Synchronic and diachronic ordering in Chinese character form patterns

James Myers https://orcid.org/0000-0002-3866-1969

Graduate Institute of Linguistics
National Chung Cheng University
168 University Road, Min-Hsiung
Chia-Yi 62102
Taiwan

Abstract

This study demonstrates that interactions among patterns in modern Chinese character form cannot be ascribed to the diachronic order in which the patterns arose. In particular, lexically idiosyncratic patterns and articulatorily or perceptually motivated patterns interact in the modern system as if ordered early and late, respectively, regardless of when they arose historically. The results help confirm that, just as with spoken and signed languages, historical change in writing systems involves a sequence of synchronic grammars, each restructured from the previous one in accordance with universal synchronic principles.

Keywords

Chinese script, orthographic grammar, rule ordering, diachronic change, productivity

Acknowledgements

Thanks to Taiwan's National Science and Technology Council (Taiwan) grant 112-2410-H-194-030-MY3, Yangru Chen, Jane Tsay, the conference audience, and of course Yannis Haralambous for making it all possible.

Presented at Grapholinguistics in the 21st Century, October 2024.
(https://grafematik2024.sciencesconf.org/)

1. Introduction

As we all know, ontogeny recapitulates phylogeny. Ernst Haeckel's aphorism makes biological sense because evolutionary changes that target earlier stages in embryonic development risk mucking up the whole developmental process; tacking something on at the end is much safer. Humans don't need gills, but trying to grow lungs without starting with gills is likely to leave you without working lungs at all. Similarly, historical sound change tends to leave sequential traces in synchronic systems, with older changes preserved as "earlier" phonological rules (King, 1973). Again this makes sense: adding a new phonetic flourish to your accent is likely to leave your speech more intelligible than distorting it morphophonologically.

Do writing systems conform to the same principle? Before attempting to answer this, we must first confront the fact that Haekel's principle leaks, even in biology; evolution can tweak embryonic development in a wide variety of ways unattested in the fossil record, as long as it all works out in the end, an insight that is part of what has become known as evo-devo (Love, 2024). Phonological rules need not exactly match the course of sound change either, as we'll see shortly. Thus we should not be surprised if writing systems also show misalignments between diachronic and synchronic order. After all, diachrony also has to make synchronic sense, whether in embryos, speech, or writing, and ultimately we would like to develop phonological and orthographic versions of evo-devo that explain why and where history and grammar do and do not match. Pushing linguistic theory in this direction, however, would require us to be relatively concrete in our examples and a bit more sophisticated in our analyses.

To this end, this paper focuses on the history and modern form of Chinese characters, a system that I have been exploring for several years now from a grammar-oriented perspective (Myers, 2016, 2019, 2021a, 2021b, 2022, 2024). In the rest of this paper, I first review the nature of ordering in phonology and its relation to sound change (section 2), then discuss a selection of form patterns in Chinese characters and argue for their synchronic ordering (section 3), and finally explore the relationship between this ordering and the historical evolution of these patterns (section 4). Section 5 offers some brief conclusions.

2. Ordering in phonology and sound change

Sound changes happen in real time, the same sort of time that historians and biologists and physicists talk about, whereas synchronic rule ordering is an abstract and controversial concept that only a subset of linguists has ever believed in. This might lead one to think that we, as putative scientists, should aim at explaining apparent synchronic ordering (the mysterious) in terms of diachronically ordered sound changes (the well-understood). However, as I hope to make seem plausible in this section, not only is historical sound change intrinsically insufficient to explain interactions within phonological systems, but phonological history itself is better understood in phonological terms, that is, in terms of grammar. These are not new ideas, but they underlie what we'll do when we come back to Chinese characters later in the paper.

We can start with King (1973), who argued for what remains the conventional view: synchronic rule ordering, by and large, follows the order of historical sound change. However, a lot depends on what we mean by "rule ordering," not to mention "by and large." King himself made the conventional caveat that phonology is universally ordered before phonetics, which allows new rules to be inserted between the two modules, including before old phonetics. Since then, of course, phonological theory (at least one of its major dialects) has abandoned the notions of rules and rule ordering in favor of the simultaneously applied

constraints of Optimality Theory (OT; Prince & Smolensky, 2004). The inability of OT to handle certain types of constraint interactions, however, eventually forced its adherents to admit a certain amount of ordering back into the theory (see review in Pruitt, 2023). In this paper I will adopt the ordering approach known as Stratal OT (Kiparsky, 2015), rather than its major rival, harmonic serialism (McCarthy, 2007; Hauser & Hughto, 2020), for two reasons. First, Stratal OT was developed with both diachronic and synchronic phonology in mind, and second, its notion of universally ordered strata explicitly implements King's conventional caveat: phonology (what Stratal OT calls stem and word phonology) applies before phonetics (phrasal phonology).

To explain why the caveat is necessary, I will unpack a notorious case originally based on observations by Joos (1942) and first analyzed in terms of rule ordering by Halle (1962): diphthong raising and flapping in Canadian English. Similar phenomena exist in other English varieties, including my own Midwest American accent, so I can personally vouch for the examples below. Diphthong raising applies before voiceless consonants, as shown in (1a), and flapping, which applies before unstressed syllables, voices coronal consonants, as shown in (1b). This means that flapping could potentially "bleed" raising (in the terminology pioneered by Kiparsky, 1968), that is, remove its conditioning environment by voicing the voiceless coronal that would otherwise trigger it. What actually happens, however, is that raising "counterbleeds" flapping, thereby "overapplying" (in the terminology of McCarthy, 1999), as in (1c): the diphthong is raised before what would otherwise be voiceless if it weren't for the effects of flapping. In other words, the interaction here is "opaque": the effects of raising are obscured by the effects of flapping.

(1) a. write [ɹʌit] ride [ɹaid]
b. late [leit] later [leiɾɚ]
c. writer [ɹʌiɾɚ] rider [ɹaiɾɚ]

As Halle (1962) pointed out, this interaction is just what we would expect if raising applies synchronically before flapping, as demonstrated in (2).

(2)

	writer	rider
Underlying	ɹaitɚ	ɹaidɚ
Raising	ɹʌitɚ	--
Flapping	ɹʌiɾɚ	ɹaiɾɚ
Surface	ɹʌiɾɚ	ɹaiɾɚ

If this rule ordering simply fossilizes the effects of diachronic sound change, raising must have emerged at an earlier date than flapping. This seems plausible at first, since immigrants from Scotland played an outsized role in Canadian history (the first prime minister was named Macdonald after all), and Scottish English has a phonological pattern quite similar to Canadian raising (Aitken, 1981) but does not flap coronal obstruents. The history of Canadian English would then be similar to what is happening now in New Zealand, another former outpost of the British Empire with a heavy Scottish influence, where flapping has only recently begun to emerge (Silby, 2008).

