𒁀 in the writing of Akkadian *sisû* in Section 2.2). I also treat the short vertical stroke 𓏲 (Gardiner sign number Z1), illustrated in the spelling of *niwt* and *kmt* earlier, as an **X**. In this case, the sign does have a function, but this is simply to indicate that the word is not spelt in the <**Ps(s)**^{n}> pattern that dominates the writing system but as <**S**> or <**Sp**> (or rather, with recognition of the sign as **X**, <**SX**> or <**SpX**>). It in itself does not have semantic or phonetic value, though. Similar to Gardiner Z1 are simple symbols used in some

Vietnamese (Hannas 1997: 81), Tay-Nung (Doan 1996: 94), or Kam texts (Liang 1980: 89) to indicate various functions, such as that the originally Chinese character that they are attached to does not represent a Chinese loanword, but is a native word (or a fully nativised older Chinese loan), either <**SX**> or <**PX**>, e.g. 木ᐟ <*mộc* + **X**> = Vietnamese *mọc* 'to grow', 風 <'wind' + **X**> = Kam *lemc* 'wind'.[9]

The **Ψ** component is useful for those cases where the logogram for one word includes a component that (a.) is itself an alternative logogram for the same word, (b.) a phonogram derived from a case of (a.), or (c.) the logogram for a closely related word so that one of the two may be considered a derivative of the other. As an example of (a.), consider Mayan *chum* 'to sit' written either with () or without () a conflated *m[u]* Complement  ; and as an example of (b.) see an Old Japanese spelling of *umey* 'plum (blossom)' as 宇梅 <*u* + *mey*/'plum'> = <**pΨ**> (Vovin 2005: 24), a loanword from Chinese, which in Chinese is written with a character 梅 that Old Japanese had adapted as a phonogram *mey*. As examples of (c.), consider the following Tangut pairs.

[9] This function is, arguably, a 'language-indicator', and we can group with Vietnamese, Tay-Nung, and Kam diacritics other indicators of the language of the word being written. These include the so-called Glossenkeil 𒀸 (Rüster & Neu 248) in Hittite writing to indicate that the word being written is not Hittite but a Luwian loanword, and the MOUTH component of Chinese character script (which indicates a noun or action related to the mouth in Chinese writing), which is widely used in Cantonese-specific characters and to a slightly lesser extent in Vietnamese-, Zhuang-, or Bai-specific characters to indicate that the word is 'dialect', i.e. a native word and not one found in (Standard) Chinese. In light of the fact that MOUTH in this use has a clear function and tends to be considered an extension of its usual use in the Chinese system (i.e. an **s**), we are hard-pressed to describe these language indicators as **X**; similarly, **s** is problematic as we are hard-pressed to see any semantic value in them, beyond 'any meaning at all as long as it is in such-and-such a language'. Arguably 'language indicators' are a fifth component type **L**(anguage).

(2) *Tangut word pairs*

a. Tangut  *bja*[2] ‘to break, end’ (intransitive) (Li4459).

a´. Tangut  *phja*[1] ‘to break/cut off’ (transitive) = <**sΨ**> (Li3708).

b. Tangut  *lwu*[1] ‘to get mixed’ (intransitive) (Li3078).

b´. Tangut  *lwụ*[1] ‘to mix’ (transitive) = <**XΨ**> (Li4850).

c. Tangut  *rjir*[1] ‘to get’ (Li1599).

c´. Tangut  *rjor*[1] ‘to get’ = ?<**sΨ**>, ?<**XΨ**> (Li23).

d. Tangut  *sjwɨ*[2] = ‘who’ (Li432).

d´. Tangut  *sjwɨ*[1] ‘who’ = <**Ψs**> (Li441).

(2a./a´.) and (2b./b´.) are intransitive/transitive pairs of verbs (reflecting characteristic voice, tonal and phonation variations used in Tangut derivation of such pairs), (2c./c´.) consist of different ablaut-based inflections of the same lexical verb, and (2d./d´.) consist of the same word, which can be pronounced apparently freely with either of the tones (tagged [1] and [2]) of the language. The Tangut writing system makes clear use of the logogram for the first word of each pair in the logogram of the second word of each pair. For example, the lower part of the logogram for *lwụ*[1] ‘to mix’ (transitive) uses the logogram for *lwu*[1] ‘to mix’ (intransitive) as a component both for its phonetic and its semantic values. A comparable Chinese example is the use of the character 分 *fēn* ‘divide’ in the character 份 *fěn* ‘a portion’, the latter analysable as <**sΨ**> = <PERSON + *fēn*/‘divide’>. One interpretation of the problematic ‘zhuanzhu’ category of the *liushu* classification of Chinese characters is that it too reflects the writing of derivation-related words in Old Chinese.

There are cases in which for aesthetic/organisational reasons the elements of a component are split by another component, such as in Tangut , (2d´.) above.

Egyptian, which in form is governed by the frame, provides various examples, such as in the usual writing of the verb *m3* ‘see’, which consists of two components and three elements: <[*m3* + *3*] + EYE>. However, as the *m3* and EYE elements are somewhat long-thin and the *3* element is somewhat tall-thin, the elements are reordered to fit the frame aesthetically in the sequence *m3* + EYE + *3*: . This reordering is purely aesthetic-formal and not functional, so the logogram is still to be analysed as <**Ps**>, not with two **p** components. Two **p** components would only be appropriate in cases where both the start and the end of a word – but not the middle – are written phonographically, as occurs sometimes in Mayan, e.g. the spelling of *winik* ‘person’ as <*wi* + ‘person’ + *k[i]*> = <**pSp**> (Thompson 0117.0521:0102).

