

Précis of *The Grammar of Chinese Characters**

James Myers

National Chung Cheng University

<http://personal.ccu.edu.tw/~lngmyers/>

2019/7/11

0. Introduction

The title of Myers (2019) is meant literally: Chinese characters really do seem to have a mentally active and productive grammar, with striking similarities to the morphology and phonology of signed and spoken languages. This paper simply sketches out the key points made in the book, one section per chapter. Section 1 outlines previous analyses of Chinese characters, which already reveal grammar-like properties. Section 2 argues that characters have morphological operations akin to affixation, compounding, and reduplication. Section 3 argues that characters also have phonology (of a silent sort, as in sign languages), which describes abstract formal regularities in strokes and overall character shape. Section 4 provides corpus-based evidence for the productivity of many of the above regularities, and Section 5 provides experimental evidence. Section 6 first considers possible explanations for character grammar and then sketches out how the idea might be useful beyond theoretical linguistics.

1. Chinese character grammar: The very idea

In this section I review the nature of Chinese characters and the nature of grammar, then link them together. Chinese characters (漢字 *hànzì*) are written symbols that almost always represent a single spoken morpheme (hence a single syllable) in Sinitic languages, including Mandarin. They have also been adapted and modified for non-Sinitic languages like Japanese (where 漢字 is pronounced *kanji*), Korean, and Vietnamese (see Handel, 2019). Most of my discussion will focus on so-called TRADITIONAL CHARACTERS, by far the longest-lived Chinese orthographic system, but now restricted mainly to Hong Kong, Macau, and Taiwan. In later sections I also discuss the system of SIMPLIFIED CHARACTERS developed in the People's Republic of China and now used across the world, including Singapore.

Though their pedigree is ancient, Chinese characters have never formed a fixed inventory, with new characters having been (and continuing to be) coined at an exponentially increasing rate; as illustrated in Figure 1, I mean this literally. This is possible because characters are constructed systematically, as recognized at least since the still immensely influential Shuowen classification system (說文解字 *Shuōwén Jiězì*, 100 CE). Of all character types, SEMANTIC-PHONETIC CHARACTERS are by far the most common (over 80%) and have been for 3,000 years (Huang, 2003); as illustrated in Table 1, these are composed of a SEMANTIC RADICAL and a PHONETIC COMPONENT. The next most important type are SEMANTIC COMPOUNDS, where character meaning relates to the meanings of its constituents. Characters formed via REDUPLICATION are traditionally considered a subtype of semantic compound, though they also have interesting properties of their own, as we will see later.

* For much more detail and nuance, see Myers (2019) [www.routledge.com/9781138290815; <http://personal.ccu.edu.tw/~lngmyers/CharGram.htm>]. My thanks to many talk audiences, book readers, lab assistants, and grants (NSC 97-2410-H-194-067-MY3, NSC 97-2410-H-194-067-MY3, MOST 103-2410-H-194-119-MY3, MOST 106-2410-H-194-055-MY3).

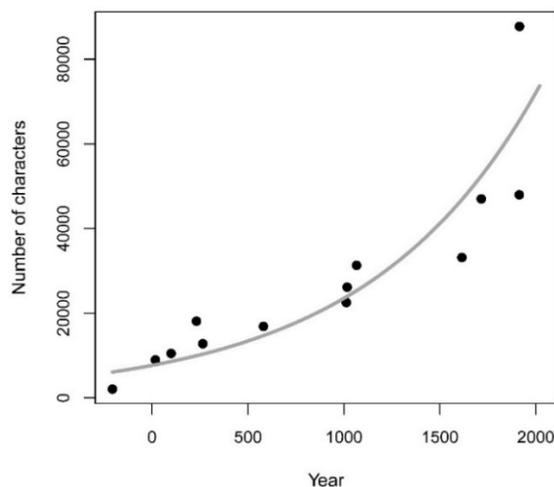


Figure 1. Chinese character dictionary sizes (data from Yip, 2000, p. 19)

Table 1. Some character types

Traditional terms	My terms	Examples
形聲 <i>xíngshēng</i>	Semantic-phonetic	媽 <i>mā</i> ‘mother’ = 女 <i>nǚ</i> ‘female’ + 馬 <i>mǎ</i> ‘horse’
會意 <i>huìyì</i>	Semantic compound	明 <i>míng</i> ‘bright’ = 日 <i>rì</i> ‘sun’ + 月 <i>yuè</i> ‘moon’
疊體字 <i>diétǐzì</i>	↳Reduplication	林 <i>lín</i> ‘forest’ (from 木 <i>mù</i> ‘wood, tree’)

How could we determine if such observations (and the many more discussed below) imply a genuine grammar? Linguists generally expect a grammar to be psychologically real, productive, and abstract. For example, we recognize that the natural sign languages of the deaf have grammars because fluent signers depend on systematic mental activity (e.g., Emmorey, 2001), they have productive knowledge that goes beyond rote memory (e.g., Berent & Dupuis, 2018), and this knowledge is abstract in three ways: it is not fully reducible to iconicity (e.g., Frishberg, 1975), it involves amodal (i.e., not solely articulatory or perceptual) mental representations (as recognized in the sign phonology literature; e.g., Brentari, 2011), and it not only comprises a mental module distinct from the rest of cognition, but is itself composed of distinct sub-modules (e.g., Padden & Perlmutter, 1987, on sign phonology vs. sign morphology).

Can writing systems also have grammars? Bloomfield (1933, p. 21) famously thought not: “Writing is not language, but merely a way of recording language by means of visible marks.” It is true that unlike spoken and signed languages, reading and writing are “unnatural”, in that they are not universal and depend on explicit teaching. Yet reading and writing also invoke undeniably real mental knowledge that productively extends what is explicitly taught (e.g., Pacton *et al.*, 2001). This knowledge is also abstract, being non-iconic (i.e., semi-independent of pronunciation or meaning; Householder, 1971), amodal (literacy depends on linking writing with reading; e.g., Naka, 1998), and modular (the Visual Word Form Area is a key part of the brain’s reading network; Dehaene & Cohen, 2011).

Linguists have also often observed how well the tools of grammatical analysis apply to orthographic systems. For example, the Greek GRAPHEME for /s/ has predictable ALLOGRAPHS <ς> word-finally and <σ> elsewhere, and slips of the pen reflect a supra-graphemic level of mental representation something like prosody, as illustrated by common misspellings like <Philippines> → <Phillipines>. As illustrated in (1), letters also alternate with each other in ways strikingly like phonological rules and constraints (after McCawley, 1994; see also Evertz, 2018).

- (1) a. $y \rightarrow i / _ X$ carry ~ carrier spy ~ spies dry ~ dried
 b. *ii carry ~ carrying spy ~ spying dry ~ drying
 c. Content word 3-letter minimum bee vs. be inn vs. in buy vs. by two vs. to
 d. Interaction of above die (*dy) ~ dying lie (*ly) ~ lying

Unsurprisingly, therefore, many researchers have argued that Chinese characters also have grammar. In the most ambitious analysis of this sort, Wang (1983) adopted the formal framework of Chomsky & Halle (1968) in accounting for a wide variety of formal observations in character structure and alternations (many of which are substantiated via my own evidence below). Other linguists have taken a more quantitative approach to character constituent combination, notably Stalph (1989) and Kordek (2013). Sproat (2000) included Chinese in his argument that all writing systems have “regular grammars”, expressible via simple chain models. Ladd (2014), building on previous work, argues that characters show duality of patterning (Hockett, 1960).

My own approach is closest to that of Wang (1983), except that I use as little formalism as possible, back up every claim with a wide variety of data, and follow the model of lexical grammars of spoken and signed languages, with separate but interacting modules for CHARACTER MORPHOLOGY (structural operations with interpretations in meaning or pronunciation), CHARACTER PHONOLOGY (regularities in uninterpreted structure, in the silent sense of sign phonology), and CHARACTER PHONETICS (gradient psychophysical influences).

2. Character morphology

CHARACTER MORPHEMES are often obvious, as in the examples in Table 1 above, but not always. Even just for traditional characters alone, estimates of the number of lexically distinct constituents range widely, from 249 (Lui *et al.*, 2010) to 667 (Morioka, 2008). An amusingly ambiguous case is shown in (2) (you need sharp eyes to see the problem). Of course, spoken and signed languages have similarly ambiguous quasi-morphemes, like the notorious nose-related /sn/ formative in English words like *sneeze*, *snore*, *snot*, and *sneer*.

