Inner structure and outer form

Analogues of morphology and phonology in writing systems

James Myers
National Chung Cheng University, Taiwan
https://lngmyers.ccu.edu.tw/

Thanks!

- National Science and Technology Council (Taiwan) grants 97-2410-H-194-067-MY3, 101-2410-H-194-115-MY3, 103-2410-H-194-119-MY3, 109-2410-H-194-096-MY3, 112-2410-H-194-030-MY3
- My chief lab assistant Yangru Chen and many other lab assistants and experimental participants over the years
- Many conference attendees and journal reviewers
- You too, maybe...?

Overview

- Grammatical approaches to writing systems
- Morphology vs. phonology
- Case studies
  - Chinese characters
  - Latin letters
  - Devanagari
  - Others

The linguistics of writing systems

> Writing is not language, but merely a way of recording language by means of visible marks.

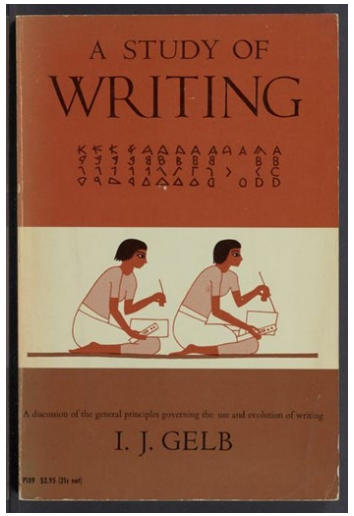
←Bloomfield (1933)

Also Bloomfield (1933)→

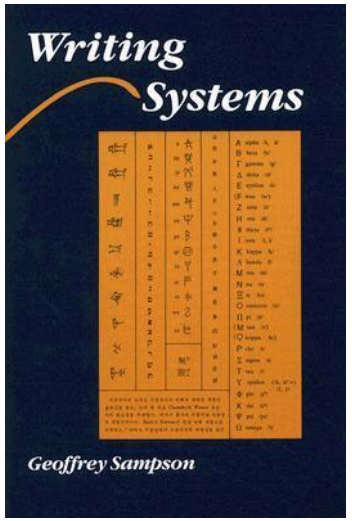
> It seems certain that these gesture languages are merely developments of ordinary gestures and that any and all complicated or not immediately intelligible gestures are based on the conventions of ordinary speech.

- There's actually a long & rich tradition of linguistic studies of writing

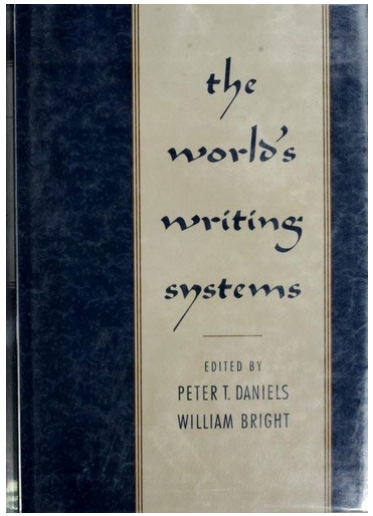
Gelb (1963 [1952])



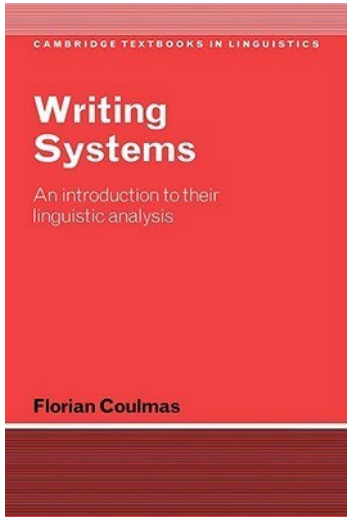
Sampson (1985)



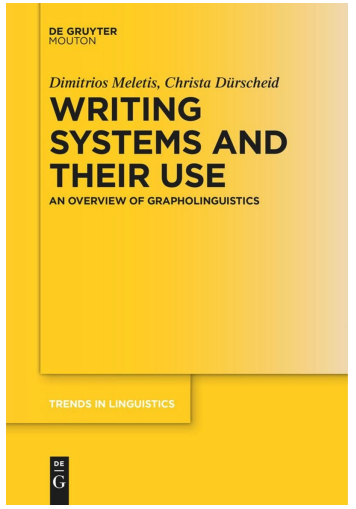
Daniels & Bright (1996)



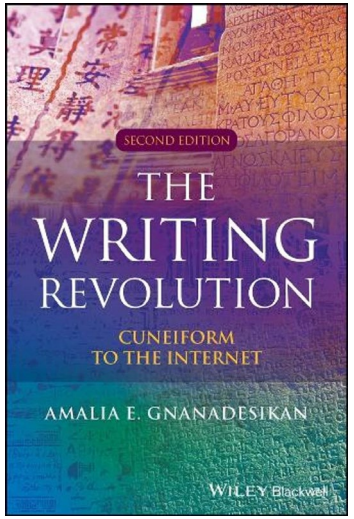
Coulmas (2002)



Meletis & Dürscheid (2022)



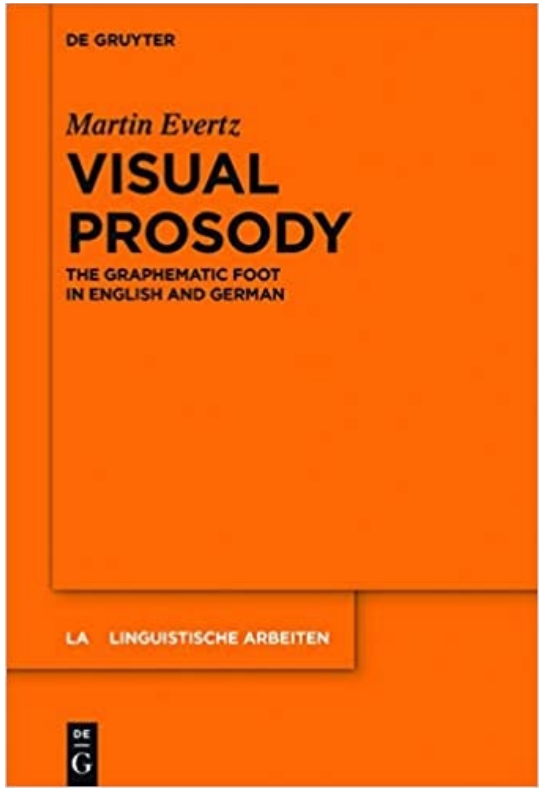
Gnanadesikan (2025)



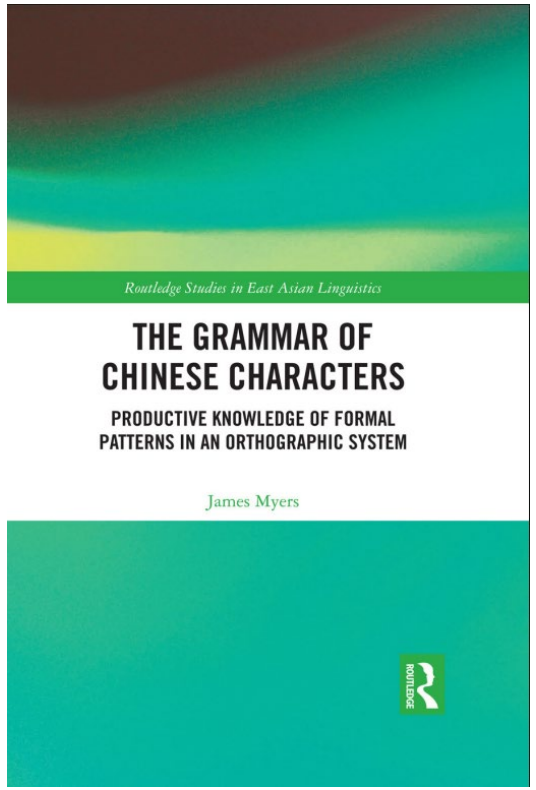
The *grammar* of writing systems?

- Some linguists work on this too

Evertz (2018)



Myers (2019)



Wang (1983)

TOWARD A GENERATIVE GRAMMAR OF

CHINESE CHARACTER STRUCTURE AND STROKE ORDER

A thesis submitted to the Graduate School of the

Univeraity of Wisconsin-Madison in partial fulfillment of

the requirements for the degree of Doctor of Philosophy

by

Jason Chia-Sheng Wang

Degree to be awarded: December 19 83 May 19____ August 19____

JAMES D. MC CAWLEY

(1994)

SOME GRAPHOTACTIC CONSTRAINTS

Segments and syllables in Hangeul and Thaana

A comparison and optimality theoretic analysis

Amalia E. Gnanadesikan
University of Maryland, retired

(2023)

Morphology vs. phonology in writing systems

- Gelb (1963): “inner structure” vs. “outer form”
- Contemporary (German?) tradition (e.g. Meletis & Dürscheid 2022)
  - Graphematics: “how the visual units of writing assume linguistic functions ... and how they relate to linguistic units”
  - Graphetics: “visual form and appearance of writing as well as its physical ... features”
- Inner structure is written “morphology”
  - Maps between written units and spoken language
- Outer form is written “phonology”
  - Realizes form physically (silently, as in sign language phonology)

Overview

- Grammatical approaches to writing systems
- Morphology vs. phonology
  - What are they and how do they differ?
  - Where do they “leak”?
- Case studies
  - Chinese characters
  - Latin letters
  - Devanagari
  - Others

Morphology: Morphemes

- Minimal mapping between forms and interpretations

  dog  *-s* PLURAL

- Segmentation challenges

  breakfast = *break* + *fast*?

  German: *gemacht* = *mach* 'make' + *ge ... t* PAST PARTICIPLE

- Opacity: noncompositional semantics

  freedom = *free* ['unbound', not 'gratis'] + *dom*

- Empty morphs: formally coherent but semantically vacuous

  receive ~ *reception*, *conceive* ~ *conception*

- Phonesthemes: insufficiently systematic form-meaning mappings

  glitter, *gleam*, *glint*, *glisten*, ... vs. *gland* (and **itter*, **eam*, **int*, **isten*)

Morphology: Affixation

- Closed class, bound, fixed position
  - *dog*, *dogs*, **s*, **sdog*
- Semantically abstract
  - *dog*  vs. *-s* PLURAL
- Formally reduced
  - *dog* is a full syllable, *-s* isn't
- Operators: Convert one syntactic/semantic class to another
  - V-*er* = $[V+AGENT]_N$
- Derivation vs. inflection
  - Derivation: Creates new lexical entry (*dancer*)
  - Inflection: Accommodates lexical entry in sentential context (*dancing*)

Morphology: Compounding

- Open class (roots)
  - *wug task*
- Generally not bound
  - Cf. many Chinese compounds: *xiàozhǎng* ‘principal’ (*xiào* ‘school’ + *zhǎng* ‘master’)
- Positions generally not fixed
  - *houseboat* *boathouse*
- Concrete semantics
  - Interpretation depends on real-world knowledge (Jackendoff 2010)

    snowballs are made of snow, but *snowplows* are used for snow

Morphology: Reduplication

- Iconic meaning (Hurch 2005)
  - Plurality and abundance*: *rénrén* ‘everyone’ (*rén* ‘person’) [Chinese]
  - Intensity: *guāngguāng* ‘completely bare’ (*guāng* ‘bare’)
  - Attenuation: *shìshì* ‘give it a try’ (*shì* ‘try’)
  - Repetition & durativity: many sign languages...
- Prosodic templates (syllables, feet, ...), e.g. [XX]
  - Kinande (Mutaka & Hyman, 1990)
    - *o-ku-gulu* ‘leg’ *o-ku-**gulu**-gulu* ‘a real leg’
    - *e-n-dwa* ‘wedding’ *e-**n-dwa-n-dwa**-n-dwa* ‘a real wedding’
  - Child speech (Dressler et al., 2005)
    - ***bebe*** ‘bear’ (< *Bär* [German])
  - Sign languages (Wilbur, 2009)
    - **SIT-SIT** ‘chair’ (ASL)

*(here, distributive)

Phonology: Phonemes & features

- Phonemes are distinctive and usually also contrastive
- Segment-sized
  - Segments can be tricky to identify
    - Cf. tones
    - Does *James* have four phonemes /dʒ eɪ m z/ or six /d ʒ e ɪ m z/?
  - Do sign languages have a direct analogue to segments?
    - Movements and positions (Perlmutter 1992) or entire signs (Channon 2002)?
- Features are not quite as universal as once thought
  - Spoken vs. signed features
  - Mielke (2008): Of 6,077 phonology classes in 561 spoken languages...
    - ... “unnatural” classes are often more common than supposedly “natural” classes
  - So what’s universal is the emergence of feature-like units through language-specific experience, not a fixed feature inventory

Phonology: Patterns

- Allophones (distributions)
  - *top*, *stop*, *tree*
- Alternations (allomorphy)
  - Phonologically conditioned: *delete* ~ *deletion*, *prettier* vs. *more beautiful*
  - Lexically conditioned: *child* ~ *children*
- Lexicalization (exceptionality & morphology-sensitivity)
  - *obscene* ~ *obscenity* *obese* ~ *obesity*
  - *electricity* vs. *panicking*
  - Lexicalized phonology makes words seem monomorphemic (Goldsmith 1990)

    Cánada *América*n *Américanize*

    órigin *oríginal* *originálity*

Phonology: Prosody

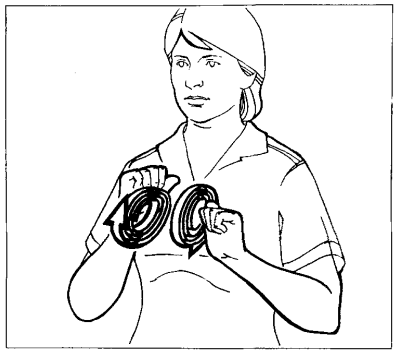
- Asymmetry
  - *blackbòard* *blàck bóard*
- Constituents: groups rather than features
  - *Do you live in* ^{L}*Canada? Yes, I live in* ^{H}*Canada.*
  - *háppy* ~ *háppy* + *er* *còmpáct* ~ *more còmpáct* (counting feet)
- Aligned with morphemes/words
  - *Américanize* *originálity*
  - *blackbòard* *blàck bóard*
- Hierarchical but not recursive
  - Syllables
  - Metrical feet
  - Prosodic words