There are two modifications to the four-way **S P X Ψ** analysis. The first is the use of ‘partial’ components, distinguished algebraically as lower case **s p** in the examples above and below, corresponding to the concept of ‘semantic determinative’ in Mesopotamian/Egyptian traditions and ‘phonetic complement’ in Mesopotamian/Mayan traditions. These are components that, within the potential of the writing system, are regularly less precise. **p** is used when only part of the word – the beginning or the end – is written, whereas **s** is used when only the semantic field is indicated, not the actual meaning. The Egyptian and Mesopotamian families use both **s** and **p**, although the repertoire of **s** in the latter is much fewer than in the former; the Chinese, Vietnamese, Zhuang, Yi, and Tangut writing systems only use **s**, essentially because the writing systems cannot indicate a unit lower than a syllable and the languages are essentially monosyllabic; while the Mayan writing system virtually only uses **p**. The ‘cartouche (+ pedestal)’ that typically encloses the names of the twenty days in the 260-Day Count in Mayan appears so far to be the only clear-cut **s** in the system, DAY, e.g. *ajaw* ‘lord’ <**S**> versus *ajaw* ‘20th day’ <**sΨ**> = <DAY × [*ajaw*/‘lord’]>.[10] The Chinese

[10] The difference between **S** and **s** can sometimes be unclear. For example, because the principle of Egyptian writing is that a word-final semantic component is always a ‘[semantic] determinative’ (i.e. **s**) and the fact that the BOAT ‘determinative’ is used to write several words with a meaning of ‘boat’ (Gardiner 1957: 498), I have treated the spelling of *dpt* ‘boat’ earlier as <**Ps**> rather than <**PS**>, even

writing system is similar in that the dominant pattern is <**sP**>/<**Ps**> with a limited number of semantic components available in the system, and <**SP**>/<**PS**> does not occur. This fact is particularly notable when we see how the Chinese writing system was adapted to write Vietnamese or Zhuang, which, in contrast, are notable for their free use of potentially *any* logogram as a semantic component of another, and for their frequent use of the <**PS**> structure not found in the Chinese system. This is illustrated in the opening line of the Vietnamese national epic *Truyện Kiều*: 𤾓𢆥𥪞𡎝𠊛些 *trăm năm trong cõi người ta* 'a hundred years in our world'. The first three logograms include a semantic component that not only gives the precise meaning (and which is a logogram in its own right with the same meaning in Chinese), but also is not in the limited set of **s**-components in the Chinese system. (It should be noted that phonetics that begin with /l/ are used in writing *tr*-initial words because in Early Modern Vietnamese these words began with *tl-/bl-*. For example, *trăm* 'hundred' is listed as *tlăm* in Rhodes (1651: 802))

(3) *Vietnamese logograms*

a. 𤾓 *trăm* 'hundred', <**SP**> = <'100' + *lâm*>

b. 𢆥 *năm* 'year', <**PS**> = <*nam* + 'year'>

c. 𥪞 *trong* 'inside', <**PS**> = <*long* + 'inside'>

A second modification is the use of 'reduced' components, where the act of reduction itself may have a function. A clear example of functional deletion is the logogram 冇 used to write *móuh* 'not to exist, not be located, not have' in Cantonese, which expresses the negative polarity of the word by deleting part of the logogram used to write its affirmative equivalent 有 *yáuh* 'to exist, be located, have'. A similar example

though BOAT appears to have the same meaning as the word *dpt*; in the many cases of names of gods or words for animal species in which the 'determinative' is unique to (and in hieroglyphic writing pictorially identifiable as) the referent of the word, it is impossible to support anything other than a <**PS**> analysis.

is reported for Zhuang (Ramsey 1987: 242), where deletion of one side of the Chinese character for 'gate' 門 produces the Zhuang character for 'side' 尸, which is listed as *mbangj* 'part' and *mbiengj* 'piece; half' in Su (1989: 314, 317).

The Tangut writing system is one that appears to make widespread use of *element* reduction for purposes of fitting the logogram to the frame, for example 𗍫 *śiwe*[1] 'honey' above. Japanese also has some examples amongst its made-in-Japan 'Chinese' characters, for example with the WIND-component 風 retaining only the outer two-stroke yet still distinctive frame of its nine-stroke form, e.g. the writing of *tako* 'kite' as 凧 <~~'wind'~~ × 'cloth'>. Where the deletion is of an element, it is below the level of component and is only a formal feature, therefore we ignore it. In an exceptional case like Cantonese *móuh*, however, the deletion is functional. I would argue that the act of deletion itself should be considered a semantic component.