Unfortunately for this simple story, we are not justified in ascribing Canadian raising to Scottish English, since diphthong raising has emerged independently across a variety of geographically dispersed English dialects (Moreton, 2021). Even worse, in some areas flapping is demonstrably older than raising, despite the fact that synchronically it is still ordered later (Fruehwald, 2016). More generally, it seems unwise to lay the burden of

explaining the present on the past when only the present is directly observable (explaining the mysterious through the even more mysterious). Even something as apparently well-established as the diachronic ordering of Grimm's law before Verner's law actually remains somewhat controversial (Noske, 2012). Such epistemological confusions should not be surprising, given that the reconstruction of dead languages depends on theoretical assumptions about what makes a plausible human language, assumptions that are derived from the study of living languages. The idea that historical linguistics should be grounded in theoretical linguistics rather than the other way around is an old one that continues to be rediscovered (e.g., Kiparsky, 2014), but it is essential to understanding the relationship between diachronic and synchronic ordering. If the latter is more fundamental, regardless of how it should be formalized or understood in psychological terms, we must recognize that every historical stage of the language has something like it, and this means that what we must actually explain is how each synchronic order morphs into the diachronically next one.

In the case of Canadian English, as with the patterns in Chinese character form that we will look at later, explaining the synchronic ordering is relatively simple: the later pattern is of a type that is universally expected to apply later. In the terms of King (1973), we might say that diphthong raising is (morpho)phonology while flapping is phonetics; in Stratal OT we would say that raising applies in the word-level stratum while flapping is phrasal. The non-phonetic status of raising is shown by lexical anomalies like those in (3), where it "overapplies" in certain cases when the diphthong is not followed by an underlyingly voiceless consonant.

(3) a. tiger [tʰʌigɚ] (Vance, 1987, p. 201)
b. didactically [dʌidæktikli] (Hall, 2005, p. 194)

The phrasal status of flapping is mostly uncontroversial, since it clearly applies across words, as in (4).

(4) get it [gɛɾ it]

The tinge of controversy comes from the observations in (5), which show the "underapplication" of flapping before an unstressed syllable in the word in (5b), presumably because this syllable is stressed in the word in (5a) that it is morphologically derived from (Withgott, 1982; Steriade, 2000). Flapping thus seems to undergo morphophonological cyclicity in a shockingly unphonetic way, or, less dramatically, perhaps there is simply enough residual stress on the relevant syllable in the derived word to block flapping as usual. An intermediate position is also possible, namely that flapping is indeed phonetic but that phonetics is sensitive to the abstract prosodic structure built on morphological structure, even when the phonetic correlates of stress are not realized. That is, we might be able to maintain a strict division between phonology and phonetics if we explode one or both into smaller parts that interact in different ways. I will apply this sort of divide-and-conquer approach a few more times before the end of this paper.

(5) a. míli[tʰ]àry
b. míli[tʰ]ărism

In any case, the synchronic ordering of raising before flapping seems to be the inevitable consequence of the different natures of these two processes, the earlier more phonological and the latter more phonetic, making irrelevant the order with which the processes emerged historically. It also does not seem arbitrary that one process behaves in a more phoneticky

way than the other in all English varieties in which both appear. Flapping is a straightforward effect of underarticulating stop closure in intervocalic position. By contrast, it is still not entirely clear what motivates diphthong raising. It had long been thought (e.g., Chambers, 1973) that the ultimate cause was the shortening of vowels before voiceless codas, which presumably compensates for the greater duration of voiceless consonants relative to voiced ones (voicing requires airflow and thus somewhere for the air to flow to, and space is limited in the restricted or even closed oral cavity created by an obstruent), and the less time there is to articulate a diphthong, the more the nucleus will be raised to shorten the articulatory distance from it to the high offglide. However, Moreton and Thomas (2007) argue for a totally different and even more complex explanation, involving not vowel shortening but instead the distinct effects of voicing and devoicing on the nucleus and offglide of diphthongs. Arguably, then, the productivity of flapping is continually renewed by its transparent articulatory support in the mouths of contemporary speakers, whereas raising took such a convoluted route from phonetics to phonology that its mental reality today is essentially cut off from its phonetic origins.

It turns out to be convenient that raising and flapping differ in their natures, since otherwise OT would not be able to handle their interaction. As OT experts can easily see, (6) confirms that there is no way to rank constraints associated with each relative to constraints that preserve features of the input form (faithfulness constraints, in OT parlance) and still yield an output that conforms to the facts. For all other readers, let me explain that the input form is in the upper left, and the forms below it represent alternative output candidates. Constraints are ranked from left to right, with the stars representing constraint violations. The winning candidate is determined by eliminating all but the most obedient candidates relative to the top-ranked constraint, then continuing to winnow the candidates down through all the other constraints, until we end up the the single winner. With the constraint ranking shown here, the winner is the candidate indicated with an arrow, but this is not the output that we want. Regardless of how the four constraint columns are scrambled, the correct form [ɹʌiɾɚ] in the last row can never win, because the raised diphthong before the voiced flap violates the raising constraint and both faithfulness constraints.

(6) writer (standard OT)

ɹaitɚ	FLAPPING	RAISING	FAITH(LOW)	FAITH(VOICE)
ɹaitɚ	*	*		
ɹʌitɚ	*		*	
→ ɹaiɾɚ				*
ɹʌiɾɚ		*	*	*

Stratal OT deals with opacity via the same sort of ordering proposed by Halle (1962), except that it is restricted to the ordering of those universal strata of stems, words, and phrases. Each stratum itself is a mini-grammar with simultaneously evaluated OT constraints, perhaps with different rankings, where the winning output of each becomes the input to the next. As applied to raising and flapping, this theory gives us an analysis like that in (7), where raising is ranked high in the word stratum but low in the phrasal stratum (note that the pattern of stars differs between strata for the faithfulness constraint for vowel lowness due to the different inputs).

(7) writer (Stratal OT)

Word stratum

ɹaitə˞	RAISING	FAITH(VOICE)	FAITH(LOW)	FLAPPING
ɹaitə˞	*			*
→ **ɹʌitə˞**			*	*
ɹaiɾə˞		*		
ɹʌiɾə˞	*	*	*	

Phrasal stratum

ɹʌitə˞	FLAPPING	FAITH(LOW)	FAITH(VOICE)	RAISING
ɹaitə˞	*	*		*
ɹʌitə˞	*			
ɹaiɾə˞		*	*	
→ ɹʌiɾə˞			*	*

3. Synchronic ordering in Chinese character form patterns

Stratal OT formalism turns out to work for Chinese characters as well. I demonstrate this in excruciating detail in Myers (2024), so here I will restrict myself to the interaction among just three formal patterns that I call (following Myers, 2019) reduplicative identity, idiosyncratic allomorphy, and regular reduction.