Morphology vs. phonology: Interactions

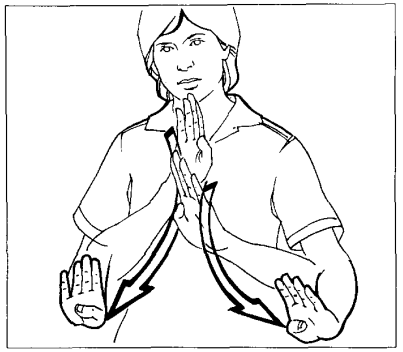
- Morphology precedes phonology (ASL: Padden & Perlmutter 1987)

MORPHOLOGY

ACTING

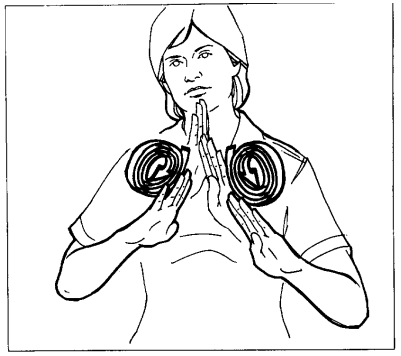


QUIET



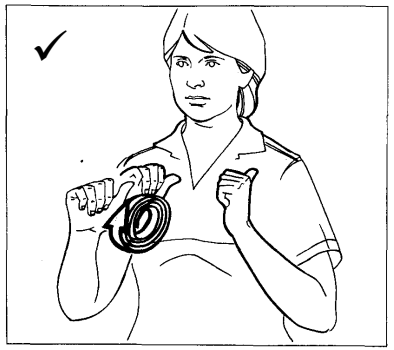
QUIET+ISH

QUIET+CHARACTERISTIC+ISH

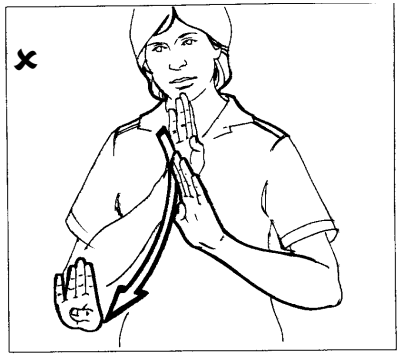


PHONOLOGY

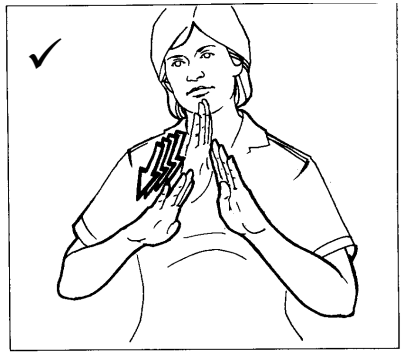
ACTING (weak freeze)



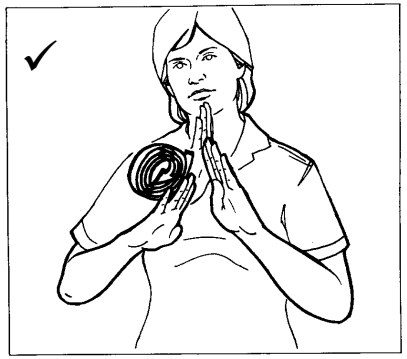
*QUIET (weak freeze)



QUIET+ISH (weak freeze)



QUIET+CHARACTERISTIC+ISH (weak freeze)



Morphology vs. phonology: Frequency effects

- Morphology and phonology show distinct frequency patterns (Bybee 2001, Phillips 2006)
- Morphological frequency effects are negative
  - More common = less regular: ***crow ~ crowed*** vs. *blow ~ blew*
  - Regular rules kick in when lexical access fails for rarer items
- Phonological frequency effects are positive if phonetically motivated:
  - More common = more regular: *mámmary* vs. ***mém'ry*** (binary left-headed foot)
  - Usage increases number of lenited exemplars (Pierrehumbert 2002)
- Phonological frequency effects are negative if lexicalized:
  - More common = less regular: ***nude /u/*** *vs. new /ju/* (*Cor-j in S. US: Phillips 2006)
  - Lexical reanalysis more likely for less accessible words (Hay et al. 2015)

Overview

- Grammatical approaches to writing systems
- Morphology vs. phonology
- Case studies
  - Chinese characters ←
  - Latin letters
  - Devanagari
  - Others

Chinese

- The poster child for logographic systems
  - Characters virtually always represent monosyllabic morphemes
  - This makes them the system's graphemes (Meletis 2019)
- Iconic characters (指事 & 象形) have quite low type frequency

  上下 *shàngxià* 'above and below'　　木 *mù* 'tree'
- Most characters are composed of two or more components
  - Phono-semantic (semantic-phonetic) characters (形聲)

    根 *gēn* 'root' = 木 'tree' + 艮 *gèn*
  - Semantic compounds (會意)

    相 *xiàng* 'see' = 木 'tree' + 目 'eye'

    林 *lín* 'forest' = 木 'tree' + 木 'tree'

Chinese: “Morphemes”

- Components are generally interpreted, hence “morphemes”
  根 *gēn* ‘root’ = 木 ‘tree’ (semantic class) + 艮 *gèn* (hints at pronunciation)
- Segmentation challenges
  施 *shī* ‘apply’ = 旆 *yǎn* ‘flying flags’ (obs.) + 也 *yě* ‘also’
- Opacity
  解 *jiě* ‘solve’ = 角 ‘horn’ + 刀 ‘knife’ + 牛 ‘cow’
- Empty morphs
  勹 in 色 *sè* ‘color’ & 角 *jiǎo* ‘horn’
- Phonesthemes
  灬 in animal-related components: 馬 ‘horse’, 鳥 ‘bird’, 魚 ‘fish’

Chinese: “Morphological” operations

- “Affixation”

  根 *gēn* ‘root’ = 木 ‘tree’ + 艮 *gèn*

  affix* stem

- “Compounding”

  相 *xiàng* ‘see’ = 木 ‘tree’ + 目 ‘eye’

  root root

- “Reduplication”

  林 *lín* ‘forest’ = 木 ‘tree’ + [XX]

  root template

*(“radical” in traditional English terminology)

Chinese: “Affixation” diagnostics

- Closed class
  - Only about 200 “radicals” for dictionary indexing, only some of which are actually semantic components
- Abstract semantics

  賬 *zhàng* ‘account’ = 貝 ‘shell’ / ‘money’ + 長 *cháng*

  嫌 *xián* ‘fed up with’ = 女 ‘female’ / ‘bad things’ + 兼 *jiān*

  嗎 *ma* (sentence-final particle) = 口 ‘mouth’ / ‘function morpheme’ + 馬 *mǎ*
- Bound

  Fully bound: 冫 冂 亠 夂 殳 彳

  Bound “allomorphs”: 心~忄‘heart’ 手~扌‘hand’ 水~氵‘water’ 艸~艹‘grass’
- Relatively fixed positions
  - Most “affixes” appear in same position in most of their characters
  - Top and left are most common (cf. suffixation preference in English)

Chinese: “Affixation” - other properties

- Reduced
  - Usually fewer strokes than “stem”: 根 嗎 賬 嫌 ... (cf. 鳥 ‘bird’ in 鴨 *yā* ‘duck’)
  - Reductive idiosyncratic “allomorphy” is common:
    心~忄‘heart’ 手~扌‘hand’ 水~氵‘water’ 艸~艹‘grass’
- Operators?
  - Identifies semantic class for whole character, but does it convert the class...?
  - No selectional restrictions on “stem”
- “Inflection” or “derivation”?
  - “Affixes” create new lexical entries...
  - ... but they tend to “agree” in (the very rare) two-character morphemes
    葡萄 *pútáo* ‘grape’

Chinese: “Compounding”

- Concrete semantics

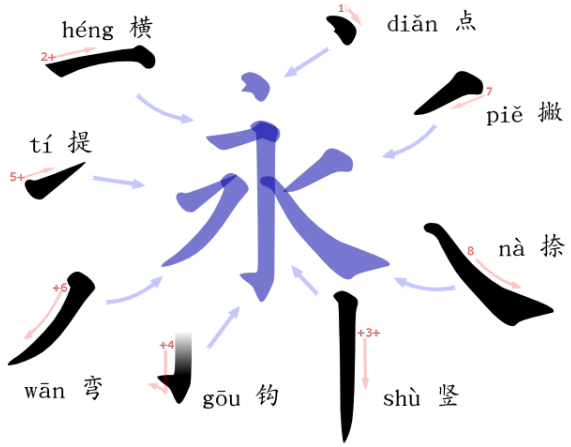
  吠 *fèi* ‘bark of a dog’ = 口 ‘mouth’ (literally) + 犬 ‘dog’

- Positions are often determined by real-world knowledge (iconicity)
  - 杲 *gǎo* ‘bright’ (sun over tree) vs. 杳 *yǎo* ‘dark’ (sun below tree)
  - 尿 *niào* ‘urine’ (水 ‘water’ under 尸 ‘body’)
    畓 *duō* ‘rice field’ (水 ‘water’ on 田 ‘field’)
    盥 *guàn* ‘wash the hands’ (hands holding water over a basin)

Chinese: “Reduplication”

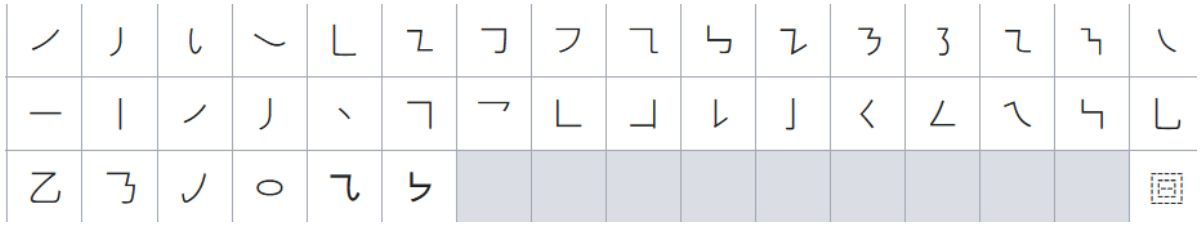
- Iconic semantics (Behr 2006)
  - Plurality/abundance: 多 *duō* ‘many’ 艸 *cǎo* ‘grass’ 森 *sēn* ‘densely forested’
  - Intensity: 晶 *jīng* ‘glittering’ 炎 *yán* ‘blazing’
  - Attenuation: 弱 *ruò* ‘weak’
- Shape templates
  - Standard (as in examples above): 
  - Rare:  (as in  *mǎng* ‘multitude of grass’)
     (as in  *líng* ‘spirit’)

Chinese: “Phonemes” & “features”



Wikimedia Commons

Complex strokes are composed of simple ones



https://en.wikipedia.org/wiki/CJK_Strokes_(Unicode_block)

Stroke	*Axis*	*Direction*	*Examples*
丶	None	(default)	太
一	Horizontal	(default)	十
丨	Vertical	(default)	十
丿	Main diagonal	(default)	木
乀	Counterdiagonal	Leftward falling	千 才
乁	Counterdiagonal	Rightward rising	子

Stroke	*Axis*	*Curving*	*Hooking*	*Examples*
亅	Vertical	Yes	No	川
亅	Vertical	No	Leftward	丁
乚	Vertical	No	Rightward	艮
乛	Horizontal	No	Downward	心
乁	Main diagonal	Yes	Upward	戈
㇇	Vertical	Yes	Leftward	犭

Note: Stroke images from https://commons.wikimedia.org/wiki/Category:CJK_strokes (by user ‘Cangjie6’), licensed under the Creative Commons CC0 1.0 Universal Public Domain Dedication.

Myers (2019)

Chinese: “Phonological” patterns

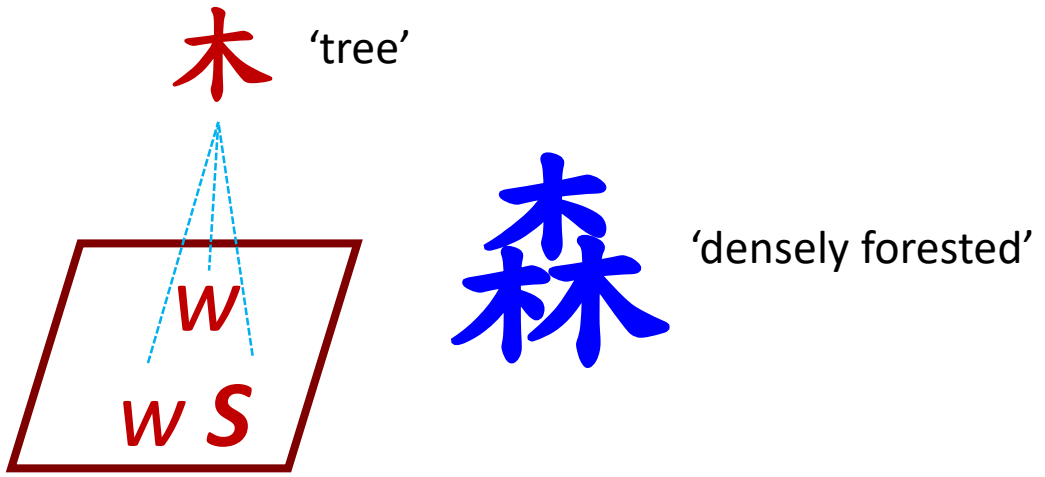
- “Allophonic” variation
  - Stroke enlargement at bottom/right
    三 *sān* ‘three’ 川 *chuān* ‘river’ 土 *tǔ* ‘soil’ (cf. 士 *shì* ‘scholar’)
  - Stroke curving at left
    川 *chuān* ‘river’ 用 *yòng* ‘use’ (cf. 同 *tóng* ‘same’)
    - “Left” defined by component edge: 訓 *xùn* ‘lesson’
- “Alternations”
  - Regular diagonalization: 土 *tǔ* ‘soil’ 地 *dì* ‘ground’
  - Idiosyncratic “allomorphy”
    心~忄‘heart’: 心 *xīn* ‘heart’ 恨 *hèn* ‘hate’
    - Idiosyncratic curving on left of whole character (makes them seem “monomorphemic”)
      辛 ‘toilsome’: 辣 *là* ‘spicy’