- (2) 鬪 *dòu* ‘struggle’ = 鬥 *dòu* ‘fight’ + 斲 *zhuó* ‘chop’

Productively formed characters are generated via three major MORPHOLOGICAL OPERATIONS very similar to those seen in spoken and signed languages: affixation, compounding, and reduplication. Some of the generally recognized properties of AFFIXATION are listed in (3), illustrated with the contrast between the English words *greenish* (affixation) and *greenhouse* (compounding).

- (3) a. Bound (-ish cannot appear on its own; *house* can)
 b. Closed class (the affix inventory is fixed, but new root morphemes are readily coined)
 c. Semantically bleached (Sweetser, 1988) (-ish has an abstract function, unlike *house*)
 d. Fixed in position (-ish only appears stem-finally; cf. *houseboat*)
 e. May be formally reduced (-ish in *gréenish* is fully unstressed; cf. *gréenhòuse*)

The same five properties also hold of the semantic radicals in semantic-phonetic characters. The set of radicals is fixed (only 214 are recognized in the modern traditional standard set by 康熙字典 *Kāngxī zìdiǎn*, 1716 CE), and as illustrated in (4), several semantic radicals never appear in isolation (many others have bound allomorphs, as we’ll discuss shortly). Radicals also show semantic bleaching, as illustrated in (5) (shells were used as currency in ancient China; see also Handel, 2018).

(4) 丿 ㇇ ㇇ ㇇ ㇇ 彳

- (5) a. 貝 *bèi* ‘shell’ 賬 *zhàng* ‘account (as of banks or bills)’
 b. 口 *kǒu* ‘mouth’ 嗎 *ma* (question particle)

While semantic radicals are not fully fixed in position, they do have strong positional preferences. This is demonstrated in Figure 2 (character counts here and elsewhere are mostly based on the 6,607 characters in the Academia Sinica Balanced Corpus of Modern Chinese [Chen, K. J., *et al.*, 1996], decomposed with the help of the Wikimedia Commons Chinese Character Decomposition page [Wikimedia Commons, 2017]; character type information is mostly from Wiktionary, 2019). The plot makes it clear that most radicals favor one position more than any alternative. Note also that the left edge (black) is by far the most common position, similar to the way that all affixation within a spoken language tends to favor one edge over the other (e.g., suffixation in English and prefixation in Navajo; Matthews, 1991). Positional variation is also systematic, with left-edge radicals favoring the bottom as the secondary option (i.e., the black area is mostly topped by light gray).

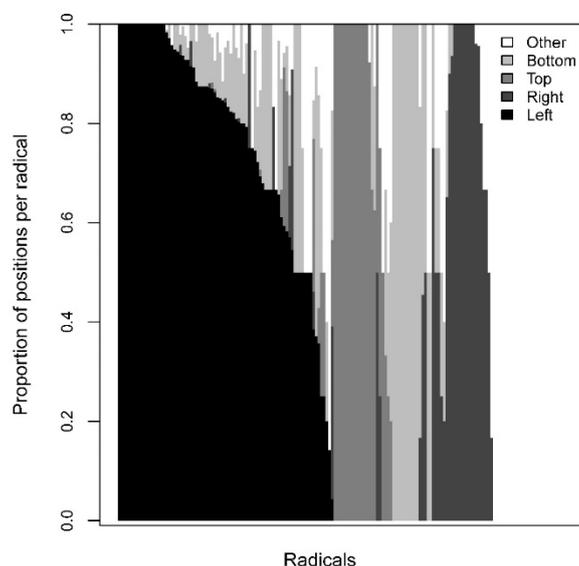


Figure 2. Semantic radical positions

Finally, semantic radicals also show formal reduction, an aspect of character phonology. Some of this reduction applies to non-radicals as well, as long as they appear in locations favored by radicals. The four major regular reduction processes are illustrated in (6), where the first example in each row involves a semantic radical. The stretching process in (6d), which enlarges the bottom-right stroke (while shrinking the constituent body at the left), will later be shown to be consistent with another aspect of character phonology.

- (6) a. DIAGONALIZATION 土：型～地 工：紅～功
 b. DOTTING 火：燙～爛 禾：銖～和
 c. SHRINKING 雨～電 尚～當
 d. STRETCHING 風～颳 支～翅

In addition to these regular phonological processes, many semantic radicals also undergo IDIOSYNCRATIC ALLOMORPHY, which is again associated with specific positions, as illustrated in (7). Moreover, as shown in Table 2, idiosyncratic allomorphy is also more likely to affect a

constituent if it is an “affix” (semantic radical in a semantic-phonetic character) rather than a “root” (in a semantic compound), except at the left edge, which, as we saw above, is the default position for semantic components.

- (7) a. Left: 心：忘～忙
 b. Right: 刀：剪～刻
 c. Top: 艸～花
 d. Bottom: 火：燈～照

Table 2. Radical idiosyncratic allomorphy rate in different positions (per totals)

Radical position	Semantic-phonetic	Semantic compound	Other
Left	46% (3086)	41% (126)	100% (3)
Right	43% (289)	22% (60)	20% (10)
Top	49% (578)	18% (114)	9% (44)
Bottom	13% (410)	8% (125)	18% (44)

In contrast to affixation, COMPOUNDING reflects unrestricted concatenation (Jackendoff, 2010). The formal consequences of this include the paucity of position-dependent idiosyncratic allomorphy in compounds, as seen in Table 2 above. Pragmatics also plays a role, with the positions of the constituents in (8) apparently dependent on the positions of their real-world referents.

- (8) a. 畚 *duō* ‘rice field’ = 水 ‘water’ + 田 ‘field’ (water over field)
 b. 尿 *niào* ‘urine’ = 尸 ‘body’ + 水 ‘water’ (water under body)
 c. 杲 *gǎo* ‘bright’ vs. 杳 *yǎo* ‘dark’ (日 ‘sun’ over vs. under 木 ‘tree’)

Finally, character REDUPLICATION differs functionally from semantic compounding in that the copying itself has meaning, and as illustrated in (9), the range of this meaning is similar to that seen in signed and spoken reduplication (Behr, 2006).

- (9) a. Plurality/abundance: 多 *duō* ‘many’ 品 *pǐn* ‘all sorts’ 蟲 *chóng* ‘insects’
 b. Intensity: 晶 *jīng* ‘glittering’ 炎 *yán* ‘blazing’
 c. Attenuation: 弱 *ruò* ‘weak’

Reduplication in characters also shares a striking formal property with that in signed and spoken languages (McCarthy & Prince, 1998; Sandler & Lillo-Martin, 2006): it is restricted to a small set of specific shapes, as shown in Table 3. The rarest type, the two-by-two square, is arguably derived via compounding of two horizontally or vertically doubled reduplicative structures (cf. 爾器毘). Linear tripling (e.g., 三川靈龠) only appears synchronically within constituents (i.e., it is not morphological at all).

Table 3. Character reduplication shapes

Shape	Count (%)	Examples
Horizontal	47 (43%)	比林單朋雙
Triangular	39 (35%)	品森蟲晶轟
Vertical	20 (18%)	多哥肉昌炎
Square	4 (4%)	覈艸燦巛

3. Character phonology and phonetics

Character phonology is distinct from character phonetics and goes far beyond the bare fact of duality of patterning (regularities in uninterpreted character structure). These points are demonstrated by the central role played by character PROSODY, or holistic character structure. The prosodic TEMPLATE for this structure is binary both horizontally and vertically, where the left and top SLOTS are WEAK (favoring reduction) and the single STRONG slot (favoring prominence) is at the bottom right. These parameters result in the structure in (10) (shown here with symmetry along the vertical axis). Aside from the unique properties ascribable to its two-dimensional nature, this template is formally and functionally similar to prosodic constituents in spoken and signed languages, particularly metrical feet.

$$(10) \begin{bmatrix} W & \\ W & S \end{bmatrix}$$

Templatic analyses for various characters are shown in (11). Note the recursion in (11a-b), the two-slot analysis of the left/top constituents in (11c), and the “exceptional” prosodic structures in (11d) for the few semantic radicals that reduce at the right or bottom (cf. exceptional word-final English stress in *batón*, as compared with regularly stressed *búttón*).