As in spoken and signed languages, the reduplication of components within Chinese characters expresses iconic meanings like abundance, as illustrated in (8) (throughout this paper I focus on the traditional characters used in Taiwan, though I will briefly allude to simplified characters in the next section). It is thus analogous to morphology, not phonology (as usual when I write about character phonology, I'm thinking of the silent visual sort like that in sign language). Nevertheless reduplication also reflects purely form-based generalizations, which I have argued are indeed analogous to phonology. One is the restriction of reduplication to a very small number of arrangements, again as illustrated in (8): horizontal doubling, vertical doubling, and doubling along both axes (the triangular grouping being far more common than the square one). This is reminiscent of prosodic constraints on the typically binary shape of reduplicative templates in spoken and signed languages, templates that encode the structure of reduplicants but not their content.

(8) a. 木 *mù* 'tree' 林 *lín* 'forest'
b. 火 *huǒ* 'fire' 炎 *yán* 'blazing'
c. 日 *rì* 'sun' 晶 *jīng* 'glittering'
d. 又 *yòu* 'again' 叕 *zhuó* 'connect'

The other phonology-like aspect of character reduplication is what fills in the content, namely the copying itself. In OT this is sort of thing is not handled by a copying rule, but rather by a species of faithfulness constraint that requires one part of a form to be identical to another (McCarthy & Prince, 1995), which here I call reduplicative identity. Thus each of the reduplicated forms in (8) is the morphological combination of a base morpheme (a simple component) with a prosodic template (a morpheme meaning 'abundance'), which then triggers the application of reduplicative identity constraints to yield the full phonological form.

So much for reduplication. The second of the three phonological patterns that I want to discuss is idiosyncratic allomorphy. This pattern is illustrated by the alternations in component form shown at the start of each row in (9), followed by full characters exemplifying these alternations. Myers (2019) calls such alternations idiosyncratic because they involve changes unique to a relatively small subset of character components (around thirty out of several hundred); for example, the deletion of the top stroke in (9b) and the dot in (9d) only happen in these specific components. The idiosyncratic nature of this allomorphy is reflected in Unicode, which has distinct code points for each variant.

(9)

a.	水 → 氵 *shuǐ* ‘water’	海 *hǎi* ‘sea’	(cf. 汞 *gǒng* ‘mercury’, 冰 *bīng* ‘ice’)
b.	手 → 扌 *shǒu* ‘hand’	拾 *shí* ‘pick up’	(cf. 掌 *zhǎng* ‘palm of hand’)
c.	糸 → 纟 *mì* ‘silk’	給 *gěi* ‘give’	(cf. 緊 *jǐn* ‘taut’)
d.	玉 → 王 *yù* ‘jade’	球 *qiú* ‘ball’	(cf. 瑩 *yíng* ‘lustrous’)
e.	艸 → ⺾ *cǎo* ‘grass’	花 *huā* flower	(cf. 卉 *huì* ‘plants’)
f.	火 → 灬 *huǒ* ‘fire’	照 *zhào* ‘view’	(cf. 燈 *dēng* ‘lamp’)
g.	刀 → 刂 *dāo* ‘knife’	到 *dào* ‘to’	(cf. 剪 *jiǎn* ‘scissors’)

Being idiosyncratic, the allomorphs must be lexically stored. Nevertheless, as with reduplication, the character analog of prosody is also relevant. Not only is the idiosyncratic allomorph always “lighter” relative to its full-character counterpart, with fewer and/or simpler strokes, but it is also restricted in its position. Namely, the reduced allomorph virtually always appears at the left or top, where even non-idiosyncratic components are usually shrunk in size, while the unreduced allomorph appears at the bottom or right, where components are usually enlarged. Since Myers (2019) argues that shrinking and enlargement are analogous to unstressed and stressed positions respectively, these observations mean that idiosyncratically reduced allomorphs can only appear in unstressed positions, just as one would expect of prosodically light elements, by analogy with spoken phonology. This general pattern can be see in (9a-e), with (9f-g) showing unusual cases of reduction in stressed positions.

Integrating the idiosyncratic and the regular requires yet another act of divide-and-conquer: the lexicon supplies the allomorphs, while grammar supplies the regularity. Phonologically conditioned allomorph selection is pretty common across the world’s languages (Nevins, 2011), and in OT it can be captured as illustrated in (10), where both heavy and light lexical allomorphs are offered as input, and the chosen one is whichever suits the prosodic constraints (*HEAVY/UNSTRESSED penalizes unstressing a component with relatively more or more complex strokes, and *LIGHT/STRESSED penalizes the complementary situation; the dashed lines indicate that constraint ranking doesn’t matter here).

(10) a. 海

{水, 氵} + 每	*HEAVY/UNSTRESSED	*LIGHT/STRESSED
水每	*	
→ 氵每		

b. 冰

冫 + {水, 氵}	*HEAVY/UNSTRESSED	*LIGHT/STRESSED
→ 冫水		
冫 氵		*

The third and final phonological pattern that I want to discuss is regular reduction, which, as its name is intended to suggest, is not restricted to a small subset of components. Two major types are what I call diagonalization, which, as shown in (11), converts the lowest horizontal stroke to a rising diagonal stroke, and dotting, which, as shown in (12), converts a falling diagonal stroke into a dot (i.e., very short falling diagonal); the examples shown here could be supplemented with many more. Unlike the case with idiosyncratic allomorphy, Unicode does not have distinct code points for the default and reduced allomorphs associated with regular reduction.

(11) 一 → ㇀

a.	土 *tǔ* ‘earth’	地 *dì* ‘earth’	(cf. 型 *xíng* ‘model’)
b.	金 *jīn* ‘gold’	鉛 *qiān* ‘lead’	(cf. 鑒 *jiàn* ‘mirror’)
c.	立 *lì* ‘stand’	站 *zhàn* ‘stand’	(cf. 童 *tóng* ‘child’)

(12) 丶 → 丶

a.	木 *mù* ‘tree’	根 *gēn* ‘root’	(cf. 宋 *sòng* the ‘Song’ [dynasty])
b.	夫 *fū* ‘husband’	規 *guī* ‘rule’	(cf. 扶 *fú* ‘help’)
c.	火 *huǒ* ‘fire’	燈 *dēng* ‘lamp’	(cf. 烫 *tàng* ‘scald’)

The influence of prosody is similar to that seen in idiosyncratic reduction, with regular reduction restricted to the unstressed left position, but its regularity points to another motivational factor at play: phonetics. This is because writers prefer to pull the writing instrument towards the writing hand, which means that they prefer to move the writing instrument downward, while right-handers also prefer to move the writing instrument rightward. This is why handwriting typically orders character components from left to right and top to bottom when composing a full character, and within each component, strokes conform to these directions as well, as do the individual strokes themselves. The result is that changing stroke type in the specific ways seen in both diagonalization and dotting shortens the distance between the end of the last stroke in the first-written component to the start of the first stroke in the next-written component. As with flapping, then, every time a writer applies regular reduction, its articulatory naturalness is reinforced in their mind, something that cannot be the case with idiosyncratic allomorphy.

