Prosodic structure in Chinese characters, revisited

James Myers
National Chung Cheng University
Lngmyers@ccu.edu.tw

ILAS Workshop on Phonetics and Phonology – October 2017

Thanks!

- Grants from Taiwan's National Science Council / Ministry of Science and Technology:
  - NSC 97-2410-H-194-067-MY3, NSC101-2410-H-194-115-MY3, MOST 103-2410-H-194-119-MY3
- Many lab assistants over the years:
  - Ko Yuguang, Yang Chentsung, Hsu Chiung-Wen, Hong Guo-Ming, Hsu Zi-Ping, Du Pei-Fen, Su Yu-Ting, Liu Yu-Jay, Chuang Wei-Chiang, Hsieh Yu-Yi, Pan Hsiao-Yin, Su Si-Qi
- Plus:
  - Chen Tsung-Ying, Zev Handel, Wolfgang Behr, Jane Tsay, Niina Zhang, and Lin Yu-Hsuan

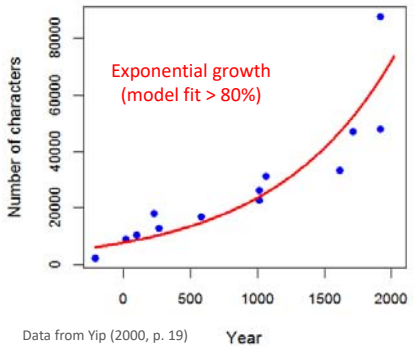
Overview

- Character "prosody"
- Radical position and reduction
- Reduplication
- Stroke shape and position
  - Lengthened strokes
  - Curved strokes
  - Hooked strokes

Overview

- **Character "prosody"**
- Radical position and reduction
- Reduplication
- Stroke shape and position
  - Lengthened strokes
  - Curved strokes
  - Hooked strokes

Chinese characters are productive



Productivity implies grammar….

Data from Yip (2000, p. 19)

Chinese character grammar

- Duality of patterning
  能 月 公 北
- Recursion (Sproat, 2000)
  鱗 魚→[米↓[夕→[十]]]

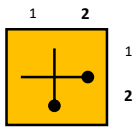
"Morphology"

- Shape regularities (Wang, 1983)
  牛→特 cf. 牢
- "Prosody": global shape constraints
  - Myers (1996, 2016)

"Phonology"

Character prosody

- Chinese characters tend to show:
  - Binarity
  - Asymmetrical edge prominence (right and bottom)
  - These have wide-ranging effects on components & strokes
- Similar to metrical feet in spoken & signed languages
  - Two-dimensional "feet":



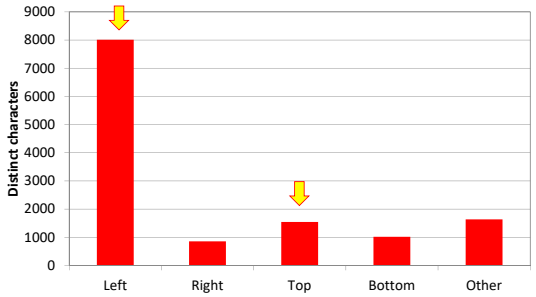
Overview

- Character "prosody"
- **Radical position and reduction**
- Reduplication
- Stroke shape and position
  - Lengthened strokes
  - Curved strokes
  - Hooked strokes

Radicals (部首) as affixes

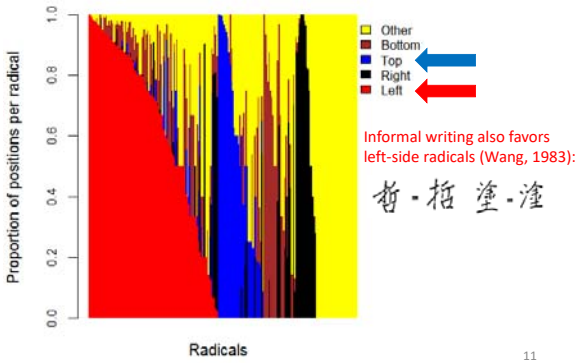
- Morphological properties
  - Closed class
  - Bound
  - Semantically bleached
- Phonological properties
  - Favor prosodically "weak" positions
  - Suppletive allomorphs
  - General reduction rules