- (11) a. 甫 [S] 浦 [W S] 蒲 $\begin{bmatrix} W \\ [W S]_s \end{bmatrix}$
 b. 烏 [S] 寫 $\begin{bmatrix} W \\ S \end{bmatrix}$ 瀉 $\begin{bmatrix} W \\ [W]_s \\ S \end{bmatrix}$
 c. 厥 居 底 扉 痛 處 $\begin{bmatrix} W \\ W \\ S \end{bmatrix}$
 d. 刻 [S W] or [S] W 熱 $\begin{bmatrix} S \\ W \end{bmatrix}$ or $\begin{bmatrix} S \\ W \end{bmatrix}$

The prosodic template thus helps explain why regular reduction only occurs at the left (diagonalization and dotting) and top (shrinking), both weak positions. Character prosody even has an analogue of WEIGHT (like stress-attracting syllables with long vowels): if tall constituents are “light” along the horizontal axis but “heavy” along the vertical axis, we can explain why they favor the left (W) or bottom (S) positions (recall Figure 2 earlier). Figure 3 illustrates this analysis for the semantic radical 言, as in 說~警. The prosodic template obviously also accounts for reduplication shape: triangular reduplication fills out the entire template, while horizontal and vertical doubling fill out part of it.

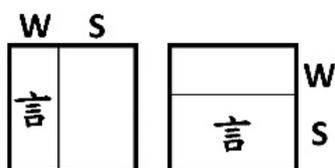


Figure 3. The relationship between the dimensions and positions of radicals

Character prosody plays a central role in stroke shapes as well. The inventory of simple strokes, as has often been observed (e.g., Wang, 1983; Peng, 2017), can be decomposed into a small set of parameters, much like distinctive features in spoken and signed phonology. These parameters describe line orientation, as in Table 4, and modifications to stroke shape like curving and hooking, as in Table 5. Complex strokes can then be analyzed as sequences of simple strokes, as in Table 6.

Table 4. Line orientation parameters for simple strokes

Stroke	Axis	Direction	Examples
丶	None	(default)	太
一	Horizontal	(default)	十
丨	Vertical	(default)	十
㇇	Main diagonal	(default)	木
㇆	Counterdiagonal	Leftward falling	千才
㇇	Counterdiagonal	Rightward rising	子

Stroke images from https://commons.wikimedia.org/wiki/Category:CJK_strokes

Table 5. Modification parameters for simple strokes

Stroke	Axis	Curving	Hooking	Examples
丿	Vertical	Yes	No	川
丨	Vertical	No	Leftward	丁
丨	Vertical	No	Rightward	艮
フ	Horizontal	No	Downward	宀
㇇	Main diagonal	Yes	Upward	戈
㇆	Vertical	Yes	Leftward	彡

See Table 4 for image credit.

Table 6. Complex strokes as sequences of simple strokes

Complex stroke	Simple components	Examples
丿	一 丨	刀
㇇	ノ 一	公
㇆	フ 丨	冫
㇇	丨 一 丨	弓
㇆	一 丨 一 丨	乃

See Table 4 for image credit.

The most obvious way in which strokes reflect prosodic structure is in PROMINENCE, whereby strokes or stroke groups are enlarged in strong positions. Examples are shown in (12), along with the relevant prosodic templates.

- (12) a. Bottom edge: 昌多哥 二土工手車里羊 $\begin{bmatrix} W \\ S \end{bmatrix}$
 b. Right edge: 珏林比炊林 川州 $\begin{bmatrix} W \\ S \end{bmatrix}$

The prosodic template also helps explain stretching, which simultaneously narrows the main body of the semantic radical at the left while lengthening its lower right stroke, as analyzed in (13).

- (13) 走~起 $\begin{bmatrix} W \\ \{S\} \end{bmatrix}$

Even when prominence on horizontal strokes is exceptionally placed elsewhere than at the bottom, it conforms to another prosodic regularity: the avoidance of CLASH, whereby a

weaker prominence is lost when adjacent to a stronger one (as in *refórm* ~ *rèformátion*, where stress moves to avoid clash, as opposed to *elícit* ~ *elicitátion*). Thus in (14a), we see that the exceptionally placed prominent stroke is adjacent to the bottommost stroke, which therefore remains short, but in (14b) the bottommost stroke remains long because the exceptionally lengthened stroke is not adjacent to it.

- (14) a. 士 末 壬 華
b. 聿 幸 重 垂 事

Further evidence for the role of character prosody in stroke form comes from the distribution of CURVING on vertical strokes. Such strokes only appear on the left edge of constituent, that is, in a weak position, as illustrated by the contrast between the curved left-edge strokes in (15a) and the straight non-left-edge vertical strokes in (15b).

- (15) a. 川 介 升 非 拜 弗 爪 月 周 角 用 片 冫 大 九 刀 力 虎 戶 底 右 看
b. 十 中 木 不 平 下 車 年 市 耳 斗

As first observed but not explained by Wang (1983, pp. 203-206), curving is disfavored in wide constituents. This is quantified in Table 7, showing that the straight-stroke variant of 冫 is more likely to appear in wide constituents (bolded examples). Such observations support the prosodic analyses in (16), where curving is blocked in strong positions.

Table 7. The relationship between curving and constituent dimensions

Stroke shape	Dominant axis		
	Horizontal	Vertical	None
Curved		月 甩 周 有 舟 角	丹 用
Straight	冊 冫 岡 巾 內 向 兩 肉 市 冫	再 甬 高 商 喬	同 冫 冉 束

- (16) a. Tall, thin constituent: 月 [W S]
b. Separate stroke groups: 冫 [S][S]
c. Wide constituent: 冊 [S][S]

Not all stroke regularities have prosodic motivations, however. In fact, RIGHTWARD HOOKING looks more like exemplar-driven analogy than phonology per se, since it generally appears only in a curiously specific environment, immediately to the left of crossed strokes, as in (17).

- (17) 氏 民 長 艮 良 衣 喪 辰 派

LEFTWARD HOOKING is even less productive, as seen in the (near) minimal pairs in (18).

- (18) a. 于 丁 事 乎
b. 干 下 聿 平

Nevertheless, as first observed informally by Wang (1983, pp. 206-210), it still conforms to a pair of statistical generalizations: leftward hooking is most common in asymmetrical constituents where the targeted stroke makes contact at the top. This is quantified in Table 8, which cross-classifies constituents with central vertical strokes (note bolded cell).

Table 8. The role of asymmetry and top contact in leftward hooking

		Asymmetrical	Symmetrical
Hooked	Top contact	17 (了...)	5 (丁...)
	No top contact	9 (事...)	1 (小)
Not hooked	Top contact	20 (下...)	18 (平...)
	No top contact	26 (牛...)	28 (中...)

The lack of full productivity is expected in a lexical grammar, and in fact character phonology shows properties typical of LEXICAL PHONOLOGY (Hargus & Kaisse, 1993). Not only do most regularities have lexical exceptions, as we have seen, but they also do not create novel structures (i.e., they are structure-preserving), as illustrated in (19).

- (19) a. Diagonalization: 子~孩 子 *zǐ* ‘child’ vs. 子 *jié* ‘remaining’
 b. Prominence: 日~昌 日 *rì* ‘sun’ vs. 日 *yuē* ‘say’

Character phonology is also sensitive to character morphology, since most of its generalizations apply morpheme-internally (within constituents), as illustrated in (20).

- (20) a. Prominence at constituent edge: 圭 (cf. 土) 順 (cf. 川)
 b. Curving at constituent edge: 所 淵 (cf. 片 兂)

Other phonological patterns are sensitive to the type of morphological operation they interact with. As we saw earlier, idiosyncratic allomorphy virtually only targets affixation (semantic radicals), not other types of constituents. It also does not apply in reduplication, as shown in (21). By contrast, regular reduction affects any type of constituent, including in reduplicative elements, as shown in (22).