There are potentially six ways these three patterns could be ordered synchronically, but it turns out that only one ordering correctly captures the facts: first idiosyncratic allomorphy, then reduplicative identity, and finally regular reduction. The ordering of idiosyncratic allomorphy before reduplicative identity is demonstrated by two related observations. Most of the time, for components that have idiosyncratically reduced allomorphs, what is reduplicated is the full allomorph, as shown in (13). That is, idiosyncratic allomorphy is generally not allowed to apply in this context.

(13)

a.	水 艹 氵 *shuǐ* ‘water’	林 *zhuǐ* ‘merged sandbars’	
b.	火 艹 灬 *huǒ* ‘fire’	炎 *yán* ‘blazing’	
c.	手 艹 扌 *shǒu* ‘hand’	拿 *nuò* ‘restrain’	

This makes sense if idiosyoncratic allomorphy has already had its chance to apply before reduplicative identity comes into effect, and either failed or was undone by the later process. In an OT framework, both scenarios would be formalized the same way. As illustrated in (14), reduplication as a morphological operation is already indicated in the input (by the

empty morpheme RED), but in the competition between idiosyncratic allomorphy (a side-effect of *HEAVY/UNSTRESSED) and reduplicative identity, the latter constraint wins out through its higher ranking (REDUPIDENT only applies to the morpheme RED). As shown by the fact that ordinary OT can handle this interaction, it is transparent: reduplicative identity "bleeds" idiosyncratic allomorphy, and nothing over- or underapplies. By the way, this analysis also avoids an apparent ordering paradox highlighted in a classic paper by Anderson (1971), which noted many cases across the world's languages where phonology appears to apply before reduplication, which, as morphology, should be universally ordered earlier. By exploding reduplication into its morphological and phonological aspects, we can say that its morphology aspect does indeed apply first (in the input), and it is only its phonological aspect, reduplicative identity, that applies later (in the constraint evaluation).

(14) a. 氷

{水, 氵} + RED	REDUPIDENT	*HEAVY/UNSTRESSED
→水水		*
氵水	*	

b. 海

{水, 氵} + 每	REDUPIDENT	*HEAVY/UNSTRESSED
水每		*
→氵每		

However, there is a very small number of components that reduplicate the idiosyncratically reduced allomorph rather than the full one, as shown in (15) and (16). It is presumably not coincidental that one of these components also supplies the only exception, in (15b), to the blocking of idiosyncractic allomorphy in reduplication. The reduced allomorph in (16a) also happens to be identical to the otherwise distinct component in (16b).

(15) a. 糸→纟 變 *biàn* 'change'
b. 糸→纟 絲 *sī* 'silk'

(16) a. 玉→王 玨 *jué* 'joined jade'
b. 王 *wáng* 'king'

Together the examples in (15) and (16) provide another argument for the ordering of idiosyncratic allomorphy before reduplicative identity. The exception to reduplicative identity in (15b) not only reconfirms that the reduced allomorph is stored in the lexicon, but also that it can combine independently, rather than having to be derived; Myers (2019) suggests that this example may actually reflect compounding rather than true reduplication. The neutralization of the reduced allomorph with another component, as seen in (16), may similarly give this allomorph an additional puff of life. Thus even though all idiosyncratic allomorphs must be stored in the lexicon, these two may be particularly predisposed to acting like ordinary underived forms that can themselves obey reduplicative identity.

The technical name for reduplication that copies reduplicant form onto base form instead of the other way around is backcopy, which is yet another phenomenon in Chinese characters that is also found across the world's languages (Inkelas and Zoll 2005). A possible OT analysis of the particular situation here is shown in (17), where the choice of reduced

allomorph is stipulated in the input by leaving out the full allomorph entirely. This analysis is again transparent, since idiosyncratic allomorphy per se doesn't come into play at all.

(17) 變

纟 + RED + ...	REDUPIDENT	*HEAVY/UNSTRESSED	*LIGHT/STRESSED
→纟纟			*
纟糸	*		

Establishing the order of regular reduction is more straightforward. Evidence that it is ordered after idiosyncratic allomorphy comes from examples like those in (18). The key observation is that the idiosyncratically reduced forms also undergo regular diagonalization in (18a-b) and regular dotting in (18c).

(18) a. 手 → 扌 *shǒu* 'hand' 拾 *shí* 'pick up'
b. 足 → 𧾷 *zú* 'foot' 路 *lù* 'road'
c. 衣 → 衤 *yī* 'clothing' 被 *bèi* 'quilt'

Since idiosyncratic allomorphy involves the selection of an entire allomorph, with all of its strokes, it must apply before regular reduction, or else there would be no strokes to diagonalize or turn into dots. This is particularly clear in (18b), which entirely lacks a horizontal stroke target for diagonalization in the full form, but gains one in the reduced form that incorporates the cognate character in (19).

(19) 止 *zhǐ* 'stop'

The example in (18c) makes the same point in a subtler way, since the stroke configuration at the bottom of this component can also show the totally different idiosyncratic allomorphy shown in (20), with the full form in (20a) reduced as in (20b), but the falling diagonal stroke is replaced by a dot.

(20) a. 艮 *gèn* (one of the Eight Trigrams)
b. 既 *jì* 'since'

In short, idiosyncratic allomorphy "feeds" regular reduction, that is, it creates the conditions for it to apply, putting the two patterns into a transparent relation. This means that their interaction can be handled by ordinary OT, as shown in (21). Note that only the first and last candidate outputs use the standard Unicode symbols for this component. The monstrosities shown in the middle two rows are intended to represent variants of 手 that either lack the topmost stroke (idiosyncratic reduction only) or diagonalize the bottommost stroke (regular reduction only).

(21) 拾

{手, 扌}+合	REGREDUCT	*HEAVY/UNSTRESSED
手合	*	*
扌合	*	
𠂆合		*
→扌合		

Finally, evidence that regular reduction also follows reduplicative identity comes from examples like those in (22), which show that neither diagonalization (22a) nor dotting (22b-c), nor size reduction for that matter, counts as violating identity. That is, the copies are considered "the same" despite their superficial differences.

(22) a. 蛩 *shēng* 'broom'
b. 林 *lín* 'forest'
c. 林 *zhuǐ* 'merged sandbars'

This is the reverse of what we saw with idiosyncratic reduction, where reduplication either disallows reduction or else, in rare cases, backcopies the reduced form. Since that was a feeding order, here reduplicative identity "counterfeeds" regular reduction, as illustrated in the derivational analysis in (23).