Most characters have left/top radicals



Computed from 13,060 traditional characters in Tsai (2006)

Most radicals favor the left or top

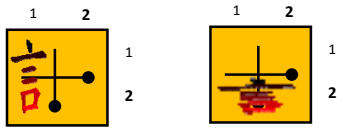


Informal writing also favors left-side radicals (Wang, 1983):

哲-揩 塗-湼

The left/bottom alternation

- Radicals that most favor the left also sometimes appear on the bottom (see previous slide)
  詞～警
- A prosodic explanation
  Left-sided radicals are "light" horizontally but "heavy" vertically:



Radicals may reduce in "weak" position

- Some radicals have suppletive forms at left/top
  心～忙　人～位　水～泊　手～拾
  艸～花　竹～箪
- But not at right or bottom (generally)
  忙～忘　泊～泉　拾～拿　狗～狀 (cf. 分～刻　燙～熱)
- Other radicals reduce in a more regular way
  - Wang (1983): Diagonalization and stroke reduction
    金：鉛～鑒　土：地～型　牛：物～犧
    木：村～果　米：精～粱　衣：被～裝
  - May even apply to non-radicals: 旦：助～宣

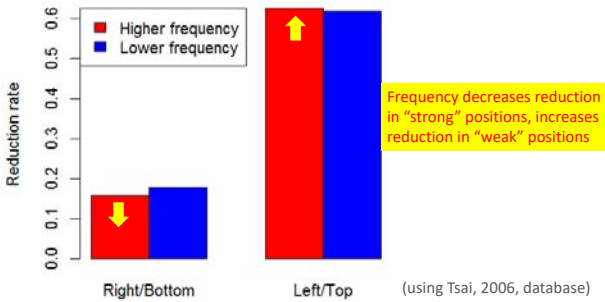
Grammaticalization of radicals

- Vietnamese (Chữ Nôm) radicals acted more like roots
  - Often "synonymic", not "taxonymic" (Handel, 2016; cf. ancient logographic systems: Rude, 1986)
  - Also often lacked formal reduction: 扡 *tay* 'arm'

Traditional characters go here

- PRC simplification: "Weak" reduction is extended
  词～警　铅～鉴　红～累
  (cf. 詞～警　鉛～鑒　紅～纍)

Frequency affects grammaticalization



Frequency decreases reduction in "strong" positions, increases reduction in "weak" positions

(using Tsai, 2006, database)

(Mixed-effects logistic regression by radical: Left/top vs. right/bottom : $p < .0001$; interaction between frequency and position only marginal: $p = .06$)

Testing radical position productivity

- Chinese readers were shown fake characters and asked to judge if they were like characters (「像中文字嗎？」)
- Lexicality: Component is/isn't a "real" radical
- Grammaticality: Is/isn't in "weak" position (left/top)

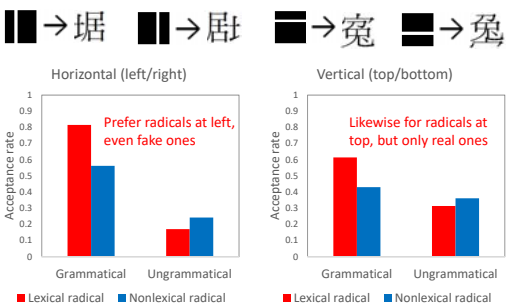
	Lex		Nonlex	
	Gram	Ungram	Gram	Ungram
Left/Right	稟	黌	牒	黷
Top/Bottom	薗	鹵	齒	鹵

- Problem: Position is confounded with reduction

So let's run a new experiment!

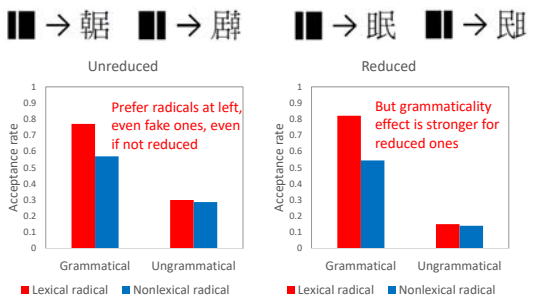
Created by Chen Tsung-Ying (Chen & Myers, 2017); this experiment was run with him too

Priming to boost position effect?