- (21) 火: 炎 (cf. 灬: 熱) 水: 淼 (cf. 氵: 江) 手: 龠 (cf. 扌: 拾)

- (22) a. Diagonalization: 珏 竝 銓 孛 鑫 弄
 b. Dotting: 林 炊 林 熯 焱 森

The lexical constraints on character phonology demonstrate that it cannot be reduced entirely to character phonetics. More generally, character phonology is not merely physical: its patterns are even preserved in mechanically produced fonts and are mostly obligatory, even for left-handed writers, and its regularities apply categorically (e.g., strokes are curved precisely at the left edge). The most we can say is that character phonetics *motivates* character phonology, at least diachronically. In particular, a constituent on the left is written before one on the right, and each constituent itself is written from left to right and top to bottom (roughly speaking; see below). Dotting and diagonalization thus have the effect of reducing the distance that the writing instrument needs to move from the lower right of the first constituent to the upper left of the next one. Yet whereas the vertical stroke is last in (23a), the final stroke in the same constituent in (23b) is the diagonalized one. In other words, just as we would expect of the synchronic relationship between phonetics and phonology, stroke order synchronically depends on diagonalization, not the other way around.

- (23) a. 牛
 b. 物

Diachronic phonetic motivations can also be seen in prominence, curving, hooking, and the prosodic template itself, but none explain away synchronic character phonology, which retains its abstract nature. Nevertheless, character phonetics is worth studying in its own right, since it is the domain of stroke direction and stroke order, widely discussed issues in the character literature; Wang (1983) even puts stroke order at the center of his analytical system. Its proper place lies outside grammar, however, because the forces that underlie it are intrinsically psychophysical.

The major influence on stroke direction and order is articulation, specifically in the right hand holding a writing instrument. It is easier to pull a writing instrument than to push it, which is why the strokes in the Chinese character 正 and the Roman letter T are written in exactly the same directions and orders: left to right and top to bottom. Conventional Chinese character stroke order also reduces overall writing distance (Lin, 2014), as seen in Figure 4.



Figure 4. Conventional stroke direction and order in writing 正

Stroke order variation arises when alternative orders have similar levels of articulatory ease. By Chinese convention, the horizontal stroke in a cross is written first, but in some Western writing conventions, including in the US, the vertical stroke is generally written first. As shown in Figure 5, both options yield equally short writing paths. When Goodnow & Levine (1973) tracked the development of the American convention across age, they found that the youngest writers (4 years 5 months) had little preference for one order over the other.

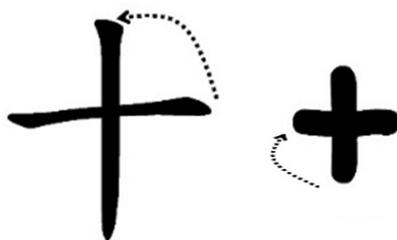


Figure 5. Equally good options for writing a cross

The same variation is also seen within Chinese character writing. The characters in (24) are identical in the traditional and simplified systems, but according to Zhang & Cheung (2013), children and adult foreign learners are taught different stroke orders in Taiwan and the People's Republic of China: in the former, the horizontal stroke is taught as written before the vertical one crossing it, while in the latter the reverse order is taught.

(24) 里 黑 冉 重

Stroke order variation can also arise from competition between articulatory and perceptual constraints. Again according to Zhang & Cheung (2013), the stroke order for both of the constituents in (25) obeys symmetry in the PRC (dots last), whereas in Taiwan, the constituent in (25b) (the idiosyncratic allomorph of the semantic radical 心) instead follows

the articulatorily motivated left-to-right order.

- (25) a. 小
- b. 忸

Since phonetics realizes rather than shapes lexical representations, stroke order may also vary across characters differing in phonological or morphological structures. Particularly striking is the difference in constituent order analyzed by Wang (1983, p. 139) between the characters in (26a-b): in those in (26a) the left/bottom-edge semantic radical is written last, but in (26b) the left/bottom-edge radical is written first. This is because only the second set of radicals derive their left/bottom-edge shape via stretching, from their default forms in (26c). The constituent order in (26a) is thus motivated by visual layout (space-filling elements first) whereas that in (26b) is articulatorily motivated (left to right).

- (26) a. 建道
- b. 趕麵飀魅
- c. 走麥風鬼

4. Corpus-based evidence for character grammar

Quantitative and qualitative corpus analyses provide additional evidence for the productivity of character grammar. For example, Figure 6 shows two productivity models based on the Sinica Corpus (Chen, K. J., *et al.*, 1996), the left one comparing semantic-phonetic characters, semantic compounds, and other character types, and the right one comparing single-edge semantic radical positions in semantic-phonetic characters. Each plot shows the increasing number of new characters (types) revealed as ever more character tokens are sampled in the corpus, yielding growth curves; the dark lines represent the observed data and the dotted and dashed lines extend these by assuming Zipf’s law of lexical frequency distributions (Evert & Baroni, 2007). The growth curves show that semantic-phonetic characters and characters with left-edge semantic radicals are not only the most common but also the most productive, as indicated by the slopes projected to rise beyond the actual corpus size.

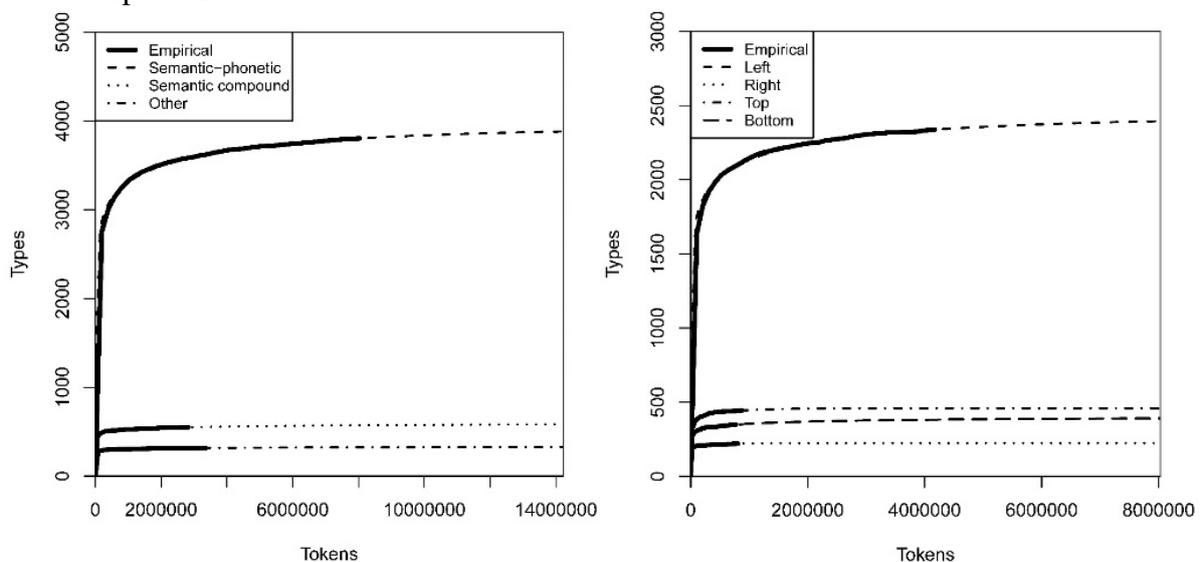


Figure 6. The relative productivity of various character types

We can search further for productivity by expanding our corpus beyond modern traditional characters, starting in the distant past with SMALL SEAL SCRIPT (小篆書 *xiǎozhuànshū*, fl. ca. 200 BCE). Character morphology has changed little since then, though as illustrated in (27a), small seal script did not yet show idiosyncratic allomorphy (X > Y means Y derives historically from X). Horizontal tripling reduplication had not yet become ungrammatical either, as seen in (27b), though it was already rare, and the several other reduplicative shapes attested in even older Chinese scripts had long since disappeared.

- (27) a. ㇀ ~ ㇁ > 心 ~ 悟 ㇂ ~ ㇃ > 水 ~ 泥
 b. ㇄ 'multitude of agricultural laborers' > 眾 *zhòng*

In the character phonology of small seal script, regular reduction was also absent except for some top-edge shrinking, as illustrated in (28a). There was also no bottom-edge or right-edge prominence, as seen in (28b), and curving was either absent or symmetrical on both edges, rather than being restricted to the left, as seen in (28c). Putting all of these observations together, it seems that prosodic phonology, which in modern characters underlies radical reduction and stroke shape, had not yet developed.