(23)

	蛩	林
Underlying	生+RED	木+RED
Reduplicative identity	生生	木木
Regular reduction	蛩	林
Surface	林	林

Counterfeeding is an opaque interaction, so if we want to do this in OT, we need to go beyond its usual formalism. If we use Stratal OT, we must accept that reduplicative identity applies in an earlier stratum than regular reduction. Myers (2019) gives a number of arguments that the strata associated with reduplicative identity and regular reduction, respectively, must be the stem-level and word-level strata, respectively. The simplest argument is that the only other universal stratum available in the theory is the phrasal stratum, but this is inapplicable to Chinese characters because characters are stored in the lexicon and so their grammar has analogs to morphology and phonology, but not to syntax. The analysis would thus be as in (24), where reduplicative identity is maximally general and thus strict, so it does not allow any deviation from identity, no matter how small. This prevents the surface form from being chosen in the first stratum where identity is high-ranked, but instead the winner waits until the next stratum for regular reduction to enforce the tiny tweaks in stroke form that we see on the surface.

(24) 林

Stem stratum

木+RED	REDUPIDENT	REGREDUCT
→**木木**		*
林	*	

Word stratum

木木	REGREDUCT	REDUPIDENT
木木	*	
→林		*

Putting all of this together, a derivational analysis of all three patterns would be as in (25). A Stratal OT analysis would work basically the same way, except that both idiosyncratic allomorphy and reduplicative identity apply within single OT mini-grammar in the stem stratum, while regular reduction applies in the word stratum.

(25)

	地	根	拾	林	林	變
Underlying	土+也	木+艮	手/扌+合	木+RED	水/氵+RED	糸+RED+...
Idiosyncratic allomorphy	--	--	扌合	--	氵水	--
Reduplicative identity	--	--	--	木木	水水	糸糸
Regular reduction	地	根	扌合	林	林	--
Surface	地	根	拾	林	林	變

As with the ordering of Canadian raising and flapping, the ordering here makes sense. Regular reduction comes last because it is the most clearly motivated in terms of phonetics. Idiosyncratic allomorphy comes first because it is anything but clearly motivated; readers and writers must simply memorize the alternating forms, even if they can predict in which prosodic context they can appear. The status of reduplicative identity lies somewhere in the middle, since despite being tied to a synchronically quite unproductive morphological process, it remains visually highly salient, and Myers (2016) showed that readers have clear intuitions on what well-formed reduplication should look like even in made-up characters. The synchronic ordering of these three processes thus follows from their relative productivity.

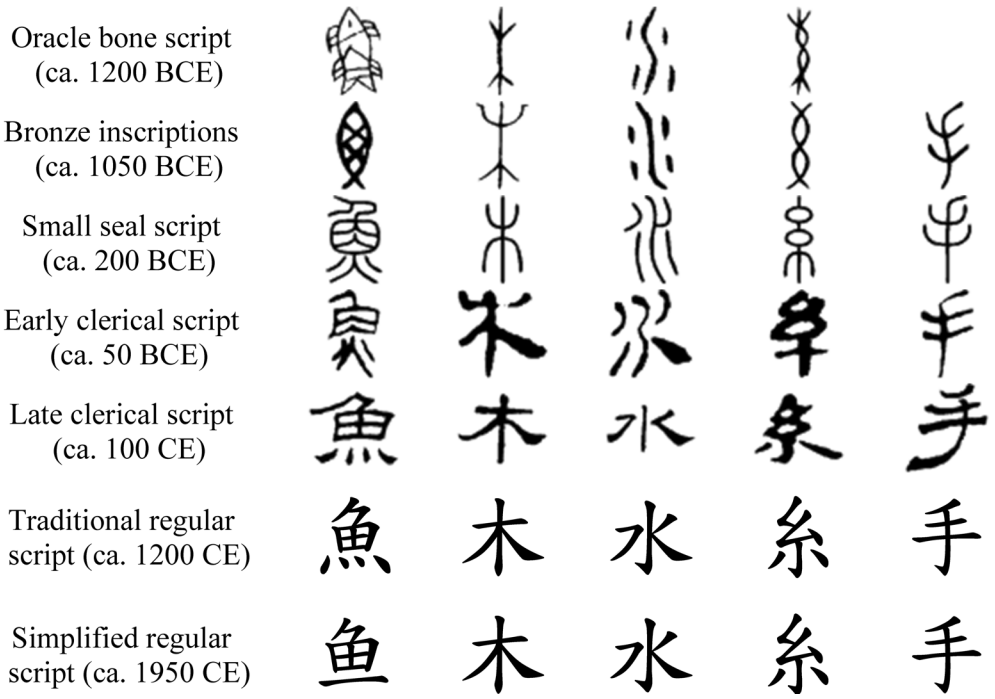
4. The history of Chinese character form patterns

The history of Chinese characters has attracted attention from curious linguists since ancient times (see Qiu, 2000, for a comprehensive overview). This is thanks in part to the fact that unlike speech or signing, writing leaves physical traces that go back thousands of years. This fact raises the possibility that the synchronic ordering in Chinese character form that we have just examined might indeed reflect nothing more than the order with which these three processes arose; perhaps here the past really is less mysterious than the present. This proves to be only partially correct, and arguably entirely wrong, in the sense that it is relative productivity that actually drives synchronic ordering, even diachronically.

Given the great length and geographical diversity of Chinese character history, I will restrict myself here to an incomplete and misleadingly linearized timeline based on the historical character images available in the online database of Juang (2017). We can get some sense of what this timeline looks like from the examples in (26) (the gaps here reflect gaps in this database). All of these examples have already been discussed except for the one in the first column (魚 *yú* 'fish'). This set of characters does not illustrate any of the processes discussed in this paper, but I include it to highlight what is by far the most common approach to "explaining" Chinese character development, namely increasing abstraction (for a recent criticism of this approach, see Han, Kelly, Winters, & Kemp, 2022). We can even see how abstraction has continued beyond the traditional characters shown in the previous section, into the modern system of simplified characters (I phrase it this way because many of these

simplified forms are ancient, but were not incorporated into a standardized system until the twentieth century).