Mixed-effects logistic regression on all items: Grammaticality ($p < .0001$); Lexicality ($p < .05$); Grammaticality x Orientation ($p < .0001$); Grammaticality x Lexicality ($p < .0001$)

Interaction with horizontal reduction



Mixed-effects logistic regression on all items: Grammaticality ($p < .0001$); Lexicality ($p < .05$); Grammaticality x Lexicality ($p < .05$); Grammaticality x Reduction ($p < .05$)

Overview

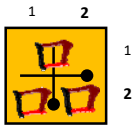
- Character "prosody"
- Radical position and reduction
- **Reduplication**
- Stroke shape and position
  - Lengthened strokes
  - Curved strokes
  - Hooked strokes

Reduplication patterns

- Binary horizontal reduplication
  林 比 兢 朋 弱 嚇 雙 遝 寐 替 質 瑩
- Binary vertical reduplication
  昌 吕 圭 炙 多 炎 哥 衆 僵 漆
- Triangular reduplication (binary both ways)
  品 淼 鑫 蠡 晶 猋 森 犇 众 磊 鑫 犇
- Square reduplication is quite rare: 毉
  - Not productive as a full template (splittable): 奐 爾 齒 繼 器
- Exceptions are quite rare and restricted:
  Non-binarity (mostly single strokes): 三 州 巡 黑 靈
  Inverted triangles (not a single template): 燚
  Splitting of horizontal reduplication (we'll come back to this): 變

Deriving reduplication

- Productive reduplication requires abstract rules or templates (Berent et al., 2014)
  - So 林 is 木 + [XX], not 木+木
  - Abstract reduplication in the PRC: 雙:双 犇:叒 鑫:垚
- Regular reduction still applies:
  根 林 烧 蛙 玨
- Triangles are derived from binarity + symmetry:



Testing reduplication productivity

- Myers, J. (2016). Knowing Chinese character grammar. *Cognition, 147*, 127-132.

Sample test items.

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

Grammar independent of lexicon

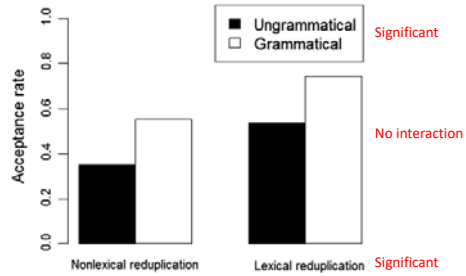
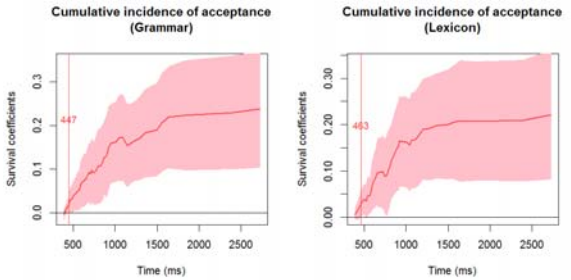


Fig. 1. Acceptance rates for fake characters containing lexical/nonlexical and grammatical/ungrammatical reduplicative configurations.

Grammar & lexicon at same time



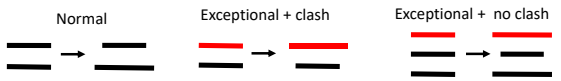
Survival analysis (Baayen & Blanche, 2017): Plots the evolving probability of acceptance vs. rejection (vertical bar marks when probability rises above chance)

Overview

- Character "prosody"
- Radical position and reduction
- Reduplication
- **Stroke shape and position**
  - Lengthened strokes
  - Curved strokes
  - Hooked strokes

Enlargement in strong positions

- Lowest component is enlarged, even if it's not free
  昌 吕 圭 炙 炎　　串 官　　二 夫 土 未
- Similarly for rightmost component
  林 川 州 井
- Exceptional top prominence triggers adjacent clash
  Clash: 士 末 畢　　No clash: 聿 重 垂 事