Chinese: “Prosody” (1)

- Asymmetry
  - Enlargement on bottom/right
- Constituents
  - Enlargement targets entire stroke groups
    㠯 *duī* ‘heap’ **飛** *fēi* ‘fly’
- Aligned with “morphemes”/“words”
  - Enlargement of entire components
    多 *duō* ‘many’ **林** *lín* ‘forest’ **大** ‘big’: **尖** ‘sharp’ vs. **奇** ‘strange’
- Hierarchical
  - **圭** *guī* ‘jade tablet’

Chinese: “Prosody” (2)

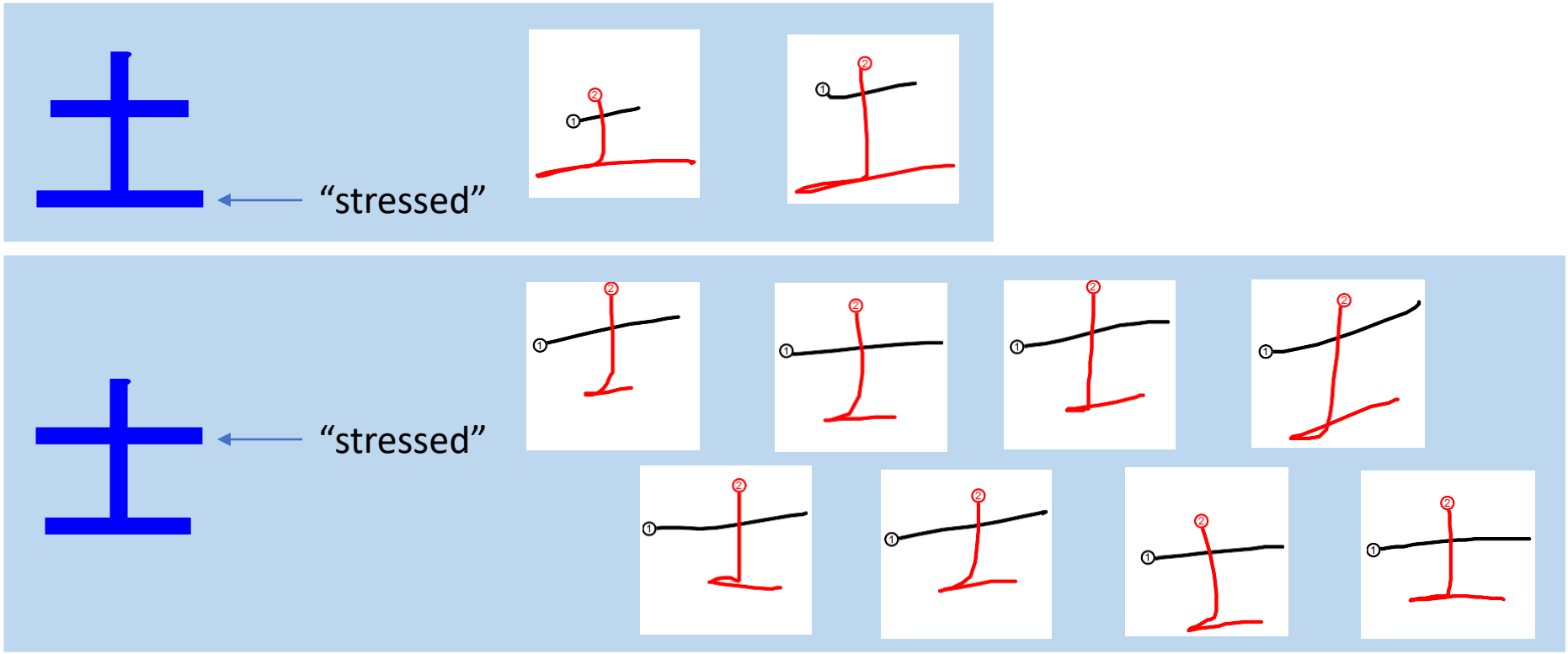
- Reduction of all kinds are preferred in “weak” top/left positions
  - “Affixes” favor them, unless they’re “heavy”: 鳥 ‘bird’ in 鴨 *yā* ‘duck’)
  - Curving & diagonalization occur in “weak” position too: 川　辣　地
  - Idiosyncratic allomorphy too: 心~忄‘heart’:　忙 *máng* ‘busy’　(cf. 忘 *wàng* ‘forget’)
- “Reduplicative templates” show same “prosodic” structure
  - Binary along horizontal and/or vertical axis
  - “Strong” slot at bottom/right
  - Other slots “weak”



Chinese: "Prosody" (3)

- "Stressed" strokes don't like to merge with others

  (Handwriting study on 34 participants)



(Myers "someday"; see also Myers 2023 for how these images were created, and Myers 2024 [2026] for the effect of "stress" on stroke curving & timing in handwriting)

Chinese: “Morphology” precedes “phonology”

- Idiosyncratic allomorphy feeds regular diagonalization
  - 足~⻊ ‘foot’: 跟 *gēn* ‘follow’
- Idiosyncratic allomorphy is blocked in “reduplication”
  - Affixation: 水~氵清 *qīng* ‘clear’ = 水 ‘water’ + 青 *qīng*
  - Compounding: 水~氵涉 *shè* ‘wade’ = 水 ‘water’ + 步 ‘walk’
  - Reduplication: 水 淞 *zhuǐ* ‘water, rivers’ = 水 ‘water’ + **[XX]**
- But diagonalization is not blocked...

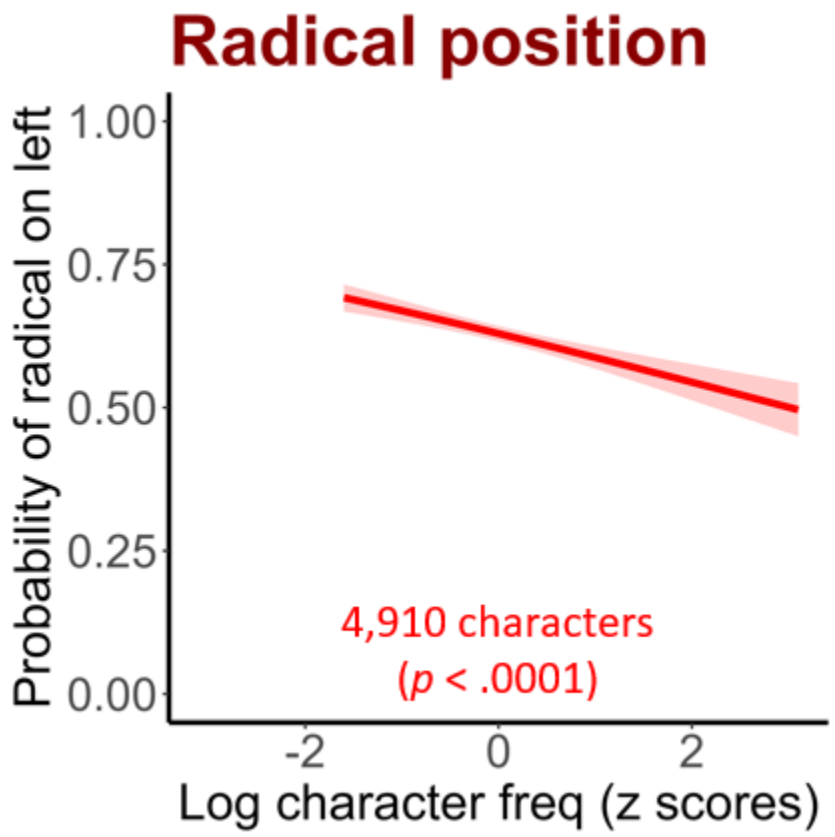
  比 牲 玨 孖 竝 鍷 垚
- ... nor is curving

  艸 辫

(See Myers 2024 for an exhausting analysis of all these “morphological” and “phonological” interactions)

Chinese: "Morphological" frequency effects

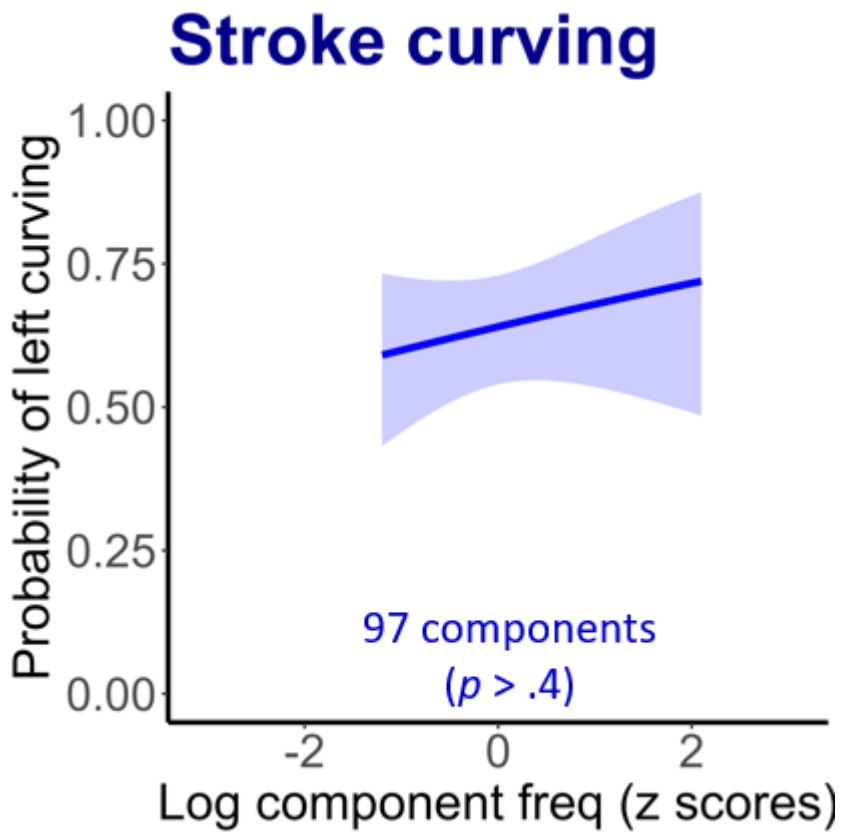
- Frequency has negative effects on component combination patterns (logistic regression)

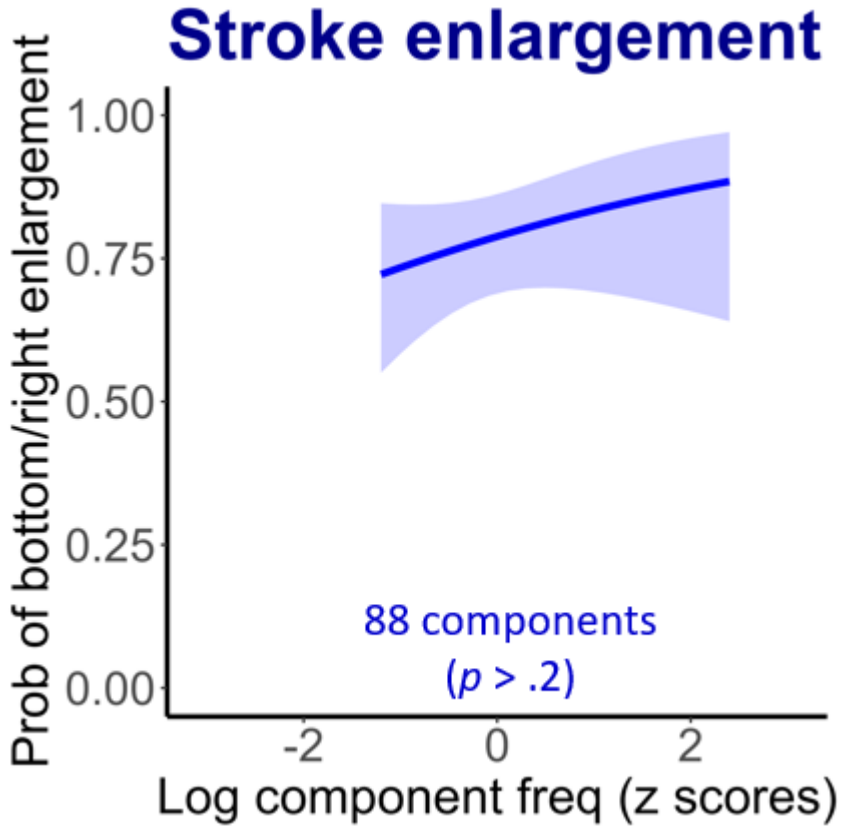


(Myers 2026)

Chinese: "Phonological" frequency effects (1)

- Frequency has positive effects on component form patterns





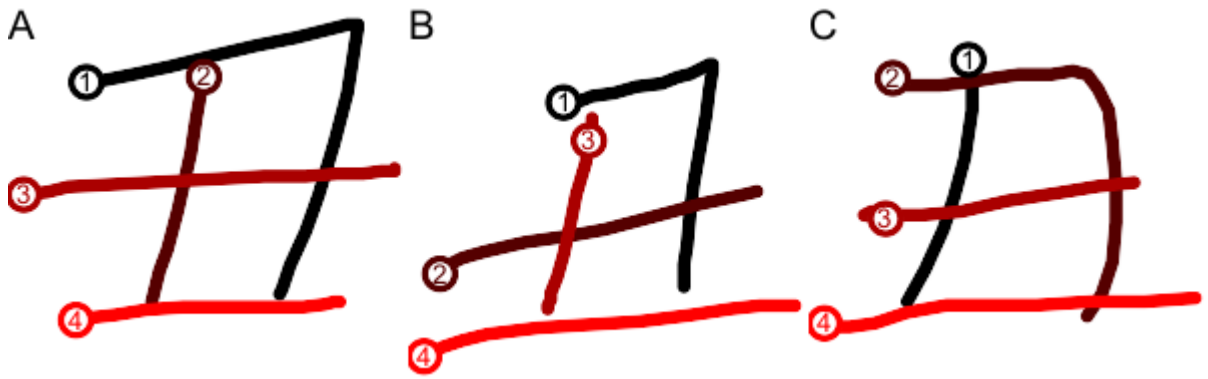
(Myers 2026)

Chinese: "Phonological" frequency effects (2)

- Stroke order relates to character form, so should be "phonology" or "phonetics"
- It is explicitly taught in school, causing it to become partially lexicalized
- Actual handwriting shows variation, though:

丑 *chǒu* 'clown'

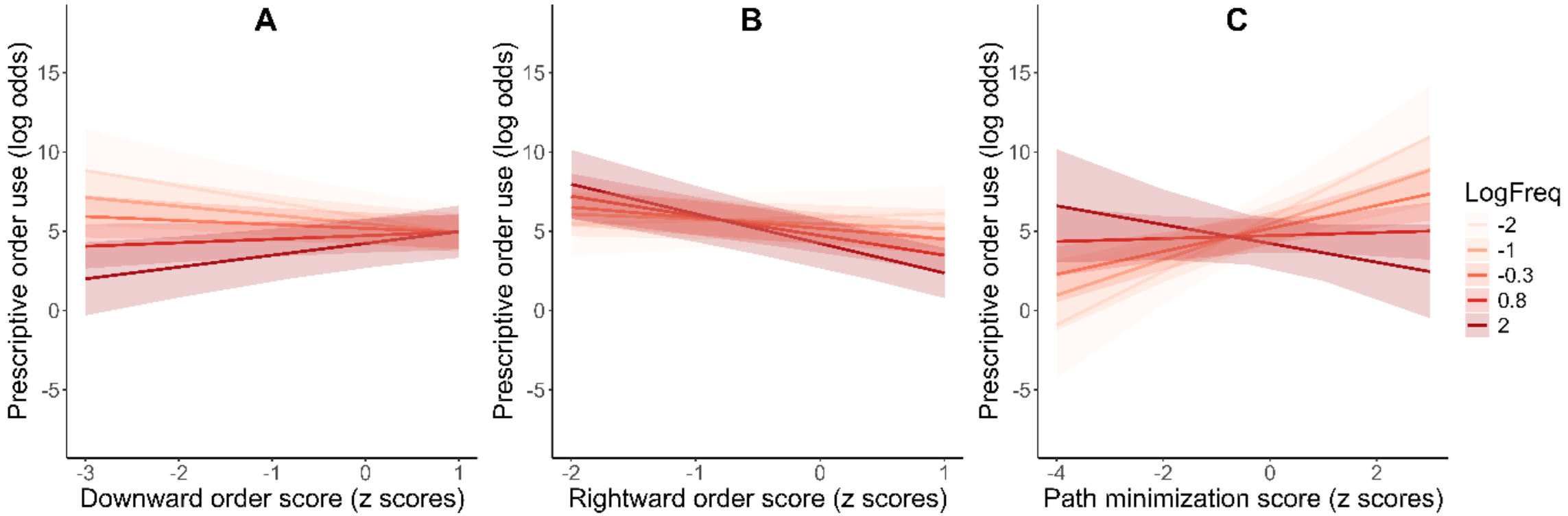
(prescriptive order) (other attested orders)



(Myers "someday")

Chinese: "Phonological" frequency effects (3)

- Lexicalized stroke order can also show negative frequency effects



(positive effect)

(negative effect)

(negative effect)

(Myers "someday")

Overview

- Grammatical approaches to writing systems
- Morphology vs. phonology
- Case studies
  - Chinese characters
  - Latin letters ←
  - Devanagari
  - Others

Latin letters

- Letters are interpreted (phonemes), so are "morphemes": <f> /f/
- Segmentation challenges: complex graphemes & nonlocality

  <ph> /f/ <x> /ks/ <i ... e> /aj/ (e.g. <ripe> /rajp/)
- Opacity: inconsistent mappings

  <cat> /k/ <city> /s/ <chip> /tʃ/ <school> /k/ <cache> /ʃ/
- Empty "morphs"

  <psychology> <pterodactyl>
- "Phonesthemes": irregular letter-internal mapping (cf. Fuhrhop et al. 2011)
  - Squat lowercase letters for sonorants: <a e i o u, m n r w>
  - Tall lowercase letters for obstruents: <b d f g h j k p t q>
  - Exceptions: <c s v x z> <l y>

Latin letters: "Morphological" operations

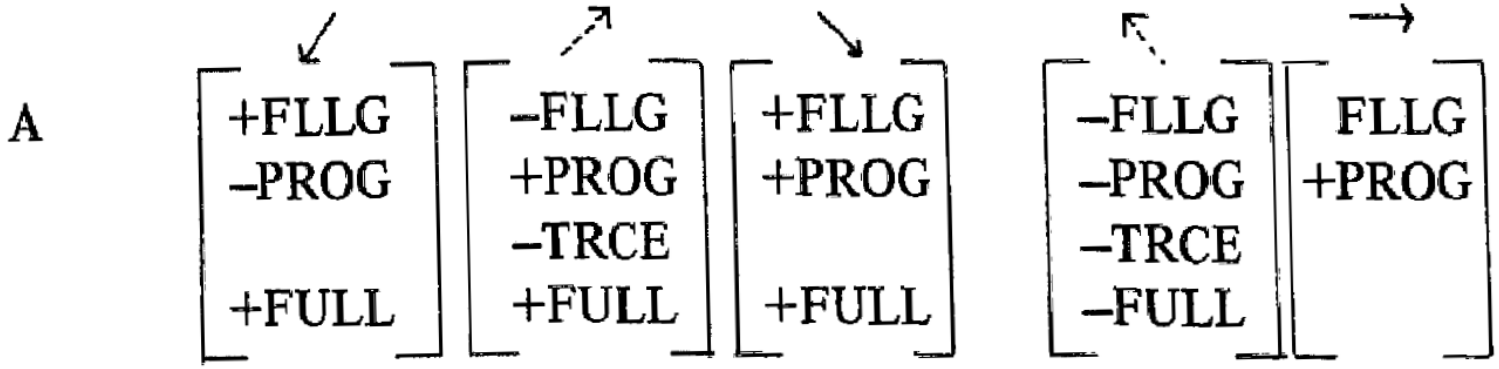
- "Compounding"
  - Virtually all letter combinations are "compounds" of various sorts
  - Interpretation depends language use, i.e. speech (Evertz 2018)
    - German speakers interpret <sasch> /saʃ/ as prosodically heavier than <sas> /sas/, as shown by stress attraction when reading aloud written nonwords
- "Reduplication"
  - Binary & iconic in many languages (e.g. Finnish): <kk> /kk/ <ii> /i:/
- "Affixation"
  - Diacritics: German <o ~ ö> /o ~ ø/ Spanish <a ~ á> /a/ in different contexts
  - Bound, fixed in position, and act as operators (e.g. umlaut, accent)
    - Umlaut as "derivation" and accent as "inflection"...?
  - (Cf. punctuation as "clitics": English *What?* vs. French *Quoi ?*)

Latin letters: "Phonemes" & "features"

- Eden (1961)

> [6] We wish to call attention to certain analogies between the structure presented here and that proposed for spoken language. Thus the segments are analogous to the distinctive features (Jakobson, Fant and Halle, *Preliminaries to speech analysis*, Cambridge, Massachusetts Institute of Technology Press, 1952); the strokes are analogous to phonemes, the letters to morphemes and the words to words. There is no counterpart in linguistics to our primitive notions of point pair ordering and angular rotation.

- Watt (1980)



Latin letters: “Phonological” patterns

- “Allophonic” patterns
  - Stroke length depends on position (Rezec 2009) 
  - Size is bottom-heavy more generally: B (cf. Chinese 自)
- Capitalization as “allomorphy” (Meletis 2020)
  - Usually depends on position (“alternations”): <Good gracious!> *<gOod>
    - So “morphology” (letter “compounding”) precedes “phonology” (letter form)
  - Otherwise capitalization is lexically conditioned
    - Common noun <gill> vs. proper name <Gill>, human <he> vs. godly <He>
    - German: adjective <arm> ‘poor’ vs. noun <Arm> ‘arm’

Latin letters: “Prosody”

- Size asymmetry (bottom-heavy)

  E B …

- Letters tend to “face” right (Watt 1980; Primus 2004)
  - < B C D E F G K L P Q R >, < b c e f h k m n p r >
  - Counterexamples: < J >, < a d g j q u y >
  - Symmetrical: < A H I M N O S T U V W X Y Z >, < i l o s t v w x z >
- The rightward facing asymmetry is “phonetically” motivated
  - Writers prefer to start strokes from other strokes (van Sommers 1984)
  - So Arabic favors left-facing letters (Primus & Wagner 2013):
    ـم /m/ (final allograph)

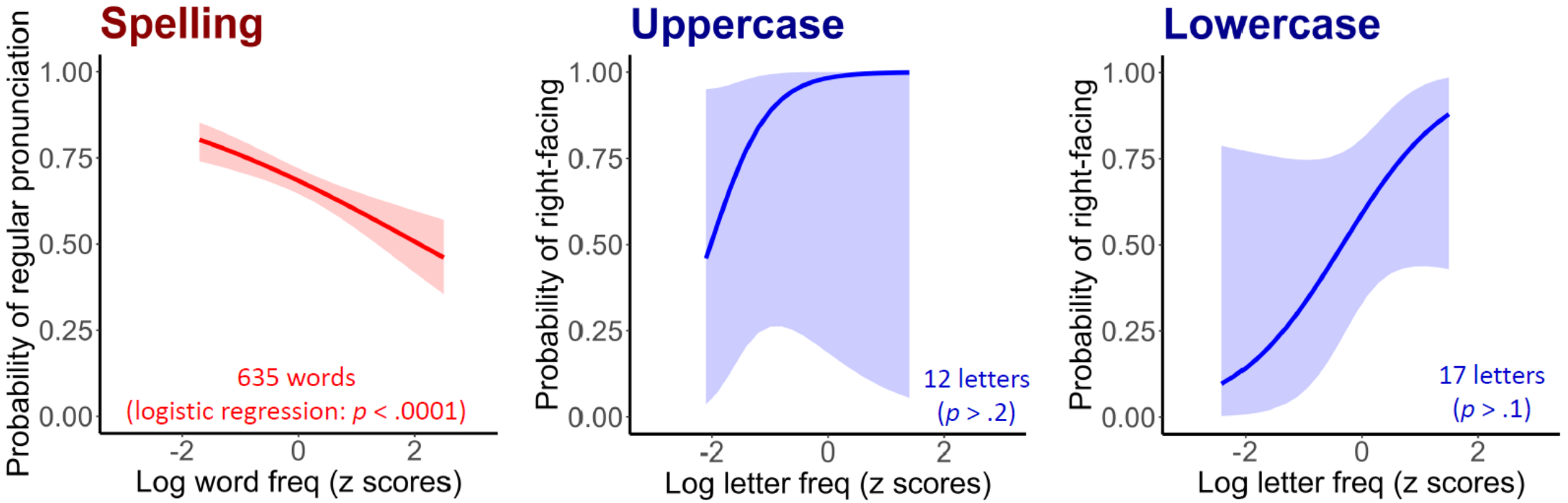
Latin letters: Frequency effects (1)

- Patterns in letter combinations ("morphology") should show negative frequency effects
  - Like regular pronunciations
  - Based on 635 English monosyllabic words in IPA (Brysbaert & New 2009, Carnegie Mellon University 2014, Constant 2012) containing 40 written rimes with regular and irregular pronunciations (Seidenberg & McClelland 1989)

    <cove> /ow/ vs. <love> /ʌ/ <braid> /ej/ vs. <said> /ɛ/
  - (See next slide)
- Patterns in letter forms ("phonology") should show positive frequency effects if "phonetically" motivated
  - Like facing rightward
  - (See next slide)

Latin letters: Frequency effects (2)

- Negative frequency effects for regularities in letter combinations
- Positive frequency effects for regularities in letter forms



(Myers 2026)

Overview

- Grammatical approaches to writing systems
- Morphology vs. phonology
- Case studies
  - Chinese characters
  - Latin letters
  - Devanagari ←
  - Others

Devanagari

- Used for Sanskrit, Hindi etc
- Abugida (AKA alphasyllabary)
  - Graphemes (akshara) represent V or $C+V_{default}$
  - So $C+V_{nondefault}$, CVC, CCV, ... require various sorts of complex graphemes

		k-	kh-	g-	gh-	ṅ-	c-	ch-	j-	jh-	ñ-	ṭ-	ṭh-	ḍ-	ḍh-	ṇ-	t-	th-	d-	dh-	n-	p-	ph-	b-	bh-	m-	y-	r-	l-	v-	ś-	ṣ-	s-	h-
a	अ	क	ख	ग	घ	ङ	च	छ	ज	झ	ञ	ट	ठ	ड	ढ	ण	त	थ	द	ध	न	प	फ	ब	भ	म	य	र	ल	व	श	ष	स	ह
ā	आ	का	खा	गा	घा	ङा	चा	छा	जा	झा	ञा	टा	ठा	डा	ढा	णा	ता	था	दा	धा	ना	पा	फा	बा	भा	मा	या	रा	ला	वा	शा	षा	सा	हा
i	इ	कि	खि	गि	घि	ङि	चि	छि	जि	झि	ञि	टि	ठि	डि	ढि	णि	ति	थि	दि	धि	नि	पि	फि	बि	भि	मि	यि	रि	लि	वि	शि	षि	सि	हि
ī	ई	की	खी	गी	घी	ङी	ची	छी	जी	झी	ञी	टी	ठी	डी	ढी	णी	ती	थी	दी	धी	नी	पी	फी	बी	भी	मी	यी	री	ली	वी	शी	षी	सी	ही
u	उ	कु	खु	गु	घु	ङु	चु	छु	जु	झु	ञु	टु	ठु	डु	ढु	णु	तु	थु	दु	धु	नु	पु	फु	बु	भु	मु	यु	रु	लु	वु	शु	षु	सु	हु
ū	ऊ	कू	खू	गू	घू	ङू	चू	छू	जू	झू	ञू	टू	ठू	डू	ढू	णू	तू	थू	दू	धू	नू	पू	फू	बू	भू	मू	यू	रू	लू	वू	शू	षू	सू	हू
e	ए	के	खे	गे	घे	ङे	चे	छे	जे	झे	ञे	टे	ठे	डे	ढे	णे	ते	थे	दे	धे	ने	पे	फे	बे	भे	मे	ये	रे	ले	वे	शे	षे	से	हे
ai	ऐ	कै	खै	गै	घै	ङै	चै	छै	जै	झै	ञै	टै	ठै	डै	ढै	णै	तै	थै	दै	धै	नै	पै	फै	बै	भै	मै	यै	रै	लै	वै	शै	षै	सै	है
o	ओ	को	खो	गो	घो	ङो	चो	छो	जो	झो	ञो	टो	ठो	डो	ढो	णो	तो	थो	दो	धो	नो	पो	फो	बो	भो	मो	यो	रो	लो	वो	शो	षो	सो	हो
au	औ	कौ	खौ	गौ	घौ	ङौ	चौ	छौ	जौ	झौ	ञौ	टौ	ठौ	डौ	ढौ	णौ	तौ	थौ	दौ	धौ	नौ	पौ	फौ	बौ	भौ	मौ	यौ	रौ	लौ	वौ	शौ	षौ	सौ	हौ
aṁ	अं	कं	खं	गं	घं	ङं	चं	छं	जं	झं	ञं	टं	ठं	डं	ढं	णं	तं	थं	दं	धं	नं	पं	फं	बं	भं	मं	यं	रं	लं	वं	शं	षं	सं	हं
aḥ	अः	कः	खः	गः	घः	ङः	चः	छः	जः	झः	ञः	टः	ठः	डः	ढः	णः	तः	थः	दः	धः	नः	पः	फः	बः	भः	मः	यः	रः	लः	वः	शः	षः	सः	हः

https://en.wikipedia.org/wiki/Devanagari

Devanagari: "Morphological" operations

- Gnanadesikan (2022)
  - "Affixation": base akshara + satellite vowels

क	का	कि	की	कु	कू	कृ	के	कै	को	कौ
ka	kā	ki	kī	ku	kū	kṛ	ke	kai	ko	kau

  - "Compounding": two base akshara

ब्य	क्र	क्ल	ल्क	ट्ट	प्न	श्ल	स्क	स्व	स्म	न्य	ष्ट	त्त्व
bya	kra	kla	lka	ṭṭa	pna	śla	ska	sva	sma	nya	ṣṭa	ttva