(26)



The dates here are very approximate, and are merely meant to indicate the temporal ordering. Oracle bone script is the oldest known, and as the name suggests, it was carved into flat animal bones, including oxen shoulder blades and turtle shells, as part of divination rituals. Incompletely written exemplars demonstrate that carvers sometimes scraped a bunch of lines in the same direction across multiple characters before rotating the bone to add lines in a different direction (Venture, 2001). Bronze inscriptions were also carved, this time with a stylus in the clay mold used for casting. Small seal script was the first to be officially standardized, since it was used to distribute official business across long distances, which meant that bones and bronze were out and bamboo strips were in, with the writing now done with ink brush pens. This change in physical medium may help explain the predilection in this script for curved strokes, which are difficult to carve. However, the physical medium cannot explain why clerical script, which was also written with a brush pen, nevertheless returned to the straighter strokes of earlier scripts. This peculiarity makes sense once one remembers that the temporal linearization here is misleading; the earlier straight-stroke styles had not disappeared during the small seal script era, and remained a source of influence on later styles. In fact, what I label here as early and late clerical script refer to the Western and Eastern Han dynasties, respectively, and thus reflect differences in space as well as time. It is an interesting coincidence that the gradual change from clerical script to traditional regular script, the modern version of which remains standard in Taiwan and Hong Kong, only began after paper came into widespread use in the first century CE, though it took another millennium before regular script became standard, starting in the Song dynasty, and in any case it is not obvious why a writing surface of greater versatility than bamboo strips would encourage strokes to become even straighter than they had already been in clerical script. Finally, the cobbling together of modern simplified script from bits of traditional regular script and alternate character forms, sometimes from stroke-deprived calligraphic styles, led to no change whatsoever in any of the patterns discussed in this paper (see Myers, 2019, for

evidence that the grammars of traditional and simplified characters are virtually identical), and so it will not be discussed further. Note as well that, this simplification, nor the many other small changes that occurred in Chinese script between the Song and Mao, and which for some reason elicit virtually no attention in the scholarly literature, did not involve any change in the physical medium either.

Yes, increasing abstraction and changes in the tools of writing have both affected the evolution of Chinese characters, but that is hardly all there is to say about it. In particular, the above array of historical forms should not merely be scanned vertically, but also horizontally: each script has its own internally consistent synchronic system, or in other words, its own grammar. A complete analysis of all these synchronic grammars and all of their diachronic changes would go far beyond what I am capable of doing in this paper (though see Myers, 2019, for some first attempts at analyzing the synchronic grammars of the small seal script and the modern simplified system), but I can at least say something meaningful just about our three modern character form regularities: idiosyncratic allomorphy, reduplicative identity, and regular reduction. Namely, when did each of these patterns first emerge, and was their real-world temporal ordering consistent with what we deduced in the previous section about their apparent ordering in the synchronic grammar of modern characters?

The answers are clear: like the synchronic ordering, regular reduction appears late in the historical record, but unlike the synchronic ordering, reduplicative identity arose long before idiosyncratic allomorphy. Reduplication as a morphological process is already attested in oracle bone script, and by definition the phonological phenomenon of reduplicative identity is attested then too; see (27), which shows oracle bone examples above the modern characters with which the corresponding Chinese morphemes are written today. However, as can be seen in (27b), the prosodic restriction on the arrangements of the copied elements did not yet exist in oracle bone script; the modern arrangements only emerged as many others were gradually lost over the course of the bronze and small seal script eras (see Behr, 2006, and Liu, 2008, for details).

(27) a.

林 *lín* ‘forest’

炎 *yán* ‘blazing’

晶 *jīng* ‘glittering’

b.

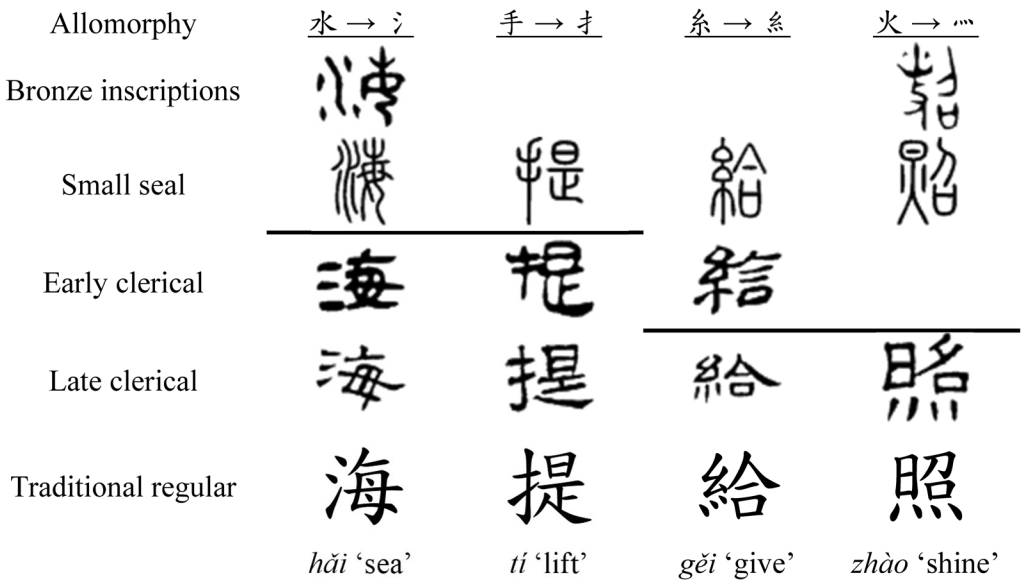
疆 *jiāng* ‘border’

少 *shǎo* ‘few’

羶 *shān* ‘rank odor’

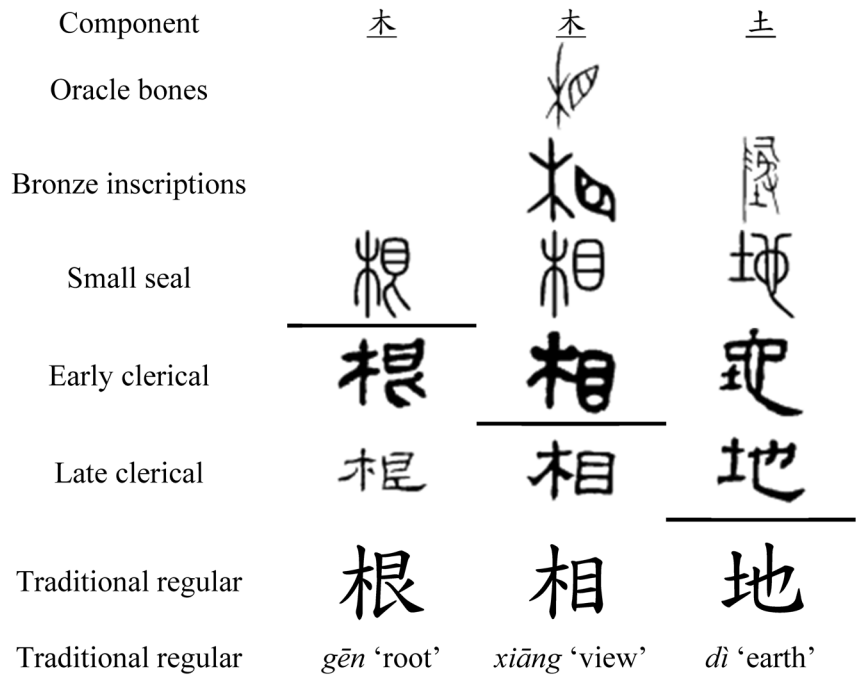
Idiosyncratic allomorphy only appeared much later, during the development of clerical script, though the precise timing depends on the specific component, as shown in (28), where the horizontal lines indicate the division between scripts before and after the change. This apparent variation in timing may partially reflect incomplete records (note the gaps, including a lack of evidence for these specific characters in oracle bone script), but perhaps also lexical diffusion, where patterns spread from one word (or word class) to another by analogy, a diachronic course the is typical for lexicalized phonological patterns (Chen & Wang, 1975; Labov, 1981).