I shall conclude this section by considering the possible permutations of semantic+semantic combinations, which application of this framework reveals to be more varied than merely the combination of two components for their meaning value. <**ss**> is used here to indicate that neither component provides the full meaning, but that the two components provide some aspect of the meaning that together logically represent the whole, as well illustrated by Sumerian, e.g. 𒊩𒆠 <'woman' + 'mountain/country'> = *géme* 'female slave', based on the fact that slaves were likely foreign captives. These correspond to the 'huiyi' category of the Chinese writing family. For example, the spelling of 明 <'sun' + 'moon'> = Chinese *míng* 'bright', or 畑 <'fire' + 'paddy field'> = Japanese *hatake* 'dry field'. Etymologically, most Chinese huiyi characters are <**Ps**> logograms (Boltz 1994: 147-149); in the case of *míng* the **P**-component has been graphically simplified to its current form.[11] However, we should deal with the synchronic, and synchronically 明 is <**ss**>. I would exclude from the category of <**ss**> such sequences that can be interpreted as the pictorial representation of

[11] Chinese does have a few, later true <**ss**> characters, e.g. 歪 <'not' + 'correct'> = *wāi* 'askew'.

a single action (and therefore <**S**>), e.g. the use of a representation of a hand + other strokes in (pre-)Seal Script Chinese (Wieger 1927: 119-140) or Egyptian (Gardiner numbers A41-A46), especially holding objects, or, for that matter, the 'scatter' hand in Mayan consisting of a hand typically with small objects falling from it (Mora-Marín 2008: 204-205). Where and how we draw the line is uncertain. Cases such as the variety of male heads in Mayan are a case in point. Do they consist of a single polyvalent sign, in which the distinguishing pictorial additions are disambiguating, as proposed by Mora-Marín (2008: 202, 204), or are they single discrete signs? If it is the former case, but the additions do not occur with similar semantic function elsewhere in the system, then we must treat them as **X**, as we did with Egyptian Z1.

<**ss**> is in contrast with <**sS**>, an attested sequence in Sumerian, and arguably a better interpretation of the differences between the formally identical Vietnamese and Zhuang characters 𡗶 <'sky' + 'above'> = V. *trời/giời* (earlier *blời*) 'sky' <**Ss**>, Z. *gwnz* 'top, above' <**sS**>.

<**ss**> is also in theory in contrast with <**SS**> that seems unlikely to occur, because it consists of two components that both independently express the meaning of the whole. Nevertheless, there are occurrences which I would argue are best represented as <<**S**> <**S**>>, following the format used for superlogograms below, where each component represents the entire meaning of a whole morpheme, but the two components represent different morphemes. The classic example would be the Mayan use of divine heads to represent the numbers '1'-'9' (e.g.  *wak* '6'), and the skeletal  to represent *lajuun* '10'; the teens '11'-'19' are expressed by the suffixation of *-lajuun* to '1'-'9' respectively, but this is represented not by a separate sign but by varying degrees of skeletalisation of the divine heads, e.g.  <<'6'> × <'10'>> = *waklajuun* '16'. Finally, Hannas's (1997: 81-82) list of three Vietnamese characters created on the 'compound ideographic' principle (i.e. our <**ss**>) includes at least one apparent true <**SS**> example 𠊚 <'to be lost' + 'to lose'> = *mất* 'lose; be lost; die'.[12]

[12] There may be Mayan examples of <**SS**> too in the representation of *ajaw* 'lord'. Firstly, there are signs consisting of several types of head – human, vulture etc. – each with a royal headband (no Thompson number as he treats it as part of a sign; Macri & Looper 2003 sign ZB3) on, that represent

4. Revelations of a functional analysis

The strict application of this analysis reveals significant differences between writing systems, including differences between systems in the same writing family. Different styles, different authors, different periods all reveal differences, as also the comparison between the 'simplified' characters used in the People's Republic of China and the 'traditional' ones used elsewhere in the Chinese-speaking world. More importantly, strict analysis on the basis of function, ignoring the formal features of how a writing system is visually presented, reveals some interesting results. It is not possible here for reasons of space to present any detailed analysis of any one writing system, which will be presented in future papers. I instead concentrate cross-linguistically on two issues: what is functionally phonographic (Section 4.1) and the existence of a functional layer above that of the logogram (Section 4.2).

4.1 Phonograms

In discussing the issue of phonograms, it is convenient to begin with Mayan studies, in which it is conventional to treat signs with CV readings as 'syllabograms' (i.e. phonograms) and those with CVC readings as 'logograms' (i.e. semantograms). Within the tradition this implies a clearcut phonetic vs. semantic distinction. Certainly CV-read signs appear to be consistently phonetic in value, while because Mayan words are all consonant-final, CVC-read signs do tend to be <**S**> or **S**-components. However, there are problems with this. Firstly, there are a number of CVC-read signs that appear to be used for (near-)homophones, such as  (T111) for *b'aak* 'bone' and *b'aak* 'captive',

---

ajaw. One interpretation is that the head sign by itself can already express *ajaw* as one of its values and the royal headband also expresses *ajaw* (Macri & Looper 2003: 27-28, 30; Mora-Marín 2008: 201-202). So the combination of head and headband *may* constitute <**SS**> writing. Less controversially, there exists another sign  (T168) that is has long been identified as an affix sign to write *ajaw* 'lord'. There occur, however, some examples of the lord-head + royal headband sign described above combined with  (a combination catalogued as AM8 in Macri & Looper (2003: 68)) that would seem to be clearcut cases of an **SS** sequence even if we ignore the headband (and **SSS** if we accept that the headband itself also represents *ajaw*).

and a degree of interchangeability between the signs for *chan*  ‘sky’ (T561c),  ‘snake’ (T764), and  ‘4’ (TIV). This is clearly the rebus principle at work. While the use of one *chan* sign for a homophone is phonographic, within the strict analysis proposed here the CVC sign  would appear to be functionally phonographic in all circumstances. These are parallel to various Old Japanese ‘kun’gana’ examples (many of which are one-off instances of visual wordplay), e.g. the use of the character for ‘anger’ 慍 for both *ikari* ‘anchor’ and *ikari* ‘anger’, or that for ‘crane’ 鶴 for both *turu* ‘crane’ and the attributive form of the perfective aspect *-turu* (Seeley 1991: 50), or the character for ‘yard’ 庭 for both *nipa* ‘garden’ and dative *-ni* + topic marker *-pa* (Seeley 1991: 189).