The variety of vertical strokes

- Vertical stroke shape (plain, curved, hooked) is never (?) lexically distinctive by itself
- Yet the contrast is apparently important enough to appear in many fonts:

十介丁　　十介丁　　十介丁

Distribution of plain strokes

- Left: 同甬行高商岡門兩竹
- Middle: 十中木東本不川州爪丫羊半下兩
  忄(忙)亻(仁)牜(牠)
  串甲申車莘幸聿年韋市午平丰
- Right: 非拜介并川州州爿

Distribution of curved strokes

- Left: 非拜介并川州州爪瓜并升飛片爿辨
  月丹舟周角用虎戶底大右力及看爿片
- Middle: …　　(does 班 count?)
- Right: …

Distribution of hooked strokes

- Left: …
- Middle: 丁于予乍乎于小水手事爭扌(把)木
- Right: 寸才可竹刂(到)
  (冂)月周同角高商兩門句弓

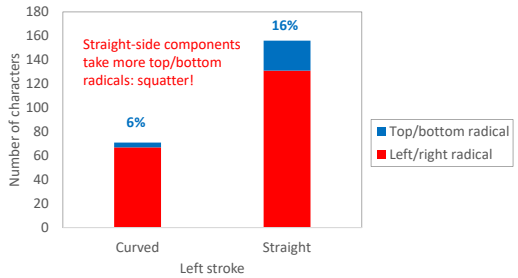
Some surface generalizations

- Plain stroke is default
  - Also allows bottom contact: 口廿上並
- Curved stroke appears on left,
  especially in "tall/thin" components (Wang, 1983): 周 vs. 同
- Hooked stroke appears on right with material above it:
  - Examples conforming to this pattern: 可周同
  - Hook on right but no material above it: 寸才扌(<手)刂(<刀)
  - Hook with material above it but not on right: 丁爭
  - Hook without material above and not on right: 手 小 水事

Curves: A deeper generalization

- Curves appear in "weak" position at left:
  - Curved stroke is also shorter than rightmost: 川
- "Left" is defined within a "prosodic" domain:
  - Separate components vs. wholes: 門 周
  - Thin/tall = one prosodic domain: 月 用 周 角
  - Fat/squat = two prosodic domains: 同 甬 高 商
    e.g. : fat/squat free 肉 vs. its tall/thin radical allomorph: 月
- A prediction (thanks to Lin Yu-Hsuan):
  - Thin/tall components favor left/right radicals more than fat/squat components, which also like top/bottom radicals
- A near minimal pair consistent with this prediction:
  角: 唷 堝 挷 桷 碏 斛 (0% top- or bottom-side radicals)
  甬: 俑 埇 蛹 踊 踴 悀 搈 楦 涌 勇 悬 筩 (25% top/bottom)

Testing curving and radical position

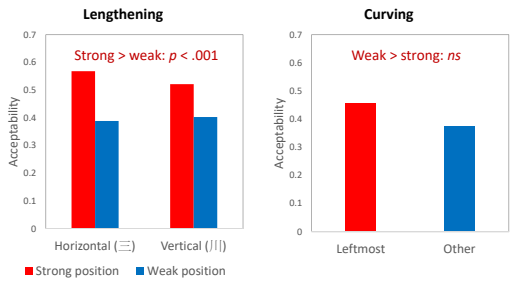


Fisher's test, $p < .05$, testing 丹月用有舟角周風 vs. 巾內冉冊冏市同向衆肉冏甬兩岡高商喬 as non-radicals in 227 characters with only one level of embedding (mixed-effects logistic regression confirms that this pattern not due to base component type frequency)

Testing the acceptability of lengthening and curving

- Another Worldlikeness experiment
  (with Chen Tsung-Ying)
- Testing 380 combinations of stroke size, shape, position, and orientation:「有中文字的感覺嗎？」

Lengthening & curving results



(Mixed-effects linear regression on arcsine-transformed Likert-scale judgments)

Hooking is more lexicalized

- No simple generalization applies to all cases
  - Near minimal pairs: 聿事 平乎
- Not predictable from small seal script (小篆):

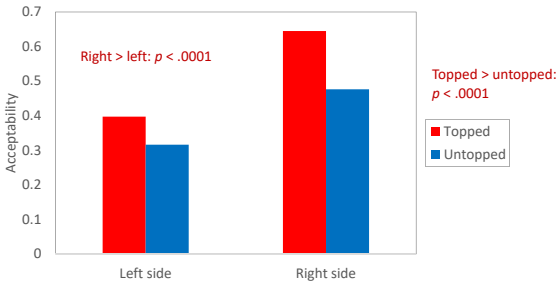
Hooks: Deeper generalizations?