- (28) a. ㇅ ~ ㇆ > 木 ~ 根 ㇇ ~ ㇈ > 土 ~ 地 ㇉ ~ ㇊ > 走 ~ 起 ㇋ ~ ㇌ > 竹 ~ 筆
 b. ㇍ > 未 ㇎ > 末 ㇏ > 土 ㇐ > 士 ㇑ > 工 ㇒ > 川
 c. ㇓ > 并 ㇔ > 井 ㇕ > 周 ㇖ > 同

We can also expand our corpus to include the modern simplified system of the PRC (created in part by reviving alternative and obsolete forms, some from calligraphic traditions). Generally speaking, simplified characters continue the trend towards greater morphological systematicity, most obviously in the replacement of many characters with the default semantic-phonetic structure, as in (29) (here X < Y means the replacement of Y by X). Position-dependent idiosyncratic allomorphy has also been extended to further radicals, as shown in (30). The simplification strategies applied to reduplication in (31) show that this morphological operation remains productive as well.

- (29) 听 < 聽 响 < 響 体 < 體 肤 < 膚 惊 < 驚

- (30) 词 ~ 警 < 詞 ~ 警 铅 ~ 釜 < 鉛 ~ 鑿 红 ~ 紫 < 紅 ~ 紫

- (31) a. Preserve template/base or just base: 𧈧 < 𧈨 虫 < 蟲 齿 < 齒
 b. Preserve template abstractly: 𧈩 < 𧈪 双 < 雙 𧈫 < 聶 𧈬 < 轟
 c. Template applied to new cases: 众 < 眾 宫 < 宮 网 < 網

As with small seal script, however, character phonology in simplified characters differs a bit more from the traditional character system. While prominence is productively extended in (32a), the curved/straight contrast ignores changes in constituent width, as seen in (32b). Rightward hooking no longer depends on crossed strokes to the right, as seen in (32c), though consistent with the generalization of Wang (1983), leftward hooking is sometimes triggered when constituents become asymmetrical, as in (32d). Nevertheless, character phonetics remains synchronically dependent on character phonology, with the simplified constituent in (32e) reversing stroke order when the lowest stroke is diagonalized.

- (32) a. 来 < 來 金 < 僉
 b. 贝 < 貝 门 < 門 风 < 風
 c. 长 < 長
 d. 东 < 東 (cf. 书 < 書 车 < 車)
 e. 车~较 < 車~較

Character grammar also helps in understanding variation across fonts. For example, the brush-pen-like regular script (楷書 *kǎishū*) applies idiosyncratic allomorphy more readily than sans serif fonts (黑體 *hēitǐ*), as seen in (33).

- (33) a. 糸~絲 示~神~禮
 b. 糸~絲 示~神~禮

Fonts can also differ in morphological decomposition, as revealed by the absence of prominence clash in regular script in (34a) and its presence in Song script (宋體 *Sòngtǐ*) in (34b), as if in the latter the adjacent prominent strokes are in separate constituents.

- (34) a. 美 ≠ 羊 + 大
 b. 美 = 羊 + 大

Variation in actual handwriting also reveal active knowledge of character morphology. As observed by Wang (1983, pp. 129-134), writers sometimes “raise” radicals to make them conform better to the canonical left-edge-radical structure, as in (35a). Among the slips of the pen collected by Moser (1991) are several that generate novel characters by misplacing constituents, as in (35b) (his Ex. 85, p. 32). The idiosyncratic avoidance of prominence clash in (35c) (from Ch’en *et al.*, 1989, p. 73) suggests that the writer considered this character monomorphemic.

- (35) a. 哲 → 摺 塗 → 塗
 b. 真没意思 *zhēn méi yìsī* ‘really uninteresting’ → 真没畜
 c. 黑 → 黑

Finally, when writers invent novel characters for metalinguistic purposes, they typically follow canonical semantic-phonetic structure, though often with an interpretation more like that of semantic compounds. Examples include the joke riffing on a TV commercial in (36a) (Mair, 2015), the failed attempt to avoid a defamation charge in (36b) (Yan, 2018), and the visual/auditory pun from the Hong Kong umbrella protests in (36c) (Ho, 2014).

- (36) a. 蹩 *duāng* (nonce interjection) (cf. 成龙 *Chéng Lóng* ‘Jackie Chan’)
 b. 蒹蒹 (cf. 妓女 *jì nǚ* ‘prostitute’)
 c. 揸 (cf. 才 ‘hand’ [手], 伞 *saan*³ ‘umbrella’, 撐 *caang*¹ ‘support’)

5. Experimental evidence for character grammar

Experiments on reading and writing are essential in establishing that character grammar is psychologically real (e.g., Liu & Wu, 2017), involves productive knowledge (see the many studies using fake characters described below), and is abstract in being non-iconic (e.g., Xiao

& Treiman, 2012, found that only 15 out of 213 simple characters have meanings guessable by non-readers of Chinese), amodal (e.g., Tan *et al.*, 2005, is one of several studies showing that the input/output systems are closely linked in character learning), and modular (the Visual Word Form Area is also essential in Chinese reading; e.g., Liu *et al.*, 2008).

Many experiments that have ostensibly focused on the real-time processing of character reading and writing actually also provide evidence of morphological knowledge. For example, when Chen, Y. P., *et al.* (1996) asked simplified Chinese readers to make same/different judgments for character pairs like those in Table 9, readers were slower to respond to characters with more constituents, suggesting automatic decomposition. Chen & Cherng (2013) found that characters are automatically decomposed in writing as well, since the writing of a traditional character was initiated faster if it had earlier been presented in a set of characters starting (in writing) with the same constituent. In both studies, the interpretability of the constituents (in meaning or pronunciation) did not seem to matter.

Table 9. Sample materials in a character difference detection task

Character type	Different pairs	Same pairs
Two constituents	裘 杂	雪 雪
Three constituents	喝 唱	读 读

Other experiments demonstrate that character constituents can be interpreted, however. Feldman & Siok (1999) presented simplified character readers with genuine characters and fake but well-formed characters and asked them to decide which was which. Each target was preceded by a prime character that did or did not share the same semantic radical and did or did not share a related whole-character meaning, as illustrated by the examples in Table 10. A key result was that lexical decisions were slowed if radical meaning and whole-character meaning mismatched across prime-target pairs (as in the bolded character, which contains *讠* but does not relate to speech).

Table 10. Sample materials in a primed lexical decision experiment

Target: 论 <i>lùn</i> ‘discuss’		Semantic radicals	
		Same	Different
Character meanings	Related	评 <i>píng</i> ‘appraise’	述 <i>shù</i> ‘narrate’
	Unrelated	诸 <i>zhū</i> ‘various’	竿 <i>gān</i> ‘pole’

Phonetic components may also be interpreted in terms of pronunciation (e.g., Lee *et al.*, 2006). Unlike affix-like semantic radicals, however, they are more stem-like in that readers seem to expect them to form an open class, with fake characters rejected in lexical decisions at roughly the same speed whether or not their phonetic components are actually used as such in real characters (Mattingly & Hsiao, 1999).

Character phonology has been studied much less frequently in the experimental literature. A rare exception is Myers (2016), where traditional Chinese readers were asked to make acceptability judgments on a set of fake characters containing reduplicative structures crossing lexical status (i.e., whether or not the exact reduplicative form appears in real characters) with grammatical status (i.e., whether or not the reduplicative shape fits the prosodic template), as illustrated in Table 11. The major finding was that two factors affected judgments independently. Moreover, reanalyses conducted after the study’s publication (see Myers, 2019, for details) showed that both factors affected judgments at almost the same time, as shown in the graphs in Figure 7, which plot the cumulative risk over time of giving a “yes” response (Scheike & Zhang, 2011), with the vertical line marking when the 95%

confidence band first rises above the baseline (after Baayen & Blanche, 2017).