Secondly, following their (1998) attempt to associate syn- vs. disharmony between the choice of ‘silent vowel’ in word-final C[V]-phonograms and the preceding actual vowel with length or complexity of the latter, Houston, Stuart & Robinson (2001) proposed a particular use of such CV signs that are characterised as having flipped readings, e.g.  *b’i > IB’*, and functioning more semantographically for specific derivational suffixes. They name these ‘morphosyllabic signs’ in contrast with ‘syllabograms’ and ‘logograms’, and transcribe them in capitals following the same convention as used for ‘logograms’. This position is supported in Macri & Looper (2003: 32-33), but the fact that, even if we accept the existence of flipping, these cases still have a transparently phonographic spelling and the fact that a form such as  *chumib’* ‘sitting place’ can be analysed without recourse to ‘morphosyllabic signs’ as <‘to sit’ + *b’[i]*> following the basic principle of ‘silent vowel’ spelling mean that in our analysis such signs are treated as phonographic. The fact that <‘to sit’> by itself is pronounced *chum* meaning that there is no vowel for *b’[i]* in writing *chumib’* to be synharmonic with is a non-argument, in my opinion, as Mayan is merely writing the phonographic end of the word minimally; parallels may be found in pre-modern Japanese writing where verb-forms such as *if-u* ‘say’-Conclusive and *if-ite* ‘say’-‘and’ were frequently written 云フ <‘to say’ + *fu*> but 云テ <‘to say’ + *te*> rather than 云

ヒテ <‘to say’ + *fite*>. Further support for treating ‘morphosyllabic signs’ as merely phonograms is the fact that some of these VC-read signs are used for more than one homophonous morpheme, e.g. the *ja/AJ* sign both for the unpossessed marker on nouns and for the passive morpheme on verbs.

In East Asia, we encounter cases of Chinese characters used in Japan purely as phonograms, a throwback to the man’yōgana system of Old Japanese. The most common example perhaps is that of *furo* ‘bath’. Though apparently monomorphemic, *furo* is frequently still written with two characters 風呂, whose Chinese-derived readings are *fū* and *ro* and whose meanings – ‘wind’ and the name of an ancient Chinese state – have nothing to do with the meaning of ‘bath’. They are clearly phonograms, despite belonging to the script repertoire of Chinese characters rather than to that of hiragana. The device in modern Japanese, known as ‘ateji’ (which we can define as Chinese characters used as Modern Japanese phonograms), is commonly used to write a very small number of words.

There is arguably one example of the opposite, namely a kana used semantographically. The hiragana symbol historically representing the syllable *wo*, を, was a phonogram in pre-1946 Japanese. However, the syllables /wo/ and /o/ appear to have merged perhaps as early as around 1000 and certainly before the sixteenth century (Martin 1987: 38, 79) and manuscripts display a degree of interchangeability. After the kana reforms of 1946 (Seeley 1991: 153-154), the *wo* phonogram was abolished except to write the accusative case morpheme /o/, and all other examples of /o/ are now written with the *o* phonogram. In effect, the original *wo* phonogram is now a semantogram, with the value of ‘Accusative’.

In addition to ateji like *furo*, Japanese still uses one made-in-Japan ‘Chinese’ character 麿 that is composed of two phonographically used characters 麻 *ma* × 呂 *ro*, used to write the morpheme *maro*, historically ‘lad’ but now mostly restricted to use in men’s names; that the frame is irrelevant to a functional analysis is shown by the fact that *maro* can also be written in two frames as two characters: 麻呂. Parallels to *maro* can be found in ‘conflation’ or ‘monograms’ in other parts of the world. Examples are Mayan *mol* ‘8th month’, typically written with a sign  (T581) conflating

phonograms  *mo* ×  *l[o]* into a single sign and treated by Mayanists as a 'logogram'; Sumerian  *ğanun* 'granary' (L244), which conflates phonograms  *ğá* ×  *nun*; or Linear B  *meri* (= Greek *méli*) 'honey' (Bennett 1996: 128), conflating phonograms *me* × *ri*. 'Monograms' are also reported for Tay-Nung for phrasal verbs, e.g. *ku pak* 'raise one's voice' written in a single frame 旬 with rebus-based phonograms *ku* × *pak* (Doan 1996: 93).

麿 *maro* is unusual in East Asia in its use of two phonetic components within a single frame, but if we follow the principle stated above of ignoring aesthetic/formal factors and only consider the functional, then there is no difference between 麿 *maro* and 風呂 *furo* functionally and when *maro* is written as 麻呂, it is clearly no different than the writing of *furo*.

