

Productive knowledge of formal regularities in Chinese characters

James Myers
National Chung Cheng University

Tsung-Ying Chen
National Tsinghua University

ICPEAL17/CLDC9 – October 2018

Thanks!

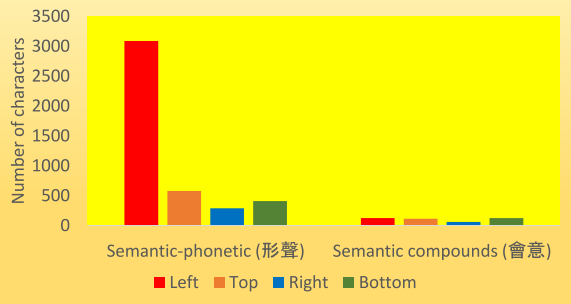
- Grants from Taiwan's National Science Council / Ministry of Science and Technology:
NSC 97-2410-H-194-067-MY3, NSC101-2410-H-194-115-MY3, MOST 103-2410-H-194-119-MY3
- Many lab assistants over the years:
Ko Yuguang, Yang Chentsung, Hsu Chiung-Wen, Hong Guo-Ming, Hsu Zi-Ping, Du Pei-Fen, Su Yu-Ting,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

Character function and form

- Chinese readers know which character constituents relate to meaning or pronunciation and where they appear (Feldman & Siok, 1997; Williams & Bever, 2010)
- 日 'day' → 時 ← 寺 sì
(semantic radical [部首]) shí 'time' (phonetic component)
- But do they also have productive knowledge of "meaningless" character patterns, a "formal grammar" of Chinese characters?
 - Yes (Wang, 1983; Myers, 2016; Myers, forthcoming)

Semantic radical position

- Semantic radicals favor the left (and top)



Characters in Sinica Corpus (<http://asbc.iis.sinica.edu.tw/>)

Radical reduction at left/top

- Left: Diagonalization of lowest horizontal stroke
土~地 金~鉛 立~站 舟~船 牛~物
且~助 工~功 重~動
- Left: Dotting of falling diagonal stroke
木~村 火~爛 米~精 耒~耗 采~釋
采~彩 禾~和
- Top: Shrinking of vertical strokes
雨~電 穴~空 竹~筆 西~覆
亦~奕 尚~當 高~膏

Stroke lengthening

- Lowest and rightmost stroke is long
二三工川州井
- Lexical exceptions, but no adjacent lengthening
土士王壬垂
- Lengthening is at edge of each constituent...
三刑
- ... and constituency is in the eye of the beholder
美 ≠ 羊 + 大
美 = 羊 + 大

Stroke curving

- Leftmost vertical stroke is curved
卅月用片片
- Constituency matters again (with usual ambiguity)
佛所門...拜
- Curving is less likely in wider constituents
肉~胖

□ vs. 冂	Wide	Narrow	Square
Curved		月用周有舟角	丹用
Straight	冊岡岡中內向兩肉市尚	再甬高商喬	同岡冉束

Stroke hooking

- Rightward hooks require crossed strokes to right
氏民長良良衣成...戌
- Leftward hooks favor asymmetry and top contact

	Asymmetrical	Symmetrical
Top contact	了可子手竹乎爭承予牙亨矛糸子孑孓	亦丁京于子
No top contact	事才水求寸事隸	小亦

Productivity in simplification

- Further radicals show regular shape changes
车~較 < 車~較
- Stroke lengthening is generalized as well
来 < 來 金 < 僉
- Curving does not generalize
非 < 非 貝 < 貝 門 < 門 风 < 風
- Rightward hooking doesn't require crossing
长 < 長 说 < 說

Testing productivity using...

Worldlikeness

A Web-based Tool for Typological Psycholinguistics

Experimenter / Participant / Researcher / About Worldlikeness

Last Update: 2018/10/5 (Update Logs)

Project funded by the MOST, Taiwan (103-2430-H-104-TD-MY2)

© Copyright 2014-2018, Language Processing Lab
Institute of Linguistics, National Chung Cheng University

Register [at] you.sdu.tw

English



<https://worldlikeness.org> (Chen & Myers, 2017)

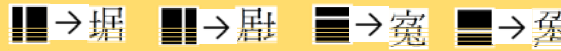
10

Experimentally testing radicals

- 98 readers gave binary Chinese-likeness judgments of 128 fake traditional characters (32 per Latin square group)
 - Grammaticality:** Is/isn't at left or top (favored positions)
 - Radical status:** Is/isn't a semantic radical (部首)

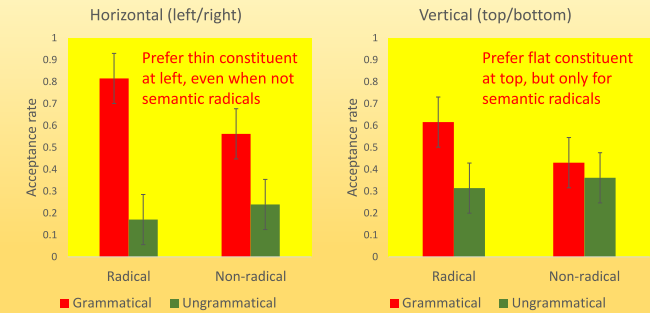
Axis	Grammatical		Ungrammatical	
	Radical	Non-radical	Radical	Non-radical
Horizontal	稜	蹀	𪛗	𪛘
Vertical	𪛙	𪛚	𪛛	𪛜