Table 11. Sample materials for a fake character acceptability task

Shape	Grammatical		Ungrammatical	
	Lexical	Non-lexical	Lexical	Non-lexical
Horizontal	棚	襟	棚	襟
Vertical	棧	璵	棧	璵
Triangular	忿	筵	忿	筵

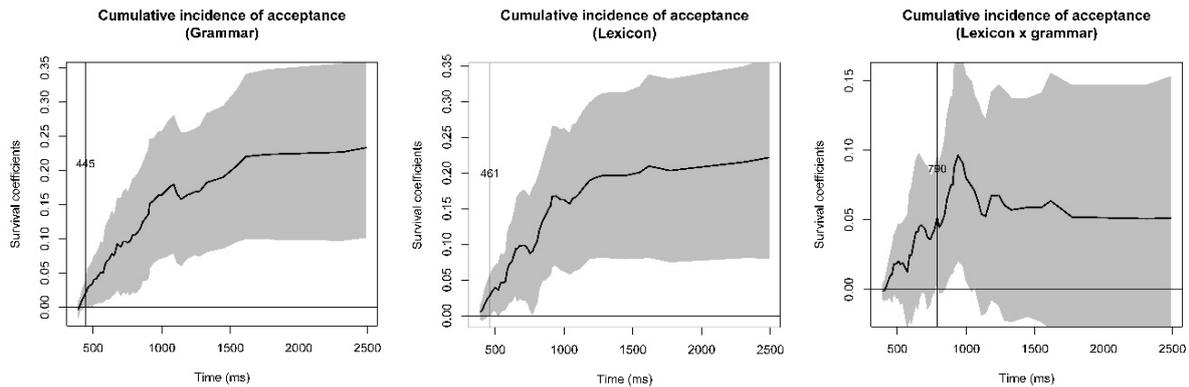


Figure 7. The time course of grammatical and lexical influences on character acceptability

Another experiment (reported in Myers, 2019) asked traditional Chinese readers to rate the three-stroke combinations on the left of Figure 8; the plot on the right shows the results just for the eight items with hooked strokes. These results show that not only do readers know that leftward hooks belong on the right edge, but consistent with the observations of Wang (1983), they also prefer this type of stroke to have topping material.

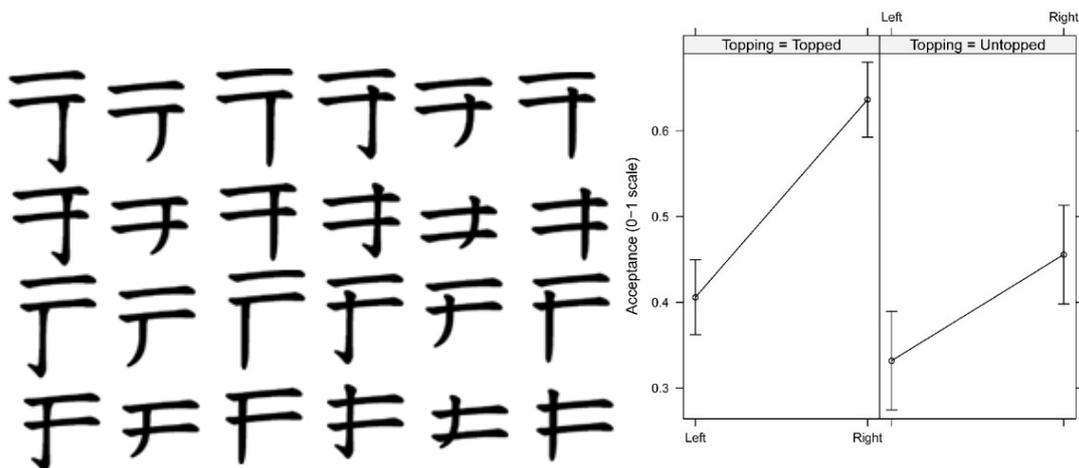


Figure 8. Stimuli and some results in a fake character acceptability task

6. Implications and applications

In this final section, we ask what the above evidence implies and how it might be made useful. Regarding the first point, there seem to be four general types of explanations for Chinese character grammar. The first notes the special cultural and linguistic environment in which the system evolved. Non-logographic systems also have visual grammars (e.g., Evertz, 2018), but with far fewer graphemes, a linear rather than two-dimensional structure, and

relatively direct interpretations in speech, these grammars are significantly less rich. Moreover, while other logographic systems also had grammars, including affix-like semantic markers (e.g., determinatives in Egyptian hieroglyphs: Ritner, 1996), the fact that Chinese writing is the sole survivor of this orthographic type may result from a unique confluence of historical factors, in particular the syllable-based phonology and root-based morphology of the spoken language, and the centralized and isolated nature of its cultural context.

As we have already hinted in earlier sections, a second class of explanations invokes very general motoric, perceptual, and cognitive constraints. For example, character “prosody” may relate to the parallel processing pathways for fine-grained versus coarse-grained information (e.g., for vision: Yamaguchi *et al.*, 2000). As Figure 9 illustrates, the coarse-grained mode highlights reduplicative structure, whereas the fine-grained mode makes the base elements more salient.

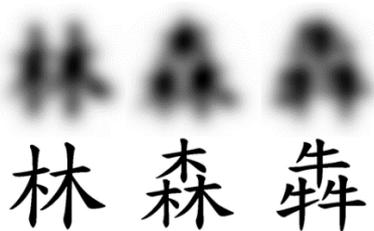


Figure 9. Coarse-grained and fine-grained views of reduplicative structures

A third class of explanations posits that the human brain evolved to process language at a level more abstract than any specific modality, an idea already consistent with the existence of natural sign languages. One argument in favor of this view is that people readily learn grammars in static two-dimensional visual displays (e.g., Pothos & Bailey, 2000). One argument against it is that the brain region most clearly associated with orthography (the Visual Word Form Area) is adjacent to the visual cortex and far from Broca’s area.

Finally, Chinese character grammar may derive from the mathematical laws (of the sort speculated on in Chomsky, 2005, and elsewhere) that underlie all complex systems, an idea independently applied to Chinese characters by Abler (2005) and Zhang (2006). However, whether acausal theories feel explanatory seems to depend on one’s personality (Myers, 2012), and the fact that human-like grammar is clearly not inevitable suggests that a full explanation is likely to require more than a few elegant equations.

Regardless of how we explain it, Chinese character grammar may still have real-world applications. For example, educators of young children should recognize that as with spoken (and signed) languages, mastering an orthographic system depends a lot on the child making his or her own (perhaps unconscious) discoveries, and in fact some children do seem to benefit when their own orthographic explorations are encouraged (e.g., Chan *et al.*, 2008). Since character grammar, like all grammar, is amodal, beginning readers also need to gain experience as writers (e.g., Tan *et al.*, 2005). More specifically, children need to learn character decomposition (e.g., Anderson *et al.*, 2013), the positional preferences of semantic radicals (e.g., Tsai & Nunes, 2003), the meaning of semantic radicals (e.g., Hung *et al.*, 2010), the pronunciation of phonetic components (e.g., Li *et al.*, 2018), and what makes novel stroke groups well-formed (e.g., Liu, 2013). However, one highly-drilled aspect of character learning, stroke order, may deserve a rethink, given that it belongs more to phonetics than to grammar per se. For example, in their study of putative stroke order errors in children learning to write traditional characters, Law *et al.* (1998) discuss several examples that actually reveal variation in character decomposition (character morphology).

Teaching characters to adult foreign students may also benefit from a grammatical perspective. As with learning a new spoken or signed language, adults need to suppress their prior knowledge when learning a new orthography (Bassetti, 2013), but adults also have

greater powers of meta-linguistic awareness than children, allowing them to benefit more from explicit grammatical analyses (as Li, 2015, argues for teaching Chinese more generally). Adults also do not need to learn characters strictly in order from least to most frequent, helping them avoid being overly exposed to irregular characters, which tend to have higher token frequencies (as with *ate* vs. *ingested* in English); this strategy is explicitly advocated by Tollini (1994) for teaching kanji to adult foreign learners of Japanese.

In clinical linguistics, the grammatical view helps explain the essential differences between developmental dyslexia in children and acquired dyslexia and agraphia in adults. In children, it is the learning of character grammar that is impaired, as shown by abnormal responses to well-formed but nonlexical characters (e.g., Tzeng *et al.* 2018), whereas in adults, character grammar is spared and evercompensates for deficits in lexical access, as shown by overreliance on phonetic components in pronunciation (e.g., Yin *et al.*, 2005) and constituent replacement errors in writing (e.g., Han *et al.*, 2007).

