Morphology-phonology mismatches predict the limits of logography in Sumerian versus Chinese and their descendant systems

Introduction   Linguists seldom study writing systems. This is natural, given that grammar is independent of writing: it is acquired without writing and can withstand unaffected a shift from one writing system to another (as evidenced by the switch from Arabic to Latin script for Turkish, from Fidel to Latin for Konso). However, there are clear cases where the structure of writing systems, or the course of their evolution, is to be explained via grammatical factors. An obvious example is that the transmission of abjads (consonantal alphabets) to non-Afro-Asiatic languages generally prompts with the innovation of vowel writing, giving alphabets, as in Greek, or akṣaras, as in Brahmi. This paper argues that logography constitutes another case where the impact of grammar is detectable in writing system design and evolution. Based on comparative study of the development and subsequent borrowing of Sumerian and Chinese writing, this paper contends that the extent of logography—use of characters with inherent meaning—is synchronically and diachronically constrained by a simple grammatical factor: the degree of (mis)alignment between syllable boundaries and morpheme boundaries, especially for functional vocabulary.

Two parallels   The starting point for this argument is the observation of two parallels, one between two "pristine" (i.e., first-in-their-region) writing systems, the second between two of their descendants. **First**, the logogram inventories of Sumerian and Chinese are highly convergent, despite a separation of some 2000 years and very different linguistic and cultural terrain. They include pictograms (CH a picture of a tree for 'tree', SU a foot for 'foot'), simplex ideograms (CH the top of a tree highlighted 'top', SU the bottom of a foot highlighted for 'foundation'), complex ideograms (CH two trees for 'grove', SU two men for 'fight'), and compound ideograms (CH sun behind a tree for 'east', SU water in an enclosure for 'marsh'). Moreover, both systems developed rebus writing (writing of homophones for meanings unconducive to pictographic and ideographic writing) and exploited in the same two ways: for writing much functional vocabulary and for a fifth shared character type, the phonosemantic, combining one character for its sound, ignoring meaning, another for its meaning, ignoring sound (CH *yǔ* 'language' is *wú* '吾' with ~~*yán*~~ 'word', SU *eme* 'language' is *me* '~~be~~' with ~~*ka*~~ 'mouth'). Moreover, phonosemantic characters imply a further parallel, namely significant levels of recursion in character creation. **Second**, when borrowed by Akkadian and Japanese, nongenetic neighbours of Sumerian and Chinese respectively, the systems underwent the same change, at first almost losing their logographic character entirely in favour of near full phonography, then relegating phonography to relatively marginal roles in favour of substantial logography.

Explicandum   These parallels seem to suggest that pristine versus borrowed are natural classes of writing systems. So, against that backdrop, it is surprising to observe that Sumerian and Japanese, a pristine system and a borrowed-modified one, have converged on a near identical division of labour between logography and phonography: the restriction of logography to lexical vocabulary and the use of phonography to functional vocabulary and lexical-functional fusions. In Japanese, this division is marked by the use of kanji (Chinese) characters versus kana (simplified Japanese syllabograms). Sumerian lacked a phonographic syllabary and used rebus writing instead, but the division was the same. For instance, in *udu bar-ka* 'of white rams", *udu* 'ram' and *bar* 'white' are written with logograms and ideograms respectively, but genitive *ka* is written via rebus *ka* 'mouth', ignoring its meaning.

Incorrect explanations   The fact that phonography is used for functional vocabulary in both systems suggests that morphological complexity militates against logography. However, comparative analysis of early Chinese and Sumerian writing reveals surprising parity in morphological complexity. On the one hand, Sumerian, though morphologically rich for both nouns and verbs, left much morphology unwritten. For instance, in an early version of the often recopied *Instructions of Šuruppak*, the sentence 'The man of Šuruppak laid down instructions to his son' lacks the ergative, possessive, dative, agreement, and imperfective of later versions, and marks only the preverb and ventive. On the other hand, Old Chinese, in addition to negation, emphasis, and aspect amongst other functional items, frequently wrote morphologically complex words. Baxter and Sagart's reconstruction posits several tetramorphemic words, like 'be near to', approximately, *s-n-kər^{ʔ}-s* TR-STAT-near-DIR. This certainly compares with level of complexity of written Sumerian.

Moreover, although sometimes scant, written morphology in Sumerian shows noteworthy linguistic awareness. For instance, several morphemes possessed C versus CV allomorphs (such as *d(a)* COM, *r(a)* DAT, *š(e)* TERM). Others (such as *ha* MOD, *mu* VENT) underwent vowel harmony. Frequently, though, their written form abstracted away from this surface variation. There were also, at times, tendencies to prefer certain sound signs for certain grammatical morphemes. In consequence, it is implausible to tie properties of how Sumerian functional vocabulary is written to their excessive abstractness or otherwise to suggest grammatical understanding lay beyond their writers' grasp.

Morphology-phonology misalignment   The explanation for the convergence between the rules of Sumerian and Japanese logography lies, I contend, in a shared degree of misalignment between syllable boundaries and morpheme boundaries. I assume that the fundamental unit of phonetic analysis that early writers bring to task of writing is the syllable. This assumption is plausible in virtue of, on the one hand, the rich phonological, psycholinguistic, and neurolinguistic literature showing the pivotal role played by syllables in language knowledge, awareness, and use, and, on the other, the role of syllables in writing systems themselves. For instance, they frequently write complex syllables like C′CV or CVC′ by embellishing the symbol for CV, not with a consonant sign, but with a further syllable sign (e.g., C′[V]-CV; CV-[V]C′, CV-C′[V]), thus preserving the syllable as the fundamental unit of writing. In this regard, there is a perfect match between Sumerian and Japanese, versus Chinese. In the latter, all functional morphology is either tautosyllabic or subsyllabic (prefixes, infixes, suffixes). None has the property of shifting syllable boundaries. In Sumerian and Japanese, by contrast, boundaries often shift. Consider Old Japanese *myi-ye-^{n}s-ar-an-aku-n-i* (see-PASS-NEG-be-NEG-NOM-DFCT-INF). Written syllabically as *myi.ye.^{n}sa.ra.na.ku.ni*, only for the first two signs do syllable and morpheme boundary coincide. As a result, the subsequent five syllables are meaningless. By definition, **meaningless syllables make syllabic logography impossible**.

That this is in fact the driving force behind differences in Chinese and Sumerian/Japanese logography is underlined by other evidence. First, resyllabification of root consonants with vowel-initial suffixes often provides phonetic disambiguation of Sumerian and Japanese roots. This does not arise in Chinese. Correspondingly, Chinese, from the Small Seal Script onwards, came to rely much more on phonosemantic characters (i.e., character-internal phonetic complements) than Sumerian. Moreover, in the Sinosphere, degree of alignment between syllables and morphemes correctly predicts the extent of logography in other borrower languages, accurately distinguishing Korean and Japanese from Vietnamese and Zhuang.

(Cuneiform/Chinese symbols, omitted here for ease of reading, will feature in the talk.)