4.2 Superlogograms

The principles above have analysed the writing systems into logogram ≥ component ≥ element; the recursiveness of logogram creation means that there may be multiple layers of analysis. However, there occasionally occurs a higher level than logogram, which I term here for convenience superlogogram. The superlogogram consists of more than one logogram, which each in turn completely writes an entire word in the language, but the sequence of logograms is grouped into a larger unit by the addition of a new component, one which applies to the *whole* sequence. We find examples in both Egyptian and Sumerian writing families. Significantly in both cases it applies especially to compound names. Consider the Egyptian phrase *wꜣḏ wr* 'the sea' (literally the 'big green'), written with the two independent logograms that write the independent words  *wꜣḏ* 'green; papyrus' (analysable in itself as <**S**>) and  *wr* 'big' (analysable in itself as <**P**> = <[*wr* + *r*]>). The phrase *wꜣḏ wr*, however, is written with the two logograms combined as a single superlogogram because of the addition of a semantic

component associated with waterways, here glossed WATERWAY, that must apply to the *whole* phrase  <<**S**> <**P**> **s**> = <<'green'> + <[*wr* + *r*]> + WATERWAY>.

Similar examples in Sumerian involve the use of the semantic component  GOD before a god's name that consists of more than one morpheme, or the semantic component  PLACE after a polymorphemic placename. The writing systems, therefore, extend the principle of Layering to a level above that of logogram, hence my term superlogogram. In the case of Egyptian, there are two additional manifestations. The first is the use of the so-called 'cartouche' component  that encloses an entire prenomen- or nomen-type of king's name, which itself consists of more than one word and therefore more than one logogram. The cartouche, therefore, is a component associated primarily with the superlogogram. The second is the principle of 'honorific transposition' (Allen 2000: 42), whereby a logogram with an honorific referent (typically 'god', 'king' or a name of a god) is fronted to the first position in a phrase, normally a head-genitive phrase that may alternatively be treated as a compound word. For example, *ḥwt-nṯr* 'temple', derived from  *ḥwt* 'enclosure' + *nṯr* 'god', is written  <<**Ψ**> <**Sps**>> = <<'god'> + < 'enclosure' + *t* + BUILDING>>.[13] That such honorific transpositions only occur within the narrow phrase and not over the scope of an entire clause or more points to the spelling of *ḥwt-nṯr* 'temple' being a superlogogram. This fronting also occurs within 'cartouche' names, which normally include the name of a god, e.g. *mry-rˁ* 'Meri-Re' (literally 'beloved of Re = the sun god') (Allen 2000: 42) is analysed here as <KINGNAME × <'sun'> + <[*mr* + *y*]>>.

[13] I tentatively treat here as **Ψ** because it is found as a phonogram with the value *nṯr* or *ntr* in words not etymologically related to *nṯr* 'god', and also as an occasional semantic determinative GOD.

5. Problems in cross-linguistic comparison

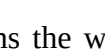
There are, of course, always going to be problems with such a strict framework. These include a degree of subjectivity in deciding at what point semantic divergence or phonetic dissimilarity constitute **X** rather than **S**/**s** and **P**/**p**. Phonetic dissimilarity is particularly an issue with Chinese characters, because the phonetic component was always somewhat approximate and over the millennia the phonetic similarity has often been lost due to regular sound change, whereas semantic dissimilarity has the additional problem that we may not actually know why a scribe chose a particular semantic component. DeFrancis (1984: 95-96) in reference to Chinese examples writes that "Fertile imaginations, to be sure can often come up with fanciful explanations in cases like these". It should be noted that DeFrancis deals with a two-way semantic/phonetic distinction, in contrast to the four-way distinction that we have suggested here. Moreover, the semantic and phonetic components in the overwhelming majority of logograms in each language should be readily explainable.

When we are working within the Chinese writing family and comparing the systems used in typologically similar languages such as (standard written) Chinese, Cantonese, Vietnamese, Zhuang etc., application of the approach laid out in this paper encounters few substantial problems other than the significant lack of information for Cantonese, Vietnamese, or Zhuang. For example, Takeuchi's (1988) explanations for some Vietnamese characters may be wrong because layering and the system are not considered, whereas Su's (1989) dictionary of sawndip has no explanation at all on the make-up of characters.

The analysis proposed here reveals interesting patterns, such as the use of <**SP**>/<**PS**> in Vietnamese in contrast to <**sP**>/<**Ps**> in Chinese, the nature of **X**, the complex variety of semantic+semantic combinations, and overall frequency. Token/type statistical analysis of these writing systems as used in actual texts should be very informative if a framework such as that proposed here is applied systematically.

However, though we can use the approach to characterise similarities and differences between, say, Chinese, Sumerian, Egyptian, and Mayan, or between members of a writing family such as Sumerian, Akkadian, and Hittite, or Chinese, Japanese, and Korean, an objective statistical analysis would seem to elude us. This is not because of incompatibility between writing systems, which a functional analysis

involving layering and component types based on the fundamental semantic/phonetic dichotomy can describe effectively, and which I would argue reflects how scribes, or at least learners, perceive(d) the system that they are/were using.

The problem instead lies with the morphological typology of the languages themselves. Whereas Chinese, Tangut, Yi, and the Vietic and Tai languages are isolating and tend to have monosyllabic morphemes, Akkadian, Hittite, Egyptian, and Mayan are inflecting, while Japanese, Korean, and Sumerian are agglutinating. Moreover, clitics can also have a significant grammatical role. Just as it is difficult to identify what actually is a 'word' of a spoken language, it is problematic to identify what an ancient scribe of Sumerian or even a modern speaker of Japanese considers to be a word. For example, Sumerian *kalam-a* 'country'-GEN and Japanese *ik-e* 'go'-IMP have a morpheme boundary that occurs within a syllable. As both languages have syllable-based phonograms the words are typically written  <'country' + *ma*> and <'go' + *ke*>. The difficult question in the case of inflecting and agglutinating languages is where to draw the line. For example, corresponding to Sumerian *kalam-a* semantically is Japanese  *kuni no* 'country-GEN', in which case the morpheme boundary is the same as the syllable boundary, *no* does not enter into any sandhi. Though they are phonologically a single word the two morphemes are typically romanised as separate words.