Finally, computational linguists should also be reminded of the value of grammar, which once lay at the heart of reading models (e.g., Dai *et al.*, 2007) but has been increasingly neglected with the growing power of analogical “deep learning” networks (Schmidhuber, 2015). However, even with superhuman amounts of training, such models still tend to be bizarrely inflexible in comparison to humans (Waldrop, 2019). For example, to learn just 1,000 Chinese characters, Cireşan *et al.* (2012) trained a network with hundreds of thousands of nodes on over a quarter of a million tokens for 14 hours; training was later sped up via the “starting small” strategy of Elman (1993), but this involved simplifying the initial training set in a way not seen in natural language acquisition. Nevertheless, as our best current models of actual brain function, artificial neural networks also have great potential in grammatical research (e.g., Silfverberg & Hulden, 2018), among other things helping to extract patterns that are not obvious to the (conscious) human eye.

I therefore believe that the notion of character grammar is not only empirically plausible and theoretically important, but also something that can be taken in any number of promising directions: analytical, historical, comparative, corpus-based, experimental, pedagogical, clinical, computational. Far from being the last word on the subject, Myers (2019) is intended to encourage continuing exploration of this rich area.

References

- Abler, W. L. (2005). *Structure of matter, structure of mind*. Sofia: Pensoft.
- Anderson, R. C., Ku, Y. M., Li, W., Chen, X., Wu, X., & Shu, H. (2013). Learning to see the patterns in Chinese characters. *Scientific Studies of Reading*, 17(1), 41-56.
- Baayen, H., & Blanche, P. (2017). Dynamic survival analysis. Talk presented at *The Workshop on New Methods in Statistics*. Tübingen, Germany, 19-20 January 2017.
- Bassetti, B. (2013). Bilingualism and writing systems. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism and multilingualism*, 2nd ed (pp. 649-670). Hoboken, NJ: Wiley-Blackwell.
- Behr, W. (2006). Homosomatic juxtaposition and the problem of ‘syssymantic’ (*huìyì*) characters. In F. Bottéro & R. Djamouri (Eds.), *Écriture chinoise: données, usages et représentations* (pp. 75-114). Paris: École des hautes études en sciences sociales, Centre de recherches linguistiques sur l’Asie orientale.
- Berent, I., & Dupuis, A. (2018). The unbounded productivity of (sign) language. *The Mental Lexicon*, 12(3), 309-341.
- Bloomfield, L. (1933). *Language*. New York: Holt.

- Brentari, D. (2011). Sign language phonology. In J. Goldsmith, J. Riggle, J., & A. C. L. Yu (Eds.), *The handbook of phonological theory, second edition* (pp. 691-721). Hoboken, NJ: Wiley-Blackwell.
- Ch'en, T. T., Link, P., Tai, Y. J., & Tang, H. T. (1989). *Chinese primer: Character workbook*. Cambridge, MA: Harvard University Press.
- Chan, L., Cheng, Z. J., & Chan, L. F. (2008). Chinese preschool children's literacy development: From emergent to conventional writing. *Early Years*, 28(2), 135-148.
- Chen, J. Y., & Cherng, R. J. (2013). The proximate unit in Chinese handwritten character production. *Frontiers in Psychology*, 4, Article 517.
- Chen, K. J., Huang, C. R., Chang, L. P., & Hsu, H. L. (1996). Sinica Corpus: Design methodology for balanced corpora. *Proceedings of the 11th Pacific Asia Conference on Language, Information and Computation (PACLIC 11)*, 167-176.
- Chen, Y. P., Allport, D. A., & Marshall, J. C. (1996). What are the functional orthographic units in Chinese word recognition: The stroke or the stroke pattern? *The Quarterly Journal of Experimental Psychology*, 49(4), 1024-1043.
- Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, 36(1), 1-22.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.
- Cireşan, D. C., Meier, U., & Schmidhuber, J. (2012). Transfer learning for Latin and Chinese characters with deep neural networks. *The 2012 International Joint Conference on Neural Networks (IJCNN)*, IEEE. 1-6.
- Dai, R., Liu, C., & Xiao, B. (2007). Chinese character recognition: history, status and prospects. *Frontiers of Computer Science in China*, 1(2), 126-136.
- Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, 15(6), 254-262.
- Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48(1), 71-99.
- Emmorey, K. (2001). *Language, cognition, and the brain: Insights from sign language research*. London: Psychology Press.
- Evert, S., & Baroni, M. (2007). zipfR: Word frequency distributions in R, *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Posters and Demonstrations Sessions*, 29-32.
- Evertz, M. (2018). *Visual prosody: The graphematic foot in English and German*. Berlin: Walter de Gruyter.
- Feldman, L. B., & Siok, W. W. (1999). Semantic radicals contribute to the visual identification of Chinese characters. *Journal of Memory and Language*, 40(4), 559-576.
- Frishberg, N. (1975). Arbitrariness and iconicity: Historical change in American Sign Language. *Language*, 51(3), 696-719.
- Goodnow, J. J., & Levine, R. A. (1973). 'The grammar of action': Sequence and syntax in children's copying. *Cognitive Psychology*, 4(1), 82-98.
- Han, Z., Zhang, Y., Shu, H., & Bi, Y. (2007). The orthographic buffer in writing Chinese characters: Evidence from a dysgraphic patient. *Cognitive Neuropsychology*, 24(4), 431-450.
- Handel, Z. (2018). The cognitive role of semantic classifiers in modern Chinese writing as reflected in neogram creation, University of Washington ms.
- Handel, Z. (2019). *Sinography: A cross-linguistic study of the borrowing and adaptation of the Chinese script*. Leiden: Brill.
- Hargus, S., & Kaisse, E. M. (Eds.) (1993). *Studies in lexical phonology*. Cambridge, MA: Academic Press.

- Ho, T. L. M. (2014). Letters from Hong Kong: Characters under the Cantonese umbrella. *Asian Review of Books*. <http://www.asianreviewofbooks.com/pages/?ID=2053>.
- Hockett, C. F. (1960). The origin of speech. *Scientific American*, 203(3), 88-96.
- Householder, F. W. (1971). The primacy of writing. In F. W. Householder (Ed.), *Linguistic speculations* (pp. 244-264). Cambridge, UK: Cambridge University Press.
- Huang, D. K. [黃德寬] (2003). 汉字构形方式的动态分析 [Dynamic analysis of the formation of Chinese characters]. *安徽大学学报(哲学社会科学版)* [Journal of Anhui University (Philosophy and Social Sciences)], 27(4), 1-8.
- Hung, K. C. [洪國鈞], Lee, S. H. [李姝慧], Chen, S. Y. [陳修元] & Chou, T. L. [周泰立] (2010). 語意部件與關聯強度對成人與國小五年級孩童漢字語意處理效應的差異 [Effects of semantic radical and semantic association on semantic processing of Chinese characters for adults and fifth graders]. *中華心理學刊* [Chinese Journal of Psychology], 52(3), 327-344.
- Jackendoff, R. (2010). The ecology of English noun-noun compounds. In R. Jackendoff (Ed.), *Meaning and the lexicon* (pp. 413-451). Oxford: Oxford University Press.
- Kordek, N. (2013). *On some quantitative aspects of the componential structure of Chinese characters*. Poznań: Wydawnictwo Rys.
- Ladd, D. R. (2014). *Simultaneous structure in phonology*. Oxford: Oxford University Press.
- Law, N., Ki, W. W., Chung, A. L. S., Ko, P. Y., & Lam, H. C. (1998). Children's stroke sequence errors in writing Chinese characters. *Reading and Writing: An Interdisciplinary Journal*, 10, 267-292.
- Lee, C. Y., Tsai, J. L., Chiu, Y. C., Tzeng, O. J. L., & Hung, D. L. (2006). The early extraction of sublexical phonology in reading Chinese pseudocharacters: An event-related potentials study. *Language and Linguistics*, 7(3), 619-636.
- Li, A. Y. H. (2015). Facilitating language learning: A generative perspective. In H. Tao, Y. H. Lee, D. Su, K. Tsurumi, W. Wang & Y. Yang (Eds.), *Proceedings of the 27th North American Conference on Chinese Linguistics, vol. 1*, 1-21, UCLA, Los Angeles.
- Li, L., Wang, H. C., Castles, A., Hsieh, M. L., & Marinus, E. (2018). Phonetic radicals, not phonological coding systems, support orthographic learning via self-teaching in Chinese. *Cognition*, 176, 184-194.
- Lin, G. S. [林桂生] (2014). 汉字笔顺规范的理论是路径法 [Chinese character stroke order theory is the path method]. *黄冈师范学院学报* [Journal of Huanggang Normal University], 34(1), 91-96.
- Liu, C., Zhang, W. T., Tang, Y. Y., Mai, X. Q., Chen, H. C., Tardif, T., & Luo, Y. J. (2008). The visual word form area: evidence from an fMRI study of implicit processing of Chinese characters. *Neuroimage*, 40(3), 1350-1361.
- Liu, I. M., & Wu, J. T. (2017). Reading characters and words, behavioral studies. In R. Sybesma, W. Behr, Y. Gu, Z. Handel, C. T. J. Huang & J. Myers (Eds.), *Encyclopedia of Chinese language and linguistics, vol. III* (pp. 532-536). Leiden: Brill.
- Liu, M. [劉敏] (2013). 中文兒童組字規則覺識發展 [Development of Chinese character orthography awareness of Taiwanese children], MA thesis, Taipei Municipal University of Education.
- Lui, H. M., Leung, M. T., Law, S. P., & Fung, R. S. Y. (2010). A database for investigating the logographeme as a basic unit of writing Chinese. *International Journal of Speech-Language Pathology*, 12(1), 8-18.
- Mair, V. (2015). Duang, *Language Log*, March 1, <http://languagelog.ldc.upenn.edu/nll/?p=17913>.
- Matthews, P. H. (1991). *Morphology, 2nd ed.* Cambridge, UK: Cambridge University Press.