  - Both operations: compounding "before" affixation

ब्या	क्री	क्ले	ल्कु	ट्टू	प्नि	श्लो	स्कृ	स्वै	स्मा	न्यौ	ष्टि	च्छे
byā	krī	kle	lku	ṭṭū	pni	ślo	skṛ	svai	smā	nyau	ṣṭi	cche

- Bound
- Reduced
- Fixed positions*
- Closed class (V < C)
- (Abstract operators…?)

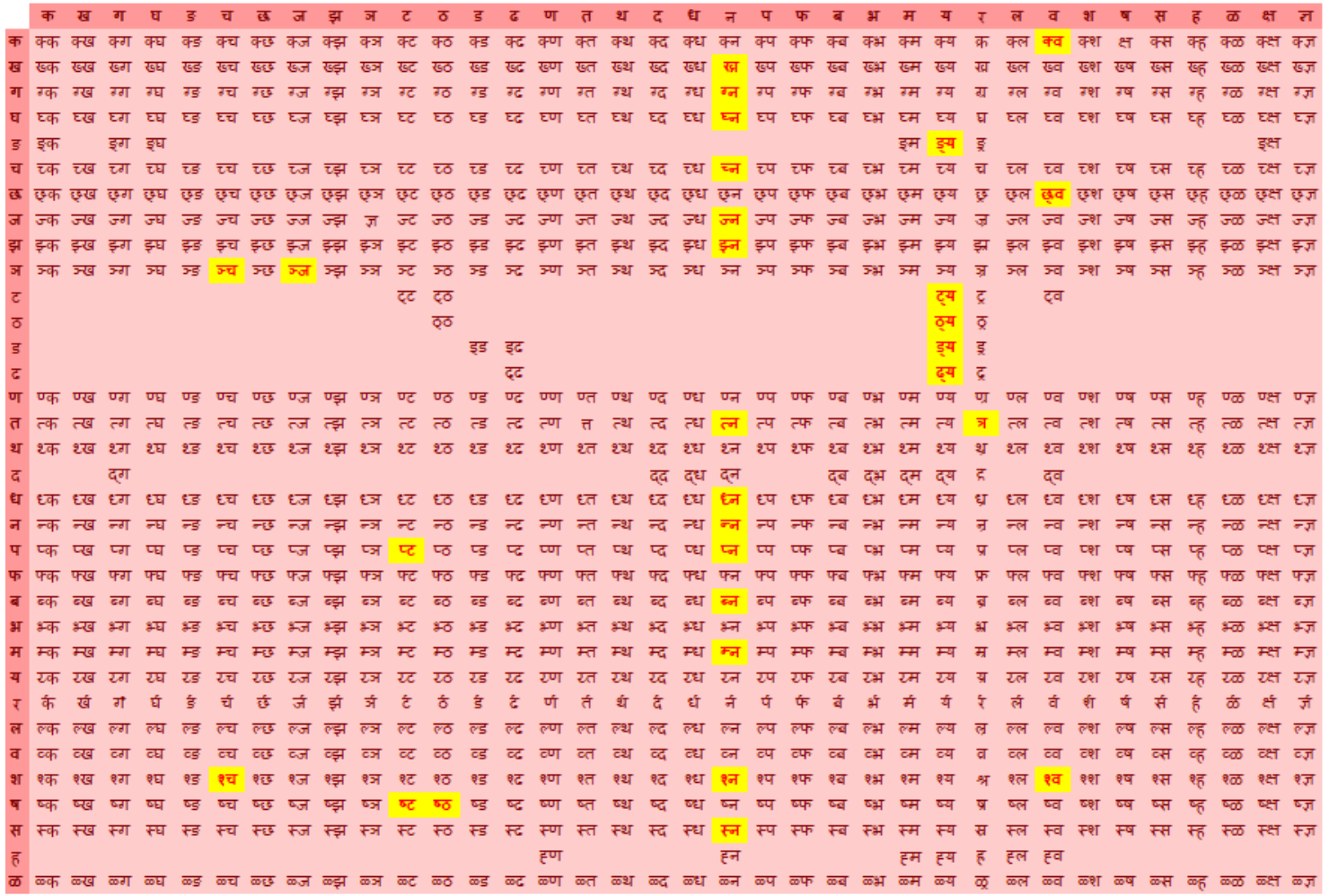
*(cf. "graphotactics": McCawley 1994)

स्म *sma* = स *sa* + म *ma*

स्मि *smi* = स्म *sma* + *i-*

Devanagari: “Compounding” irregularity

- “Compounding” has lexical exceptions



https://en.wikipedia.org/wiki/Devanagari_conjuncts

Regular

त *ta* + म *ma* = त्म *tma*

Irregular

त *ta* + र ra = त्र *tra*

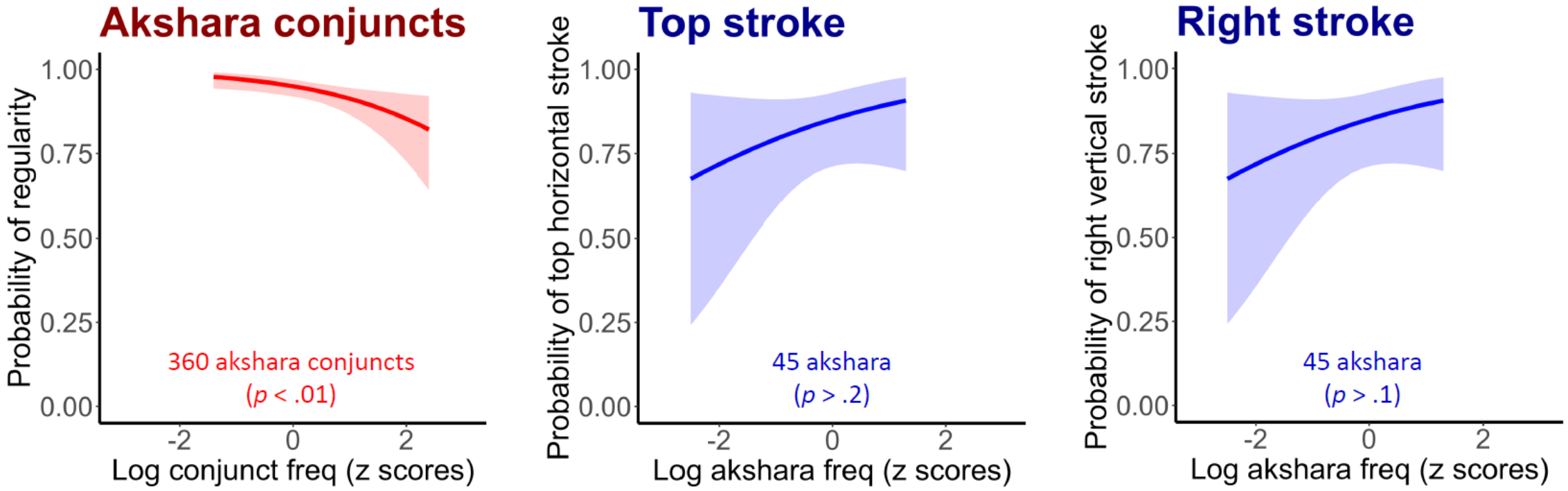
Devanagari: "Phonological" patterns

- Horizontal stroke (shirorekha) all the way across the top of each akshara
  - Shirorekha "assimilates" within a word: "phonetically" motivated?
- Vertical stroke on the right - written last, but "prosodically" motivated...?

क	ख	ग	घ	ङ	च	छ	ज	झ	ञ	ट	ठ	ड	ढ	ण	त	थ	द	ध
k-	*kh-*	*g-*	*gh-*	*ṅ-*	*c-*	*ch-*	*j-*	*jh-*	*ñ-*	*ṭ-*	*ṭh-*	*ḍ-*	*ḍh-*	*ṇ-*	*t-*	*th-*	*d-*	*dh-*
न	प	फ	ब	भ	म	य	र	ल	व	श	ष	स	ह					
n-	*p-*	*ph-*	*b-*	*bh-*	*m-*	*y-*	*r-*	*l-*	*v-*	*ś-*	*ṣ-*	*s-*	*h-*					
अ	आ	इ	ई	उ	ऊ	ए	ऐ	ओ	औ	अं	अः							
a	*ā*	*i*	*ī*	*u*	*ū*	*e*	*ai*	*o*	*au*	*aṁ*	*aḥ*							

Devanagari: Frequency effects

- Negative frequency effects for regularities in akshara combinations
- Positive frequency effects for regularities in akshara forms



(Myers 2026)

(frequency calculations based on Sanskrit: *Sanskrit frequency list* n.d., Stiehl 2007)

Overview

- Grammatical approaches to writing systems
- Morphology vs. phonology
- Case studies
  - Chinese characters
  - Latin letters
  - Devanagari
  - Others ←

Egyptian hieroglyphs: Morphology but no phonology?

- "Morphology"
  - Like Chinese, a closed class of abstract semantic "determinatives" affixed to phonetic forms in the same position, typically on the end (Ritner 1996)
- But was there any "phonology"...?
  - Emoji-like detail and iconicity:  'eat'  'embrace'  '40,000'
  - Some global constraints (e.g. facing left):  'frog'  'bee'
  - And a (very) few non-iconic (?) minimal pairs:  'great'  'bad'
- ...Not necessarily
  - Phonology has to evolve, as in natural sign languages (Sandler et al. 2011)
  - And iconicity can complement formal phonology, again as in natural sign languages (Eccarius & Brentari 2010)

Yi: Phonology but no morphology?

- Standardized script of some Tibeto-Burman languages in SW China
- A true syllabary (Walters et al. 2015, *Yi script index book* 1984)

  bit, *pit*, *bi*, *pi*, *bot*, *pot*, *lop*, *jot*, ...
- Syllabograms are combined in accordance with the spoken language, so there are no written "morphological" operations
- "Phonology"
  - Syllabogram shapes are classed by conventional indices

    E.g. *njyt* and *jip* both have index
  - Index classes vary in size (number of syllabogram types)

    E.g. *siep* [index , index size 6] vs. la [index , index size 70]
  - Syllabograms in larger indexing classes have less marked "phonology"
  - And they show the expected positive frequency effects (see appendix)

Conclusions (or are they…?)

- Writing has distinct morphology-like and phonology-like subsystems
  - The same distinction is found in otherwise quite different writing systems
- Why???
  - Did the biological evolution of language result in a relatively amodal system?
  - Is the cultural evolution of language restricted by an intrinsic solution space?
- What are “morphology” and “phonology” anyway?
  - Sign languages forced a rethink (still too often neglected)
  - Writing systems may require further rethinking

References (1/5)

Althaus (1973) Graphetik. In Althaus et al. (Eds.), *Lexikon der germanistischen Linguistik*. Niemeyer.

Altmann (1980) Prolegomena to Menzerath's Law. In Grotjahn (Ed.), *Glottometrika* 2. Brockmeyer.

Behr (2006) Homosomatic juxtaposition and the problem of 'syssymantic' (huìyì) characters. In Bottéro & Djamouri (Eds.), *Écriture chinoise: données, usages et représentations*. Paris: École des hautes études en sciences sociales, Centre de recherches linguistiques sur l'Asie orientale.

Bloomfield (1933) *Language*. Holt.

Browman & Goldstein (1988) Some notes on syllable structure in articulatory phonology. *Phonetica* 45.

Brysbaert & New (2009) Moving beyond Kučera and Francis. *Behavior Research Methods* 41.4.

Bybee (2001) *Phonology and language use*. Cambridge University Press.

Carnegie Mellon University (2014) The CMU pronouncing dictionary. http://www.speech.cs.cmu.edu/cgi-bin/cmudict.

Channon (2002) Signs are single segments: Phonological representations and temporal sequencing in ASL and other sign languages. *Sign Language & Linguistics* 5.1.

Constant (2012) The CMU pronouncing dictionary in IPA. https://people.umass.edu/nconstan/CMU-IPA/

Coulmas (2002) *Writing systems: An introduction to their linguistic analysis*. Cambridge University Press.

Daniels & Bright (1996) *The world's writing systems*. Oxford University Press.

Dehaene & Cohen (2011) The unique role of the visual word form area in reading. *Trends in Cognitive Sciences* 15.6.

Dressler et al. (2005) Reduplication in child language. In Hurch (Ed.) *Studies on reduplication*. De Gruyter Mouton.

References (2/5)

Eccarius & Brentari (2010) A formal analysis of phonological contrast and iconicity in sign language handshapes. *Sign Language & Linguistics* 13.2.