- Hooks tend to be in "strong" positions:
  - At right (月寸) or alone in the middle (丁)
- Hooks also tend to have material above:
  - Usually contact at top: 可周同丁爭手乎 (cf. 寸才小水事)
- Stroke coordination as gesture coordination?
  - Syllables are the domain of gesture coordination
  - Complex rimes are heavier (more gestures to coordinate)
  - Complex onsets may also be heavier (see Hsieh, yesterday)

Testing the acceptability of hooking

- Yet another Worldlikeness experiment
  (again with Chen Tsung-Ying)
- Testing 24 combinations of shape and position:

Hooking judgment results



(Mixed-effects linear regression on arcsine-transformed Likert-scale judgments)

More on lexicalization

- Left component in horizontal reduplication shows regular reduction or curving (single "foot"): 玨 林 艸 辫
- But a lexical subset are still splittable: 班 辦
- Reduplication also doesn't trigger suppletive reduction:
  從 鲞 惢 (*亻 *才 *忄)
  cf. 獄 (reanalysis as affixation) 腧 (different elements)
- In spoken phonology this would imply ordering/opacity:

[XX] →	Suppletion →	Redup →	Reduc/Curv →	Infixation	
[XX] →	[才巴] →	… →	… →	… →	[把]
[XX] →	… →	[木木] →	[林] →	… →	[林]
[XX] →	… →	[王王]$_{split}$ →	[玨]$_{split}$ →	[班]	
[XX] →	… →	[辛辛]$_{split}$ →	[辨]$_{split}$ →	[辦]	

Summary and conclusions

- The prosodic hypothesis accommodates many facts:
  Radicals, reduplication, strokes all show:
  Left is weaker than right, top is weaker than bottom,
  in binary groupings
- Why does it work?
  - Universal Grammar applies to any sufficiently large and complex communication system (e.g., also sign language)?
  - Extra-linguistic constraints on motor control, perception, memory, and cognition?

B 8

References (1/2)

Baayen, H., & Blanche, P. (2017). Dynamic survival analysis. Talk presented at the Workshop on New Methods in Statistics, Tübingen, Germany, January 19-20.

Berent, I., Dupuis, A., & Brentari, D. (2014). Phonological reduplication in sign language: Rules rule. *Frontiers in Psychology, 5*, Article 560, 1-15.

Chen, T.-Y., & Myers, J. (2017). Worldlikeness: A Web-based tool for typological psycholinguistic research. *University of Pennsylvania Working Papers in Linguistics, 23*(1), Article 4, 20-30.

Handel, Z. (2016). *Sinography: A cross-linguistic study of the borrowing and adaptation of the Chinese script*. University of Washington ms.

Myers, J. (1996, June). Prosodic structure in Chinese characters. Poster presented at the Fifth International Conference on Chinese Linguistics, National Tsing Hua University, Taiwan.

References (2/2)

Myers, J. (2016). Knowing Chinese character grammar. *Cognition, 147*, 127-132.

Rude, N. (1986). Graphemic classifiers in Egyptian hieroglyphics and Mesopotamian cuneiform. In C. Craig (Ed.) *Noun classes and categorization* (pp. 133-138). John Benjamins.

Sproat, R. (2000). *A computational theory of writing systems*. Cambridge, UK: Cambridge University Press.

Tsai, C.-H. (2006). Frequency and stroke counts of Chinese characters. http://technology.chtsai.org/charfreq/.

Wang, J. C.-S. (1983). *Toward a generative grammar of Chinese Character structure and stroke order*. University of Wisconsin-Madison, Ph.D. dissertation.

Yip, P. C. (2000). *The Chinese lexicon: A comprehensive survey*. Routledge.