6. Conclusion

This paper has taken the universally recognised semantic vs. phonetic dichotomy in such scripts with a view to analysing such writing systems both functionally and cross-linguistically. Crucially, a cross-linguistic approach to how such writing systems function must be on comparable grounds. For this reason, contrary to Coulmas (2003: 59), we start not from individual signs but from the 'word' of the relevant language and we identify how it is encoded in writing, defined as the logogram. Secondly, we apply a layering analysis to written forms in context, identifying what we have called components and ignoring the lower element layer. This avoids having to explain the values of elements when they do not directly provide a value in writing the logogram in

question and allows us to deal with the phonographic component of different phonographic systems on comparable terms. Thirdly, we ignore formal features such as the frame, which in any case is not used in all systems and corresponds to very different linguistic levels in those systems in which it is used.

Within this approach, almost all logograms can be explained as permutations of four component types based on the fundamental semantic/phonetic dichotomy (**S**, **P**, **X**, **Ψ**). We only allow modification to the above system in the form of full vs. partial components, defined within the terms of the writing system itself (for example, the phonographic component of Egyptian *dpt* is a full spelling, because Egyptian writing ignores vowels) and the concept of the superlogogram.

Certain issues thrown up by the range of writing systems considered have not been dealt here. These include: incomplete writing; nonsequentiality (including the encoding of one language in the syntax of another, such as Old Japanese *hentai kanbun* and Old Korean *pyŏnch'e hanmun* styles: Aldridge 2000; Nam 2012: 42); the use of repetition marks (particularly at the logogram-level as found in Egyptian and the Chinese writing family, which I would treat simply as <**R**>); and punctuation. Nevertheless, this framework, applied strictly, despite the typological problems of comparing isolating, inflecting, and agglutinating languages, reveals a range of complexity within writing systems as illustrated throughout this paper. The framework is in contrast both to such formal-functional mixed approaches as argued by DeFrancis (1989), with his emphasis on what exactly is a 'grapheme', and to the conventional analysis of each writing system in its own academic tradition which often have epigraphy or decipherment to consider. However, this framework is not meant as criticism of either, because our goals are quite different. This paper has sought to provide a broad cross-linguistic overview of the theory presented, but much further research is needed, particularly in its application to texts in individual languages.

References

Aldridge, Edith (2000). Principles of *hentai kanbun* word order: Evidence from the *Kojiki*. In Thomas E. McAuley (ed.), *Language change in East Asia*, 207-232. Richmond, UK: Curzon.

Allen, James P. (2000). *Middle Egyptian: An introduction to the language and culture of hieroglyphs*. Cambridge: Cambridge University Press.

Backhouse, Anthony E. (1984). Aspects of the graphological structure of Japanese. *Visible Language* 18: 219-228.

Bauer, Robert S. (1988). Written Cantonese of Hong Kong. *Cahiers de Linguistique – Asie Orientale* 17(2): 245-293.

Baxter, William H. (1992). *A handbook of Old Chinese phonology*. Berlin: Mouton de Gruyter.

Bennett, Emmett L. (1996). Aegean scripts. In Peter T. Daniels & William Bright (eds.), *The world's writing systems*, 125-133. New York/Oxford: Oxford University Press.

Boltz, William G. (1994). *The origin and early development of the Chinese writing system*. New Haven, Connecticut: American Oriental Society.

Bonnamy, Yvonne & Ashraf Sadek (2010). *Dictionnaire des hiéroglyphes*. Arles: Actes Sud.

Boone, Elizabeth Hill (2000). *Stories in red and black: Pictorial histories of the Aztecs and Mixtecs*. Austin: University of Texas Press.

Borger, Rykle (2004). *Mesopotamisches Zeichenlexikon*. Münster: Ugarit-Verlag.

Chen, Ping (1999). *Modern Chinese: History and sociolinguistics*. Cambridge: Cambridge University Press.

Chiang, William W. (1995). *"We two know the script; we have become good friends": Linguistic and social aspects of the women's script literacy in Southern Hunan, China*. Lanham, Maryland: University Press of America.

Chingertai [Qinggeertai], Fengzhu Liu, Naixiong Chen, Baolin Yu & Fuli Xing. (1985). *Qidan xiaozi yanjiu* [Research on Khitan lesser script]. Beijing: Zhongguo Shehui Kexue Chubanshe.

Cooper, Jerrold S. (2004). Babylonian beginnings: the origin of the cuneiform writing system in comparative perspective. In Stephen D. Houston (ed.), *The first writing: script invention as history and process*, 71-99. Cambridge: Cambridge University Press.

Coulmas, Florian (1989). *The writing systems of the world*. Oxford/Cambridge, Mass.: Blackwell.

Coulmas, Florian (1996). *The Blackwell encyclopedia of writing systems*. Oxford/Cambridge, Mass.: Blackwell.

Coulmas, Florian (2003). *Writing systems: An introduction to their linguistic analysis*. Cambridge: Cambridge University Press.