Eden (1961) On the formalization of handwriting. In Jakobson (Ed.), *Structure of language and its mathematical aspects*. American Mathematical Society.

Evertz (2018) *Visual prosody: The graphematic foot in English and German*. Walter de Gruyter.

Feldman & Siok (1999) Semantic radicals contribute to the visual identification of Chinese characters. *Journal of Memory and Language* 40.4.

Fuhrhop et al. (2011) The length hierarchy and the graphematic syllable: Evidence from German and English. *Written Language & Literacy* 14.2.

Fuhrhop & Peters (2016) *Einführung in die Phonologie und Graphematik*. J.B. Metzler

Gelb (1963) *A study of writing* (revised edition). University of Chicago Press.

Gnanadesikan (2022) Amodal morphology: Applications to Brahmic scripts and Canadian Aboriginal syllabics. In Haralambous (Ed.), *Grapholinguistics in the 21st Century 2022*, Fluxus Editions.

Gnanadesikan (2023) Segments and syllables in Hangeul and Thaana: A comparison and optimality theoretic analysis. *Written Language & Literacy* 26.2.

Gnanadesikan (2025) *The writing revolution: Cuneiform to the internet* (second edition). Wiley Blackwell.

Goldsmith (1990) *Autosegmental and metrical phonology*. Basil Blackwell.

Hay et al. (2015) Tracking word frequency effects through 130 years of sound change. *Cognition* 139.

Ho (2016) Letters from Hong Kong: Characters under the Cantonese umbrella. *Cha: An Asian Literary Journal* 33. www.asiancha.com/content/view/2483/564/

References (3/5)

Huang [黄德宽] (2003) 汉字构形方式的动态分析. 安徽大学学报(哲学社会科学版) 27.4.

Hurch (Ed.) (2005) *Studies on reduplication*. Walter de Gruyter.

Jackendoff (2010) The ecology of English noun-noun compounds. In Jackendoff (Ed.) *Meaning and the lexicon*, Oxford University Press

Liu [刘曼] (2008) 叠体字的历时考察与认知比较研究. Ph.D. thesis, Tsinghua University, Beijing.

Martinussen & Scheike (2006) *Dynamic regression models for survival data*. Springer.

McCawley (1994) Some graphotactic constraints. In Watt (Ed.), Writing systems and cognition. Springer.

Meletis (2019) The grapheme as a universal basic unit of writing. *Writing Systems Research* 11.1.

Meletis (2020b) Types of allography. *Open Linguistics* 6.

Meletis & Dürscheid (2022) *Writing systems and their use: An overview of grapholinguistics*. De Gruyter Brill.

Mielke (2008) *The emergence of distinctive features*. Oxford University Press.

Mutaka, N., & Hyman, L. M. (1990) Syllables and morpheme integrity in Kinande reduplication. *Phonology* 7.1.

Myers (2016) Knowing Chinese character grammar. *Cognition* 147.

Myers (2019) *The grammar of Chinese characters*. Routledge.

Myers (2021) Levels of structure within Chinese character constituents. In Haralambous (Ed.) *Grapholinguistics in the 21st Century: From graphemes to knowledge, Part II*. Fluxus Editions.

References (4/5)

Myers (2023) Nonstatistical things that linguists can do with R: A review with a case study on Chinese handwriting. In Lin et al. (Eds.) *Proceedings of the 34th North American Conference on Chinese Linguistics* (NACCL-34).

Myers (2024) Interactions among patterns in Chinese character form. In Myers & Tai (Eds.) *Looking at language from all sides*. Crane Publishing.

Myers (2024 [2026]) Perception and production of a feature correlation in Chinese characters. *Written Language and Literacy* 27.2.

Myers (2026) Two levels of regularity in writing systems: Evidence from frequency effects. CLA, Calgary.

Pacton et al. (2001) Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General* 130.3.

Padden & Perlmutter (1987) American Sign Language and the architecture of phonological theory. *Natural Language & Linguistic Theory* 5.3.

Perlmutter (1992) Sonority and syllable structure in American Sign Language. *Linguistic Inquiry* 23.

Phillips (2006) *Word frequency and lexical diffusion*. Palgrave MacMillan.

Pierrehumbert (2002) Word-specific phonetics. In Gussenhoven & Warner (Eds.), *Laboratory phonology* 7. De Gruyter Mouton.

Primus (2004) A featural analysis of the Modern Roman Alphabet. *Written Language & Literacy* 7.2.

Primus & Wagner (2013) Buchstabenkomposition. In Ôhashi & Roussel (Eds.) *Buchstaben der Welt – Welt der Buchstaben*. Wilhelm Fink Verlag.

Reinken (2023) Handwritten letters and grammatical structures in German: A framework. *Written Language & Literacy* 26.2.

Rezec (2009) *Zur Struktur des deutschen Schriftsystems*. Ludwig-Maximilians-Universität Ph.D. thesis.

References (5/5)

Ritner (1996) Egyptian writing. In Daniels & Bright.

Sampson (1985) *Writing systems: A linguistic introduction*. Stanford University Press.

Sandler et al. (2011) The gradual emergence of phonological form in a new language. *Natural Language & Linguistic Theory* 29.2.

Sanskrit frequency list (n.d.) en.wiktionary.org/wiki/Appendix:Sanskrit_frequency_lists.

Seidenberg & McClelland (1989) A distributed, developmental model of word recognition and naming. *Psychological Review* 96.4.

Sorace (2003) Near-nativeness. In Doughty & Long (Eds.), *The handbook of second language acquisition*. Blackwell.

Stiehl (2008) *Sanskrit-Kompendium*. Economica Verlag.

Tsai (2006) Frequency and stroke counts of Chinese characters. technology.chtsai.org/charfreq/.

van Sommers (1984) *Drawing and cognition*. Cambridge University Press.

Walters et al. (2015) A preliminary study of Nuosu Yi syllable frequency in text. SIL International ms. www.sil.org/resources/archives/61553.

Wang (1983) *Toward a generative grammar of Chinese character structure and stroke or*der. University of Wisconsin-Madison Ph.D. thesis.

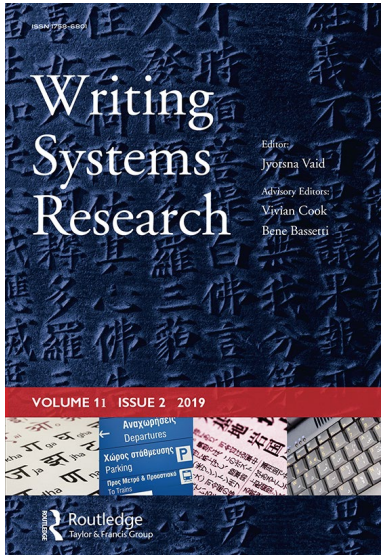
Watt (1980) What is the proper characterization of the alphabet? II: Composition. *Ars Semeiotica* 3.1.

Wilbur (2009) Productive reduplication in a fundamentally monosyllabic language. *Language Sciences* 31.

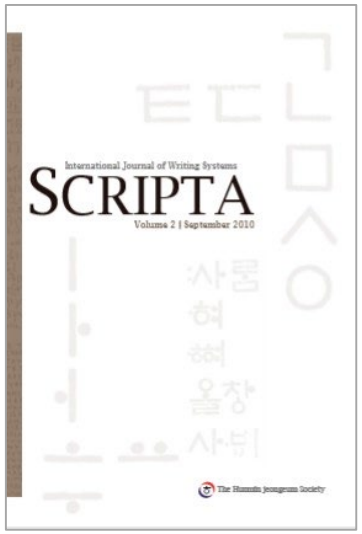
Yi script index book (1984) 彝文检字本. Sichuan Nationalities Publishing House.

The rest of the story…

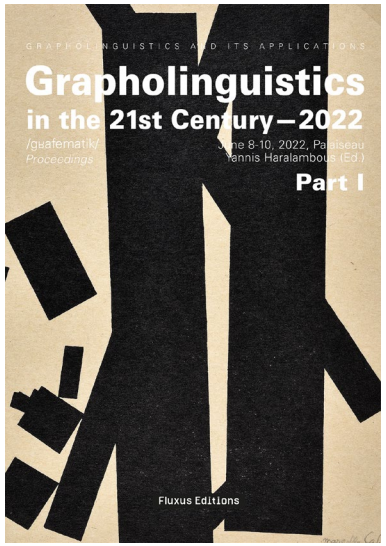
The linguistics of writing systems (2)

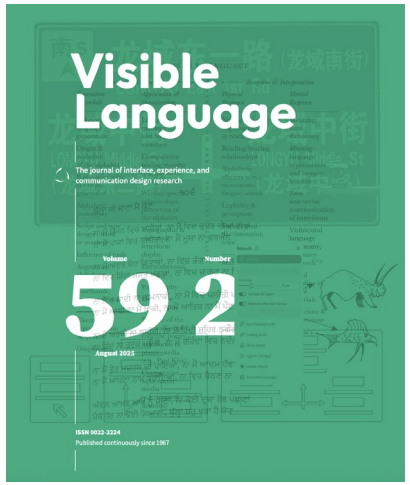


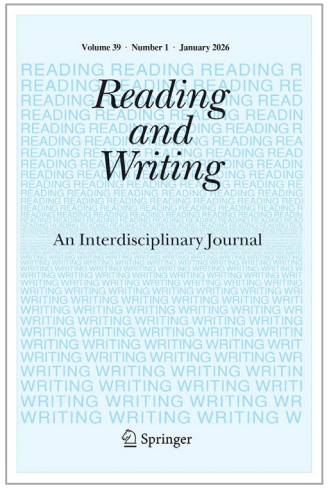
(defunct!)

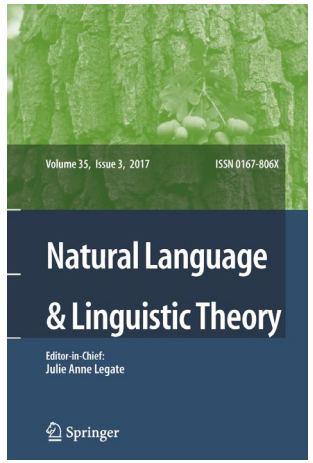


(defunct?)









(the current editor is “one of us”)

Association for Written Language and Literacy
Ассоциация письменного языка и письменности
Ενωση για τη γραπτη γλωσσα και τη βασικη εκπαιδευση
书面语言和识字协会
Association for Written Language and Literacy

15th AWLL international workshop on writing systems and literacy

Pisa, 29-31 October 2025

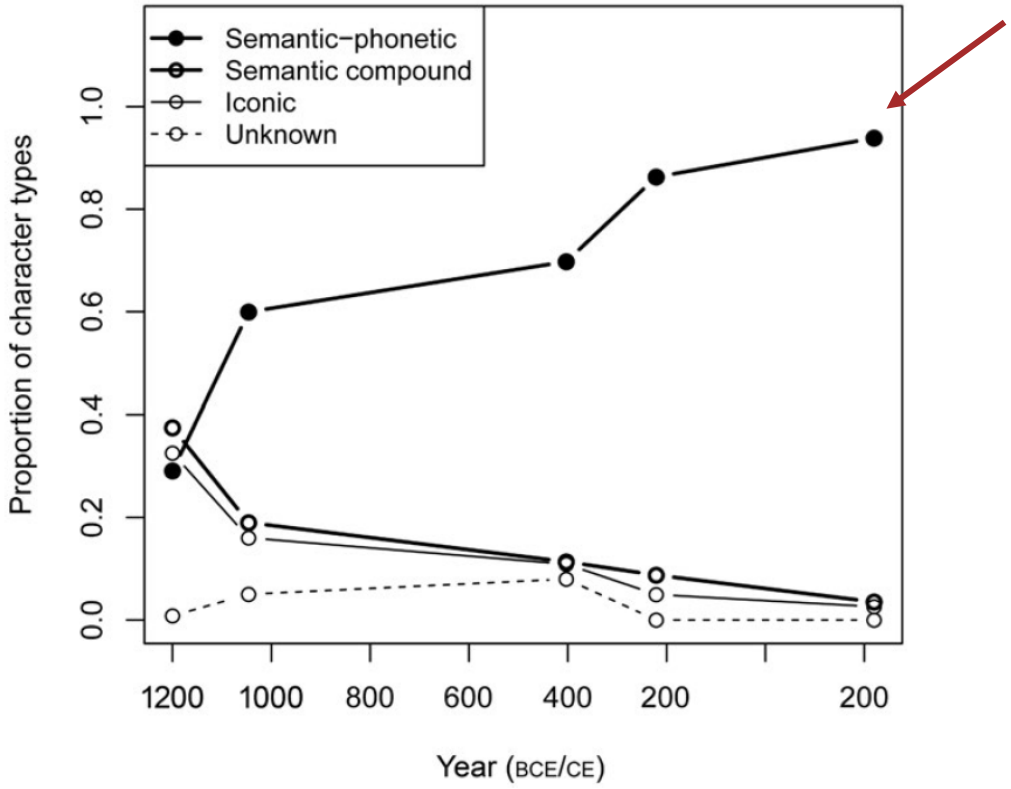
Istituto di Linguistica Computazionale "Antonio Zampolli"
Consiglio Nazionale delle Ricerche