(28)



Regular reduction also emerged late, but dotting first appeared around the same time as the earliest idiosyncratic allomorphs, in early clerical script, whereas diagonalization did not appear until regular script. As shown in (29), dotting also showed apparent variation across characters containing the same component, as can be seen by comparing the first and second columns of characters.

(29)



Variation in the development of regular reduction may again indicate lexical diffusion, but given that it is difficult to tell precisely when a stroke becomes short enough to consider a dot, nor whether the deviation of a stroke from the strict horizontal is or is not intentional, these variations could instead reflect phonetics. In favor of the lexical diffusion interpretation is the fact that dotting first appears around the same time that idiosyncratic allomorphy is also replacing full strokes with dots, as seen in the previous set of examples. It may also matter

that the example of early dotting in the first column above happens to be in a phono-semantic character, which is of a much more productive morphological type than the semantic compound showing late dotting in the second column. These subtleties make a difference to the analysis because, as we saw in the previous section, regular reduction has stronger phonetic motivations than idiosyncratic allomorphy, which would lead us to expect it to appear in the historical record in the classic all-at-once Neogrammarian fashion, rather than word by word as with more lexicalized patterns. I will need to take a closer look at the data in future research, but at least we can be sure that the final position of regular reduction in the modern synchronic ordering is only partially consistent with the historical chronology, given that it and idiosyncratic allomorphy began to emerge around the same time.

Setting all the inconvenient complexities aside, we can summarize the historical development of our three patterns with the misleadingly clean timeline in (30), where the examples are ordered chronologically from top to bottom but the rows do not necessarily represent historical simultaneity. Not only does this diachronic order deviate a bit from the synchronic ordering argued for in the previous section, but it also reveals that grammar played an active role in history, rather than merely being a fossilized remnant of it. In particular, note the interaction between reduplicative identity and idiosyncratic allomorphy in the rightmost two columns, in one of which these processes appeared simultaneously (with backcopy) and in the other of which reduplicative identity applied first, then was undone by idiosyncratic allomorphy, and finally reduplicative identity prevailed after all. The persistence of reduplicative identity across multiple scripts shows that it retained some degree of synchronic productivity even many centuries after its first appearance, whereas idiosyncratic allomorphy essentially lost whatever psychological backing it had had soon after its first appearance. The distinct ways in which these processes behave is more consistent with general psychological (or grammatical) principles choosing which pattern lives or dies, rather than the fickle vagaries of history.

(30)

Earliest forms	 (bronze)	 (seal)	 (seal)	 (bronze)	 (oracle)
Reduplicative identity: pre-seal				 (seal)	 (seal)
Idiosyncratic allomorphy: clerical					
Regular reduction: (post)clerical	地 (modern)	根 (clerical)	拾 (modern)		
Modern forms	地	根	拾	變	冰

Among the many ways in which this timeline is misleading is that it ignores the synchronic coherence within each of the historical scripts. Indeed, if something like

traditional rule-ordering theories or Stratal OT have any psychological reality, each of these scripts itself had its own grammar with its own synchronic ordering. In particular, it would be interesting to see if the interaction between reduplicative identity and regular reduction has always been as opaque as it is in the modern grammar, which in Stratal OT terms would mean that they apply in universally distinct strata. That is what we expect to see if regular reduction has always been a phonetically motivated process (argued in the previous section to apply in the word stratum), in contrast to morphologically driven reduplication (in the stem stratum). As noted earlier, however, figuring this out would take more careful analysis of a richer database, not to mention establishing grammars for each of the individual scribes in this database, in order to distinguish random phonetic variation from systematic character-by-character lexical diffusion. Readers who have glanced down to the end of this paper already know that I am not going to do any of this here.

5. Conclusions

Just as in biology, the synchrony of Chinese character structure does not recapitulate its diachrony, but instead, at least as far as we can tell given the limited data discussed here, each synchronic stage has its own internal and apparently natural (rather than ad hoc) ordering of processes. This makes the evolution of this writing system similar to what has been observed for human language more generally, and presumably also for other writing systems. Concepts like phonological rules and prosodic OT constraints have been shown to be applicable to a wide variety of writing system patterns (see, e.g., Evertz, 2018, for German and English scripts, and Gnanadesikan, 2023, for Korean and Maldivian scripts), but I would be curious to know about synchronic patterns in other writing systems that only make sense if analyzed in terms of synchronic ordering mechanisms, whether or not the ordering matches history. A crude example would be the historical evolution of lowercase Roman letters from the much older uppercase, even though in the modern system the lowercase acts as the default, with capitalization restricted to special contexts (Primus, 2004).

Bausman and Weber (2025) argue that the biologist's evo-devo framework finds genuine analogs in the cultural evolution of human language, with the evo part relating to change over historical time and the devo part reflecting the cognitive constraints that shape the learning and use of the individual languages at specific historical stages. Based on the results in this paper, it seems that the same analogs can be extended to writing systems as well.

References

Aitken, Adam J. (1981). "The Scottish Vowel-length Rule." In: *So meny people longages and tonges: Philological essays in Scots and Mediaeval English presented to Angus McIntosh*. Ed. by Michael Benskin and M. L. Samuels. Edinburgh: The Middle English Dialect Project, pp. 131–157.

Anderson, Stephen R. (1975). "On the interaction of phonological rules of various types." In: *Journal of Linguistics* 11, pp. 39–62.

Bausman, William C. and Marcel Weber. (2025). "What you can do for evolutionary developmental linguistics." In: *European Journal for Philosophy of Science* 15.1, pp. 1–18.

Behr, Wolfgang. (2006). "Homosomatic juxtaposition and the problem of 'syssymantic' (huìyì) characters". In: *Écriture chinoise: données, usages et representations*. Ed. by Françoise Bottéro and Redouane Djamouri. École des hautes études en sciences sociales, Centre de recherches linguistiques sur l'Asie orientale, pp. 75–114.