Daniels, Peter T. (1991). Is a structural graphemics possible? In Ruth M. Brend (ed.), *The eighteenth LACUS forum*, 528-537. Lake Bluff, Illinois: Linguistic Association of Canada and the United States.

DeFrancis, John (1984). *The Chinese language: Fact and fantasy*. Honolulu: University of Hawaii Press.

DeFrancis, John (1989). *Visible speech: The diverse oneness of writing systems*. Honolulu: University of Hawaii Press.

Ding, Chunshou (1993). *Yiwen lun* [Discourse on Yi writing]. Chengdu: Sichuan Minzu Chubanshe.

Doan, Thien Thuat [= Đoàn Thiện Thuật] (1996). *Tay-Nung language in the North Vietnam*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

Faulkner, Raymond O. (1999 [1962]). *A concise dictionary of Middle Egyptian*. Oxford: Griffith Institute, Ashmolean Museum.

Fedorova, Liudmila L. (2009). The emblematic script of the Aztec codices as a particular semiotic type of writing system. *Written Language & Literacy* 12(2): 258-275.

Gardiner, Alan (1927). *Egyptian grammar: Being an introduction to the study of hieroglyphs*. Oxford: Griffith Institute.

Gardiner, Alan (1957). *Egyptian grammar: Being an introduction to the study of hieroglyphs*. Third edition, revised. Oxford: Griffith Institute.

Gelb, Ignace J. (1963). *A study of writing*, 2nd edition. Chicago/London: University of Chicago Press.

Gragg, Gene B. (1996). Mesopotamian cuneiform: other languages. In Peter T. Daniels & William Bright (eds.), *The world's writing systems*, 58-70. New York/Oxford: Oxford University Press.

Grinstead, Eric (1972). *Analysis of the Tangut script*. Lund: Studentlitteratur.

Hannas, William C. (1997). *Asia's orthographic dilemma*. Honolulu: University of Hawai'i Press.

He, Jiren & Zhuyi Jiang (eds.) (1985). *Naxiyu jianzhi* [Sketch of Naxi]. Beijing: Minzu Chubanshe, Zhongguo Shaoshu Minzu Yuyan Jianzhi Congshu.

Houston, Stephen, David Stuart & John Robertson (1998). Disharmony in Maya hieroglyphic writing: linguistic change and continuity in classic society. In Andrés Ciudad Ruíz et al. (eds.), *Anatomía de una civilización: Aproximaciones interdisciplinarias a la cultura maya*, 275-296. Madrid: Sociedad Española de Estudios Mayas.

Houston, Stephen D., David E. Stuart & John S. Robertson (2001). Quality and quantity in glyphic nouns and adjectives. *Research Reports on Ancient Maya Writing* 47: 1-56.

Houston, Stephen, John Baines & Jerrold Cooper (2003). Last writing: Script obsolescence in Egypt, Mesopotamia, and Mesoamerica. *Comparative Studies in Society and History* 45(3): 430-479.

Jin, Jicong (ed.) (1984). *Nüzhen wen cidian* [Character dictionary of Jurchen writing]. Beijing: Wenwu Chubanshe.

Justeson, John S. & Terrence Kaufman (1993). A decipherment of epi-Olmec hieroglyphic writing. *Science* 259(5102): 1703-1711.

Kara, György (1996). Kitan and Jurchin. In Peter T. Daniels & William Bright (eds.), *The world's writing systems*, 230-238. New York/Oxford: Oxford University Press.

Kychanov, Evgenij Ivanovich (1996). Tangut. In Peter T. Daniels & William Bright (eds.), *The world's writing systems*, 228-230. New York/Oxford: Oxford University Press.

Labat, René (1976). *Manuel d'épigraphie akkadienne: Signes, syllabaire, idéogrammes*, revised edition. Paris: Paul Geuthner.

Li, Fanwen (2008). *Xiahan zidian* [Tangut-Chinese dictionary], 2nd revised edition. Beijing: Zhongguo Shehui Kexue Chubanshe.

Liang, Min (ed.) (1980). *Dongyu jianzhi* [Sketch of Kam]. Beijing: Minzu Chubanshe, Zhongguo Shaoshu Minzu Yuyan Jianzhi Congshu.

Liu, Fengzhu & Baolin Yu (1990). Fangkuai Baiwen [Bai characters]. In Zhongguo Minzu Guwenzi Yanjiuhui (eds.), *Zhongguo minzu guwenzi tulu* [Compendium of old writing systems of China's minorities], 247-249. Beijing: Zhongguo Shehui: Kexue Chubanshe.

Macri, Martha J. & Matthew G. Looper (2003). *The new catalog of Maya hieroglyphs, Vol. 1: The Classic period inscriptions*. Norman: University of Oklahoma Press.

Martin, Samuel E. (1987). *The Japanese language through time*. New Haven/London: Yale University Press.

Melchert, H. Craig (1996). Anatolian hieroglyphs. In Peter T. Daniels & William Bright (eds.), *The world's writing systems*, 120-124. New York/Oxford: Oxford University Press.

Montgomery, John (2002). *Dictionary of Maya hieroglyphs*. New York: Hippocrene Books, Inc.

Mora-Marín, David F. (2008). Full phonetic complementation, semantic classifiers, and semantic determinatives in ancient Mayan hieroglyphic writing. *Ancient Mesoamerica* 19(2): 195-213.

Nam Hoài-Bão (2001). *The Vietnamese writings through the ages*. Warrington, PA: Privately published.