The “naturalness” question

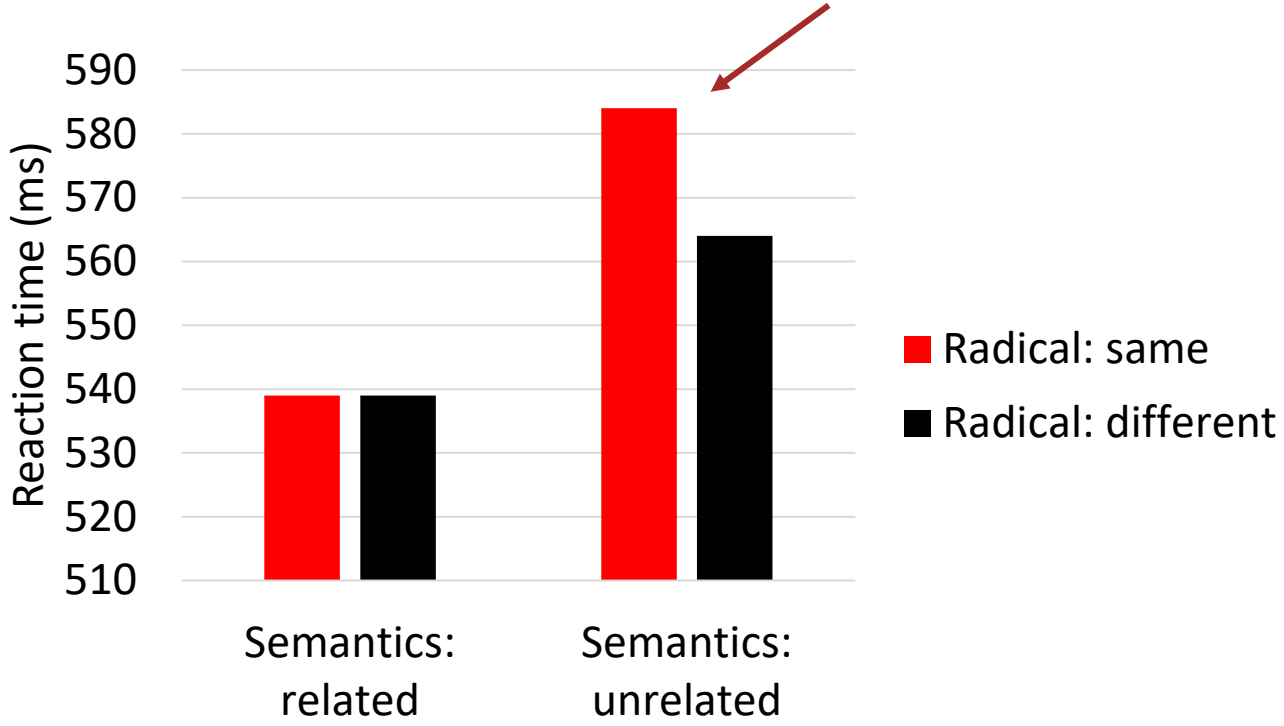
- Writing is unnatural...
  - All human cultures have speech/signing, but not all have writing
  - Writing was invented too recently to have affected human brain evolution
  - Children readily acquire speech/signing, but writing requires effortful teaching
- ... but not all *that* unnatural
  - In practice, reading and writing must also engage the innate linguistic capacity
  - L2 learning is also effortful, yet still engages innate capacities (Sorace 2003)
  - Children learn some writing system patterns implicitly (Pacton et al. 2001)
  - The brain has coopted a universal spot for reading: the Visual Word Form Area (Dehaene & Cohen 2011)

Chinese: "Affixation" is highly productive

- "Affixation" has been the dominant operation for centuries
  - And readers expect "affixes" to relate to character meaning



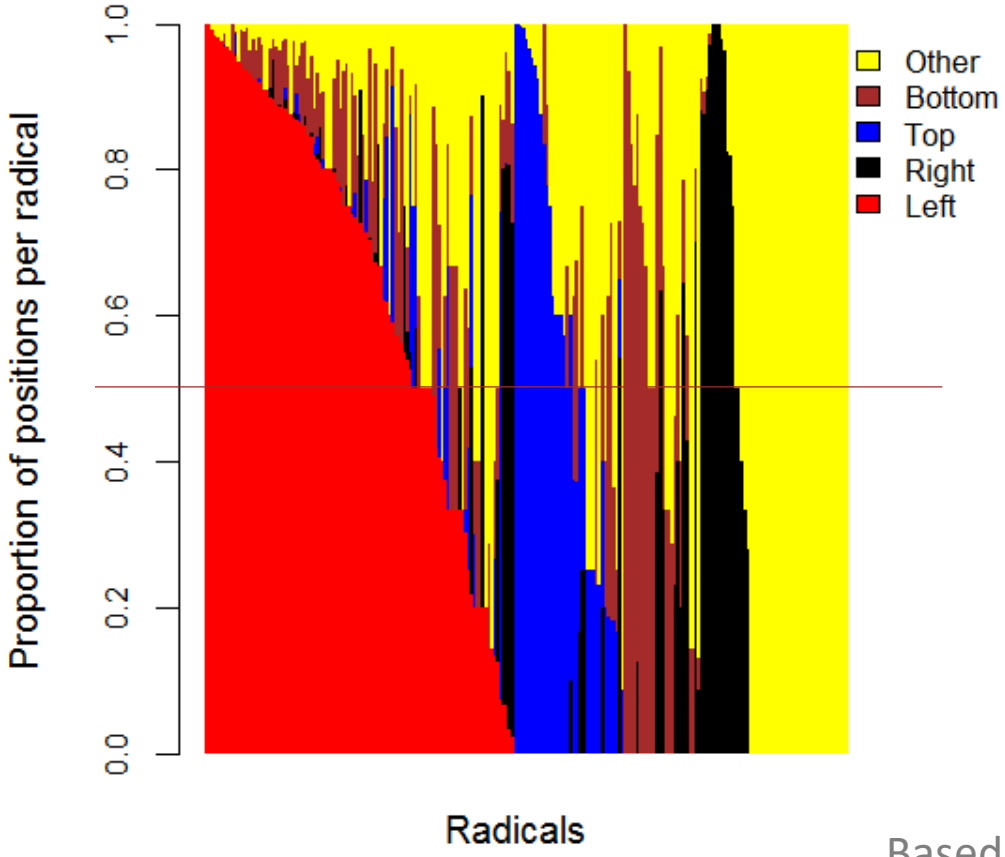
Myers (2019), based on data in Huang (2003)

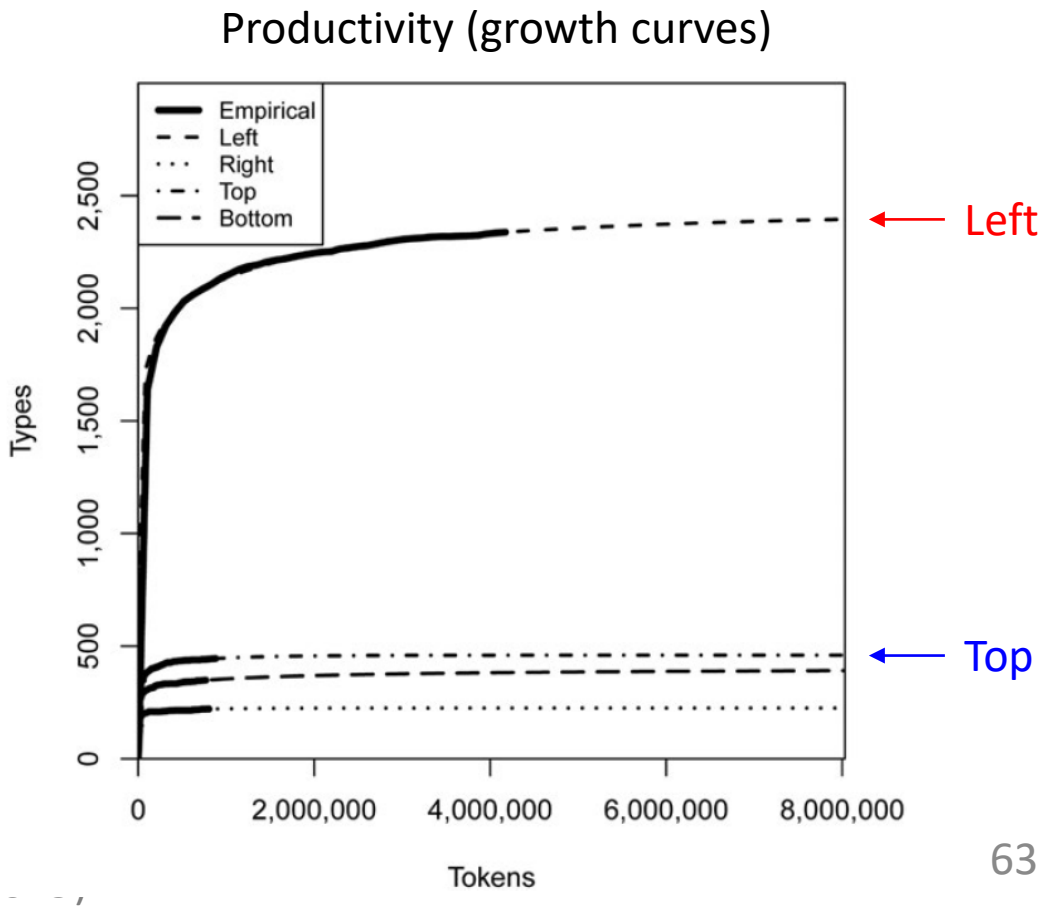


Based on data in Feldman & Siok (1999)

Chinese: "Affixation" prefers certain positions

- "Affixes" are usually (>50%) in the same position
  - And the left edge has a special status





Based on Myers (2

Chinese: “Compounding” is sort of productive

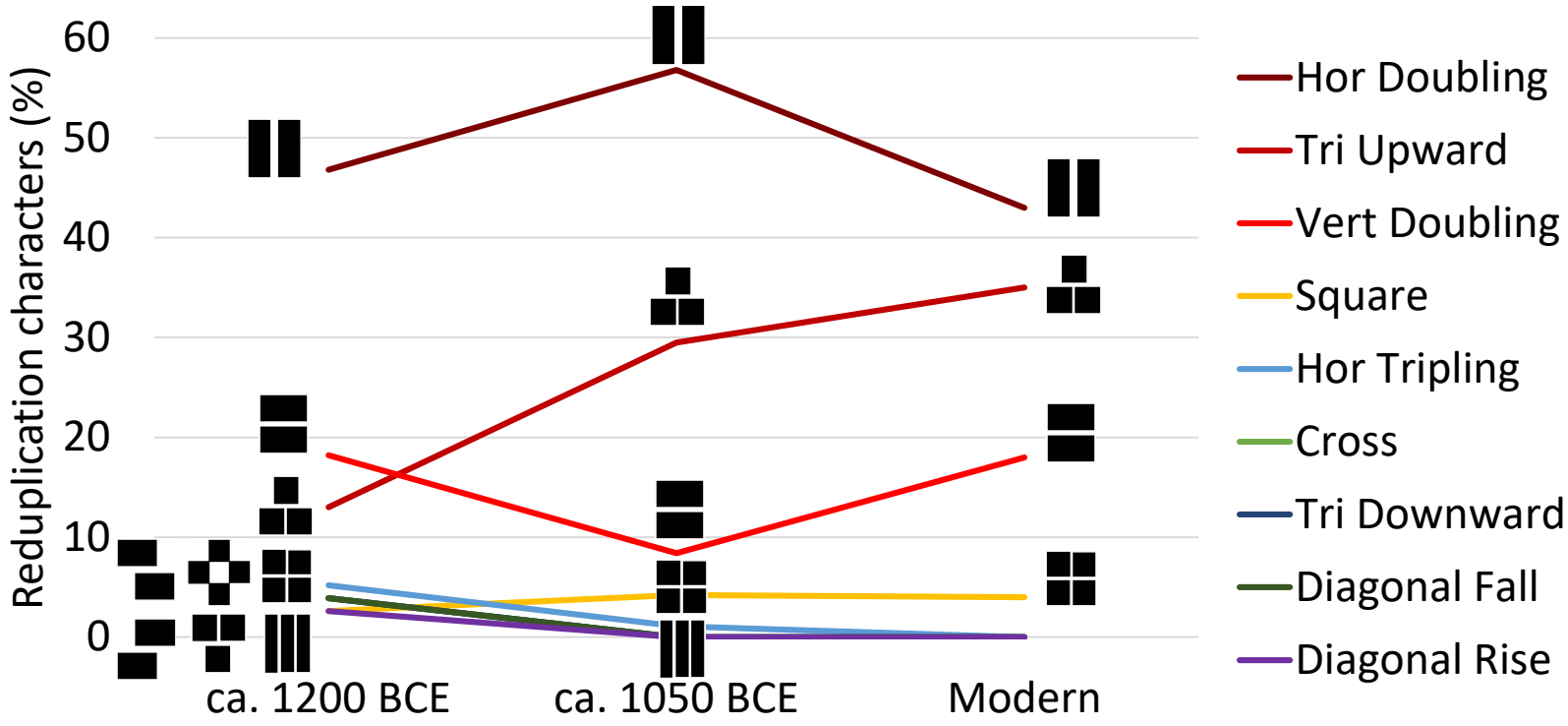
- Most “naturally” coined characters are “affixed”
- But “creatively” coined characters are often compounds
  - 圖 *túshūguǎn* ‘library’ (口 ‘enclosure’ + 書 ‘book’)
  - 招財進寶 *zhāo cái jìn bǎo* (expression of wishes for wealth and success)
    - 撐 *caang1* ‘support’ + 傘 *saan3* ‘umbrella’

https://commons.wikimedia.org/wiki/File:招財進寶 AR PL UKai CN 288pt.svg

(photo by Jason S Polley, 2014)

Ho (2016)
https://www.asiancha.com/content/view/2483/564/

Chinese: “Reduplication” is productive (1)



(data from Behr 2006; Liu 2008; Tsai 2006)

- Simplified characters may fill these templates with empty morphs

  双 < 雙 *shuāng* ‘pair’ 轰 < 轟 *hōng* ‘bang!’

Chinese: "Reduplication" is productive (2)

Sample test items.