- Mattingly, I. G., & Hsiao, P. L. (1999). Are phonetic elements in Chinese characters drawn from a syllabary? *Psychologia*, 42(4), 281-289.
- McCarthy, J. J., & Prince, A. (1998). Prosodic morphology. In A. Spencer & A. M. Zwicky (Eds.), *The handbook of morphology* (pp. 283-305). Oxford: Blackwell.
- McCawley, J. D. (1994). Some graphotactic constraints. In W. C. Watt (Ed.), *Writing systems and cognition: Perspectives from psychology, physiology, linguistics, and semiotics* (pp. 115-127). Berlin: Springer.
- Morioka, T. (2008). CHISE: Character processing based on Character ontology. In T. Tokunaga & A. Ortega (Eds.), *Large-scale knowledge resources. Construction and application* (pp. 148-162). Berlin: Springer.
- Moser, D. (1991). Slips of the tongue and pen in Chinese. *Sino-Platonic Papers*, 22, 1-45.
- Myers, J. (2012). Cognitive styles in two cognitive sciences. In N. Miyake, D. Peebles & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, 2067-2072. Austin, TX: Cognitive Science Society.
- Myers, J. (2016). Knowing Chinese character grammar, *Cognition*, 147, 127-132.
- Myers, J. (2019). *The grammar of Chinese characters: Productive knowledge of formal patterns in an orthographic system*. London: Routledge.
- Naka, M. (1998). Repeated writing facilitates children's memory for pseudocharacters and foreign letters. *Memory & Cognition*, 26(4), 804-809.
- Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General*, 130(3), 401-426.
- Padden, C. A., & Perlmutter, D. M. (1987). American Sign Language and the architecture of phonological theory. *Natural Language & Linguistic Theory*, 5(3), 335-375.
- Peng, X. (2017). Stroke systems in Chinese characters: A systemic functional perspective on simplified regular script. *Semiotica*, 2017(218), 1-19.
- Pothos, E. M., & Bailey, T. M. (2000). The role of similarity in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(4), 847-862.
- Ritner, R. K. (1996). Egyptian writing. In P. T. Daniels & W. Bright (Eds.), *The world's writing systems* (pp. 73-84). Oxford: Oxford University Press.
- Sandler, W., & Lillo-Martin, D. (2006). *Sign language and linguistic universals*. Cambridge, UK: Cambridge University Press.
- Scheike, T. H., & Zhang, M. J. (2011). Analyzing competing risk data using the R timereg package. *Journal of Statistical Software*, 38(2), 1-15.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85-117.
- Silfverberg, M., & Hulden, M. (2018). Initial experiments in data-driven morphological analysis for Finnish. In T. A. Pirinen, M. Rießler, J. Rueter & T. Trosterud (Eds.), *Proceedings of the Fourth International Workshop on Computational Linguistics of Uralic Languages*, 100-107, Association for Computational Linguistics, Stroudsburg, PA.
- Sproat, R. (2000). *A computational theory of writing systems*. Cambridge, UK: Cambridge University Press.
- Stalph, J. (1989). *Grundlagen einer Grammatik der sinojapanischen Schrift* [Foundations of a grammar of the Sino-Japanese script]. Wiesbaden: Harrasowitz Verlag.
- Sweetser, E. E. (1988). Grammaticalization and semantic bleaching. In K. Lambrecht, S. Axmaker, A. Jaisser & H. Singmaster (Eds.), *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, 389-405, Berkeley Linguistics Society, Berkeley.

- Tan, L. H., Spinks, J. A., Eden, G. F., Perfetti, C. A., & Siok, W. T. (2005). Reading depends on writing, in Chinese. *Proceedings of the National Academy of Sciences*, 102(24), 8781-8785.
- Tollini, A. (1994). The importance of form in the teaching of kanji, 世界の日本語教育 [Japanese Education in the World], 4, 107-116.
- Tsai, K. C., & Nunes, T. (2003). The role of character schema in learning novel Chinese characters. In C. McBride-Chang & H.-C. Chen (Eds.), *Reading development in Chinese children* (pp. 109-125). London: Praeger.
- Tzeng, Y. L., Hsu, C. H., Lin, W. H., & Lee, C. Y. (2018). Impaired orthographic processing in Chinese dyslexic children: Evidence from the lexicality effect on N400. *Scientific Studies of Reading*, 22(1), 85-100.
- Waldrop, M. M. (2019). What are the limits of deep learning? *Proceedings of the National Academy of Sciences*, 116(4), 1074-1077.
- Wang, J. C. S. (1983). *Toward a generative grammar of Chinese character structure and stroke order*. Ph.D. thesis, University of Wisconsin-Madison.
- Wikimedia Commons. (2017). *Chinese character decomposition*. https://commons.wikimedia.org/wiki/Commons:Chinese_characters_decomposition.
- Wiktionary. (2019). *Wiktionary*. <https://www.wiktionary.org/>.
- Xiao, W., & Treiman, R. (2012). Iconicity of simple Chinese characters. *Behavior Research Methods*, 44(4), 954-960.
- Yamaguchi, S., Yamagata, S., & Kobayashi, S. (2000). Cerebral asymmetry of the 'top-down' allocation of attention to global and local features. *Journal of Neuroscience*, 20(9), Article RC72.
- Yan, H. J. [顏宏駿] (2018). 自創文字罵人「妓女」判拘 30 天 [Person who invented characters to call somebody a prostitute receives 30-day sentence]. 自由時報 [Liberty Times], March 20, <http://news.ltn.com.tw/news/society/paper/1185367>.
- Yin, W., He, S., & Weekes, B. S. (2005). Acquired dyslexia and dysgraphia in Chinese. *Behavioural Neurology*, 16(2-3), 159-167.
- Yip, P. C. (2000). *The Chinese lexicon: A comprehensive survey*. Abingdon: Routledge.
- Zhang, H. Y. (2006). The evolution of genomes and language. *EMBO Reports*, 7(8), 748-749.
- Zhang, X. [張小衡] & Cheung, W. K. [張煥淇] (2013). 兩岸漢字規範筆順比較 [A Mainland-Taiwan comparative study on standard stroke order of Chinese characters]. 中國語文通訊 [Newsletter of Chinese Language], 92 (1), 17-26.