Nam, Pung-hyun (2012). Old Korean. In Nicolas Tranter (ed.), *The languages of Japan and Korea*, 41-72. Abingdon: Routledge.

Nishida, Tatsuo (1964-1966). *Seikago no kenkyū: Seikago no saikōsei to Seika moji no kaidoku* [Research into Tangut: reconstruction of the Tangut language and decipherment of Tangut characters], 2 volumes. Tokyo: Zauho Kankōkai.

Nishida, Tatsuo (1980). *Seika moji: kaidoku no purosesu* [Tangut characters: the decipherment process]. Tokyo: Tamagawa Daigaku Shuppanbu.

Norman, Jerry (1988). *Chinese*. Cambridge: Cambridge University Press.

Ramsey, S. Robert (1987). *The languages of China*. Princeton, NJ: Princeton University Press.

Reiman, Etsuko Obata (1983). The status of kokuji: usage of ideographs unique to Japan. *Research on Language and Social Interaction* 16: 185-232.

Rhodes, Alexandre de (1651). *Dictionarium annamiticum, lusitanum, et latinum*. Rome: Sacra Congregatio de Propaganda Fide in Lucem.

Rüster, Christel & Erich Neu (1989). *Hethitisches Zeichenlexikon: Inventar und Interpretation der Keilschriftzeichen aus den Boğazköy-Texten*. Wiesbaden: Harrassowitz.

Sampson, Geoffrey (1985). *Writing systems*. London: Hutchinson.

Sasse, Walter (1980). «Chinesische» Zeichen erfunden in Korea. *Asiatische Studien* 34: 189–205.

Seeley, Christopher (1991). *A history of writing in Japan*. Leiden: E.J. Brill.

Shi, Dingxu (1996). The Yi script. In Peter T. Daniels & William Bright (eds.), *The world's writing systems*, 239-243. New York/Oxford: Oxford University Press.

Smith, Janet S. (Shibamoto) (1996). Japanese writing. In Peter T. Daniels & William Bright (eds.), *The world's writing systems*, 209-217. New York, Oxford: Oxford University Press.

Su, Yongqin (ed.-in-chief) (1989). *Gu Zhuangzi zidian / Sawndip sawdenj* [Old Zhuang characters dictionary]. Nanning: Guangxi Minzu Chubanshe.

Takeuchi, Yonosuke (ed.) (1988). *Chunomu jiten* [Chữ nôm dictionary]. Tokyo: Daigaku Shorin.

Taylor, Insŭp & Maurice Martin Taylor (1995). *Writing and literacy in Chinese, Korean, and Japanese*. Amsterdam: John Benjamins.

Thompson, J. Eric S. (1962). *A catalog of Maya hieroglyphs*. Norman: University of Oklahoma Press.

Thompson, Laurence C. (1965). *A Vietnamese grammar*. Seattle: University of Washington Press.

Thomsen, Marie-Louise (1984). *The Sumerian language*. Copenhagen: Akademisk Forlag.

Tranter, Nicolas (2008). Non-conventional script choice in Japan. *International Journal of the Sociology of Language* 192: 133-151.

Vovin, Alexander (2005). *A descriptive and comparative grammar of Western Old Japanese*, volume 1. Folkestone: Global Oriental.

Wei Qingwen (ed.) (1980). *Zhuangyu jianzhi* [Sketch of Zhuang]. Beijing: Minzu Chubanshe, Zhongguo Shaoshu Minzu Yuyan Jianzhi Congshu.

Wei, Qingwen (1989). Guangxi Zhuangzu-de fanghuai wenzi [Character script of the Zhuang of Guangxi]. In Qihuang Xie, Zaizhong Guo, Junqing Mo & Hongmei Lu (eds.), *Lingwai Zhuangzu huikao*. 581-597. Nanning: Guangxi Minzu Chubanshe.

Wieger, Léon (1927 [1965]). *Chinese characters: Their origin, etymology, history, classification and signification. A thorough study from Chinese documents*, 2nd edition. China: Catholic Mission Press [New York: Paragon Book Reprint Corp. & Dover Publications Inc.].

Xu, Lin & Yansun Zhao (eds.) (1984). *Baiyu jianzhi* [Sketch of Bai]. Beijing: Minzu Chubanshe, Zhongguo Shaoshu Minzu Yuyan Jianzhi Congshu.

Xu, Lin & Yansun Zhao (1990). Fangkuai Baiwen [Bai characters]. In Zhongguo Minzu Guwenzi Yanjiuhui (eds.), *Zhongguo minzu guwenzi tulu* [Compendium of old writing systems of China's minorities], 240-246. Beijing: Zhongguo Shehui: Kexue Chubanshe.

Yunnan Sheng Lunan Yizu Zizhixian Wenshi Yanjiushi (1984). *YiHan jianming cidian* [Concise Yi-Chinese character dictionary]. Kunming: Yunnan Minzu Chubanshe.

Zhang, Yuansheng (1990). Fangkuai Zhuangzi [Zhuang characters]. In Zhongguo Minzu Guwenzi Yanjiuhui (eds.), *Zhongguo minzu guwenzi tulu*, 227-232. Beijing: Zhongguo Shehui: Kexue Chubanshe.

Corresponding author

Nicolas Tranter
School of East Asian Studies
University of Sheffield
6-8 Shearwood Road
Sheffield S10 2TD
United Kingdom

n.tranter@sheffield.ac.uk