	Lexical grammatical	Lexical ungrammatical	Nonlexical grammatical	Nonlexical ungrammatical
Horizontal	森	蒸	鼓	鼓
Reduplication	林 *lín* 'woods'		NA	
Element	木 *mù* 'wood'		支 *zhī* 'branch'	
Vertical	侈	侈	侇	侇
Reduplication	多 *duō* 'more'		NA	
Element	夕 *xī* 'evening'		夫 *fū* 'husband'	
Triangular	媂	媂	嫈	嫈
Reduplication	晶 *jīng* 'crystal'		NA	
Element	日 *rì* 'sun, day'		欠 *qiàn* 'owe'	

(Myers, 2016)

Chinese: "Reduplication" is productive (3)

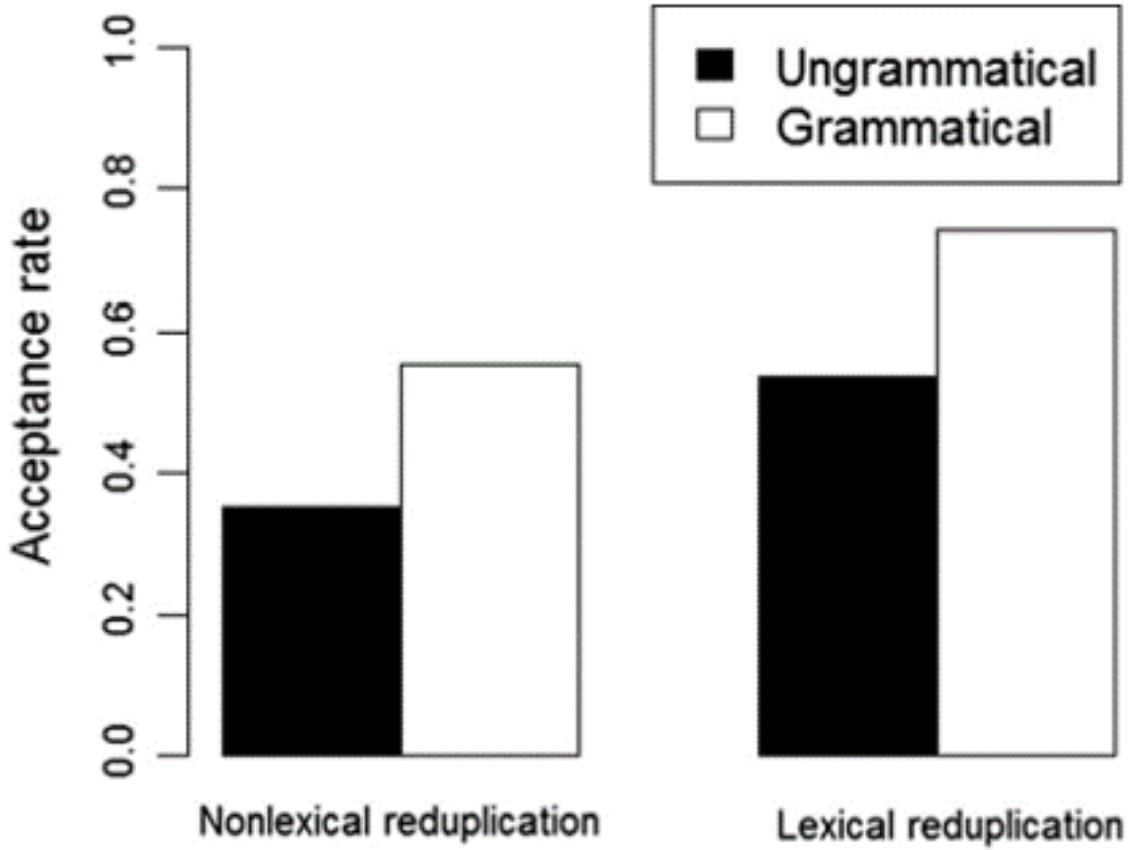
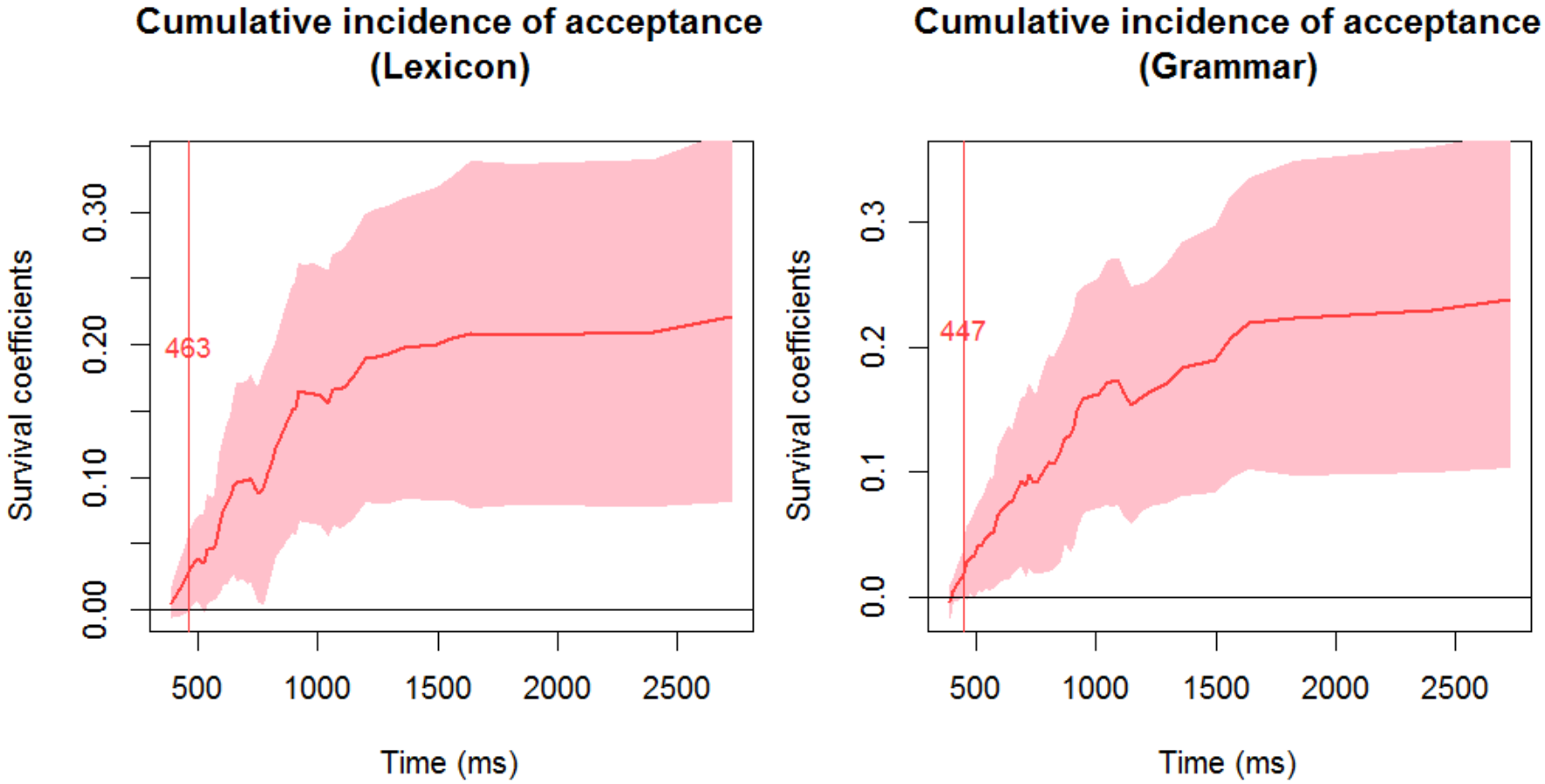


Fig. 1. Acceptance rates for fake characters containing lexical/nonlexical and grammatical/ungrammatical reduplicative configurations.

(Myers, 2016)

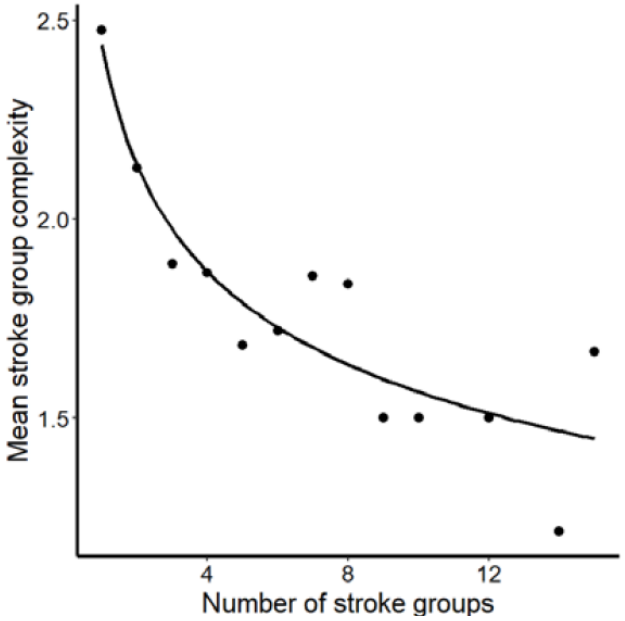
Chinese: "Reduplication" is productive (4)



(dynamic survival analysis [Martinussen & Scheike 2006];
plot after Myers 2019)

Chinese: "Prosody" (4)

- Stroke groups as "syllables"? (Myers 2021)
  - Targets of "stress": 自 *duī* 'heap' 由 *yóu* 'from' 甲 *jiǎ* 'helmet'
  - Gestural complexes
    - Spoken rimes are coordinated with onsets (Browman & Goldstein 1988)
    - Writers prefer to start strokes from other strokes (van Sommers 1984)
  - So 工 *gōng* 'work' is "disyllabic": 丅一
- Stroke groups compete with each other in accordance with Menzerath's law (Altmann 1980)



Chinese: “Morphological” interactions

- “Reduplication” only applies to base components (including bound)
  - Base components: 多 昌 炎 圭　　林 朋 比 弱 羽 艸　　品 森 蟲 晶 轟
- “Compounding” only applies to bases (including bound) & reduplication
  - Base components: 明 相 尋
  - Reduplicated forms: 區 雙 器 替 琵
- “Affixation” can apply to any type of form (except bound)
  - Base components: 住 根 英
  - Reduplicated forms: 錢 琳 臨
  - Compounds: 請 想 份
  - Affixed forms (recursion): 燙 榴 瞇

Latin letters from a Chinese perspective

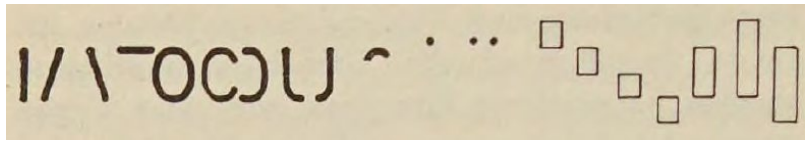
- I'm going against the German graphematic tradition here
  (including Fuhrhop et al. 2011 and all the other citations in these slides)
  - They see phonology, spelling, and even letter form as one system
  - But this seems too alphabet-centric to me
- So I'm trying to see how far I can push the idea that "morphology" = interpreted form system
  - In Chinese characters, semantic components are interpreted semantically, like real "morphemes"
  - But other components are also interpreted, just in terms of pronunciation
- Alphabetic letters are also interpreted in terms of pronunciation

Fuhrhop & Peters (2016)

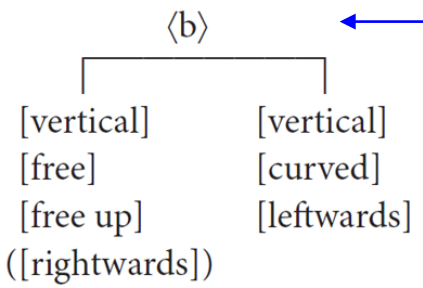
Latin letters: "Prosody" (2)

- Stroke positions have long been recognized as crucial in letter forms
  - Althaus (1973)

Elements Positions



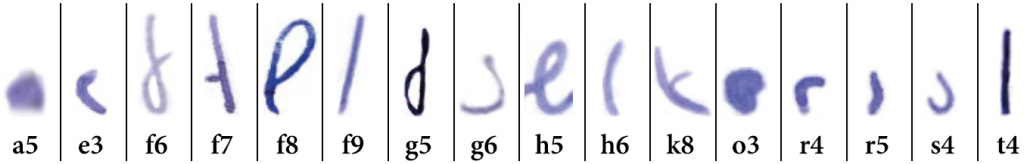
Primus (2004)



l in "head" position (must be vertical)
ɔ in "coda" position

(features)

- Only the head is obligatory (Reinken 2023)



- My speculation: Analogues to specific prosodic levels?
  - Size asymmetry = "binary stress feet"?
  - Body/coda asymmetry = "syllables" (as in Chinese stroke combinations)?
    - Writers prefer to start strokes from other strokes: **T E**
    - So letter heads = "onsets", codas = "rimes" (start from heads)...?

Devanagari: “Inflection” vs. “derivation”

- Gnanadesikan (2022)
  - “Inflection” via satellite vowel doesn’t change base akshara interpretation

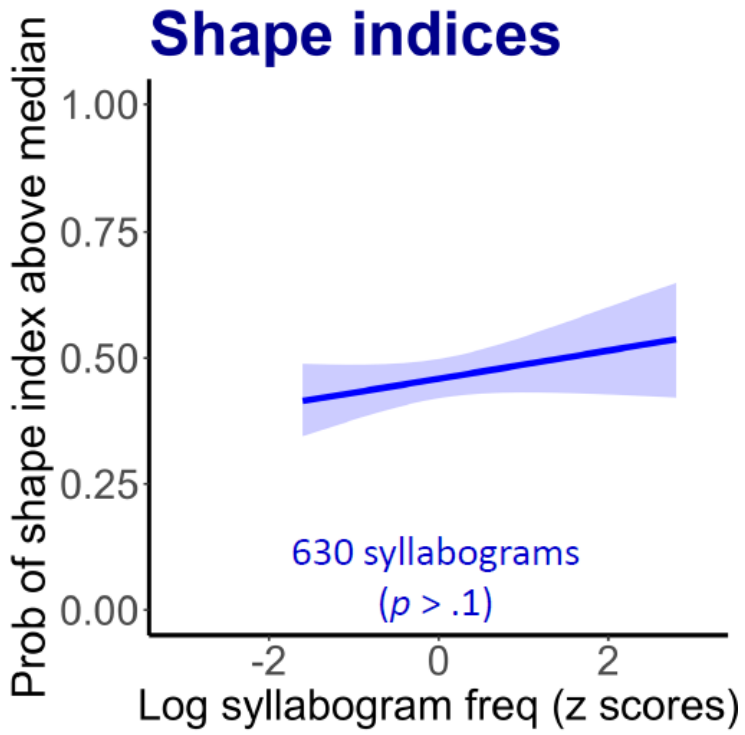
क	का	कि	की	कु	कू	कृ	के	कै	को	कौ
ka	kā	ki	kī	ku	kū	kṛ	ke	kai	ko	kau

  - “Derivation” via dotting does change base akshara interpretation

अ	अं	अः
a	*aṁ*	*aḥ*

Yi: Frequency effects

- Positive frequency effects for syllabogram index class sizes
  - The higher the token frequency, the larger the class



(Myers 2026)