Chambers, Jack K. (1973). “Canadian Raising.” In: *Canadian Journal of Linguistics* 18, pp. 113–135.
Chen, Matthew Y. and William S.-Y. Wang. (1975). “Sound change: Actuation and implementation.” In: *Language* 51, pp. 255–281.
Evertz, Martin (2018). *Visual prosody: The graphematic foot in English and German*. Berlin: Walter de Gruyter.
Fruehwald, Josef. (2016). “The early influence of phonology on a phonetic change.” In: *Language* 92.2, pp. 376–410.
Gnanadesikan, Amalia E. (2023). “Segments and syllables in Hangeul and Thaana: A comparison and optimality theoretic analysis.” In: *Written Language & Literacy* 26.2, pp. 239–266.
Hall, Kathleen Currie. (2005). “Defining phonological rules over lexical neighbourhoods: Evidence from Canadian raising.” In: *Proceedings of the 24th West Coast Conference on Formal Linguistics*, vol. 24. Ed. by John Alderete, Chung-hye Han, and Alexei Kochetov, Somerville, Mass: Cascadilla Proceedings Project, pp. 191–199.
Halle, Morris (1962). “Phonology in generative grammar.” In: *Word* 18, pp. 54–72.
Han, Simon J., Piers Kelly, James Winters, and Charles Kemp. (2022). “Simplification is not dominant in the evolution of Chinese characters.” In: *Open Mind* 6, pp. 264–279.
Hauser, Ivy and Coral Hughto. (2020). “Analyzing opacity with contextual faithfulness constraints.” In: *Glossa* 5.1, pp. 82.1–33.
Inkelas, Sharon and Cheryl Zoll. (2005). *Reduplication: doubling in morphology*. Cambridge: Cambridge University Press.
Joos, Martin. (1942). “A phonological dilemma in Canadian English.” In: *Language* 18, pp. 141–144.
Juang Der-Ming. (2017). *The development and application of the Xiaoxuetang Wenzixue Database* [in Chinese]. Taipei: Academia Sinica Center for Digital Cultures. URL: https://xiaoxue.iis.sinica.edu.tw/ (visited on January 13, 2025).
King, Robert D. (1973). “Rule insertion.” In: *Language* 49.3, pp. 551–578.
Kiparsky, Paul. (1968). “Language universals and linguistic change.” In: *Universals in linguistic theory*. Ed. by Emmon Bach and Robert Harms. New York: Holt, Rinehart & Winston, pp. 170-202.
———— (2014). “New perspectives in historical linguistics.” In: *The Routledge handbook of historical linguistics*. Ed. by Claire Bowern and Bethwyn Evans. Abingdon: Routledge, pp. 64–102.
———— (2015). “Stratal OT: A synopsis and FAQs.” In: *Capturing phonological shades within and across languages.* Ed. by Yuchau E. Hsiao and Lian-Hee Wee. Newcastle upon Tyne: Cambridge Scholars Publishing, pp. 1–44.
Labov, William. (1981). “Resolving the Neogrammarian controversy.” In: *Language* 57.2, pp. 267–308.
Liu Man. (2008). “The diachronic study and cognitive comparison of stacked characters” [in Chinese]. Ph.D. thesis. Tsinghua University (Beijing).
Love, Alan C. (2024). *Evolution and development*. Cambridge, UK: Cambridge University Press.
McCarthy, John. (1999). “Sympathy and phonological opacity.” In: *Phonology* 16.3, pp. 331–399.
———— (2007). *Hidden generalizations: Phonological opacity in Optimality Theory*. London: Equinox Publishing.
———— and Alan Prince. (1995). “Faithfulness and reduplicative identity.” In: *University of Massachusetts Occasional Papers in Linguistics 18*. Ed. by Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk. Amherst, MA: GLSA, pp. 249–384.

Moreton, Elliott. (2021). “Phonological abstractness in English diphthong raising.” In: *American Speech* 106.1, pp. 13–44.
———— and Erik R. Thomas. (2007). “Origins of Canadian Raising in voiceless-coda effects: A case study in phonologization.” In: *Laboratory Phonology* 9, ed. by Jennifer S. Cole and José Ignacio Hualde. Berlin: Mouton, pp. 37–64.
Myers, James. (2016). “Knowing Chinese character grammar.” In: *Cognition* 147, pp. 127–132.
———— (2019). *The grammar of Chinese characters: Productive knowledge of formal patterns in an orthographic system*. Abingdon: Routledge.
———— (2021a). “Levels of structure within Chinese character constituents.” In: *Grapholinguistics in the 21st Century: From graphemes to knowledge*, Part II . Ed. by Yannis Haralambous. Paris: Fluxus Editions, pp. 645–582.
———— (2021b). “Areal script form patterns with Chinese characteristics.” In: *Written Language & Literacy* 24.2, pp. 259–283.
———— (2022). “Teaching Chinese character grammar.” In: *Chinese Language Learning and Technology* 2.2, pp. 1–37.
———— (2024). “Interactions among patterns in Chinese character form.” In: *Looking at language from all sides*. Ed. by James Myers and James H-Y. Tai. Taipei: Crane Publishing, pp. 129–152.
Nevins, Andrew. (2011). “Phonologically-conditioned allomorph selection.” In: *The Blackwell companion to phonology, Volume IV: Phonological interfaces*. Ed. by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice. Malden, MA: Wiley-Blackwell, 2357–2382.
Noske, Roland. (2012). The Grimm-Verner push chain and Contrast Preservation Theory. In: *Phonological explorations: Empirical, theoretical and diachronic issues*. Ed. by Bert Botma and Roland Noske. Berlin: De Gruyter, pp. 63–86.
Primus, Beatrice. (2004). “A featural analysis of the Modern Roman Alphabet.” In: *Written Language & Literacy* 7.2: 235–274.
Prince, Alan and Paul Smolensky. (2004). *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA, and Oxford, UK: Blackwell.
Pruitt, Kathryn. (2023). “Serialism and opacity in phonological theory.” In: *Annual Review of Linguistics* 9, pp. 497–517.
Qiu, Xigui. (2000). *Chinese writing*. Compiled and translated from Chinese by G. L. Mattos and J. Norman. Berkeley: The Society for the Study of Early China and The Institute of East Asian Studies, University of California.
Silby, Winston. (2008). “t-flapping in New Zealand English: A change over time.” In: *New Zealand English Journal* 22, pp. 24–35.
Steriade, Donca. (2000). “Paradigm uniformity and the phonetics-phonology boundary.” In: *Papers in Laboratory Phonology* 5. Ed. by Michael Broe and Janet Pierrehumbert. Cambridge, UK: Cambridge University Press, pp. 313–335.
Vance, Timothy J. (1987). “‘Canadian Raising’ in some dialects of the northern United States.” In: *American Speech* 62, pp. 195–210.
Venture, Olivier. (2001). “Quelques observations au sujet de la mise en page des textes de divination sur plastron”. In: *Actes du colloque international commémorant le centenaire de la découverte des inscriptions sur les carapaces*, ed. by Yau Shun-chiu and Chrystelle Maréchal. Paris: Éditions Langages Croisés, pp.71–90.
Withgott, M. Margaret. (1982). “Segmental evidence for phonological constituents.” Ph.D. thesis. University of Texas at Austin.