- 500 ms primes indicating relative constituent dimensions



11

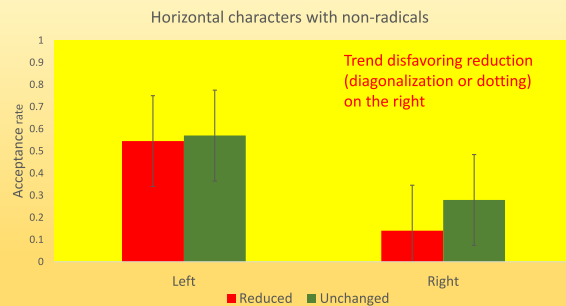
Effect of position on judgments



Mixed-effects logistic regression: Gram ($p < .0001$); Rad ($p > .1$); Gram x Axis ($p < .0001$); Gram x Rad ($p < .001$); Log positional frequency ($p = .002$) factored out

12

Effect of reduction on judgments



Mixed-effects logistic regression: Gram ($p < .0001$); Rad ($p > .1$); Reduct ($p = .09$); Gram x Reduct ($p > .1$); Log positional frequency ($p = .05$) factored out

13

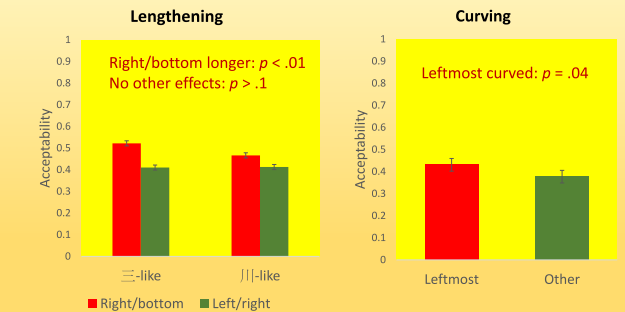
Lengthening & curving acceptability

- 80 readers saw 320 five-stroke non-lexical constituents (80 per Latin square group) similar to 卅卌丰手
- Chinese character "feeling" on a seven-point Likert scale

Main axis	Crosser edge	Crosser contact	Crosser hooking	Curving/length	Stroke position

14

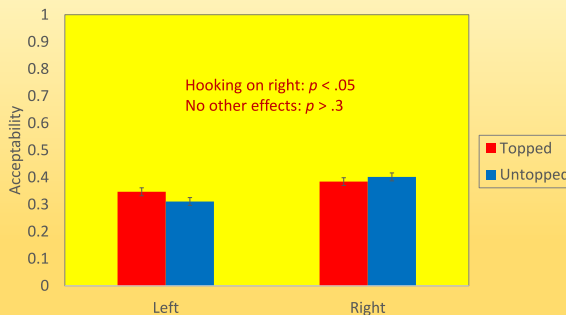
Lengthening & curving results



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

15

Hooking results



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

16

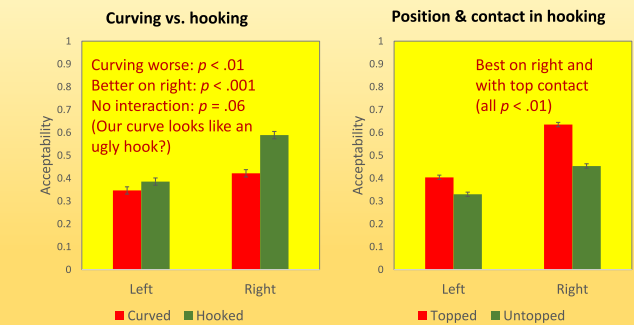
Retry with simpler materials

- 77 readers saw 24 three-stroke non-lexical constituents (80 per Latin square group) similar to 卅卌丰手
- Chinese character "feeling" on a seven-point Likert scale



17

New curving & hooking results



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

18

Conclusions

- Chinese characters do have a formal grammar
- The regularity of the formal patterns suggest productivity during character history
- Simplified characters maintain and extend many of these formal patterns
- Acceptability judgments on fake traditional characters show that readers “know” these patterns
- Further corpus-based and experimental data are needed in this underexplored area

19

Further reading

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.

Feldman & Siok (1997). The role of component function in visual recognition of Chinese characters. *JEP: LMC*, 23(3), 776-781.

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

Myers, J. (forthcoming). *The grammar of Chinese characters: Productive knowledge of formal patterns in an orthographic system*. Routledge.

Wang (1983). *Toward a generative grammar of Chinese Character structure and stroke order*. U. of Wisconsin diss.

Williams & Bever (2010). Chinese character decoding: A semantic bias? *Reading and Writing*, 23(5), 589-605.

20

Form drives stroke order



(gifs created by Wiktionary user Micheletb)

21

Historical innovations and reanalyses

- Small seal script (小篆書) favored the left and top for radicals, but still didn't show reduction reliably

木~觀 > 木~根 土~地 > 土~地 (艸~艸) > 竹~筆

- No bottom/right lengthening or left-only curving

二 > 二 土 > 土 士 > 士 月 > 丹 井 > 井

- Rightward hooking spread by analogy?

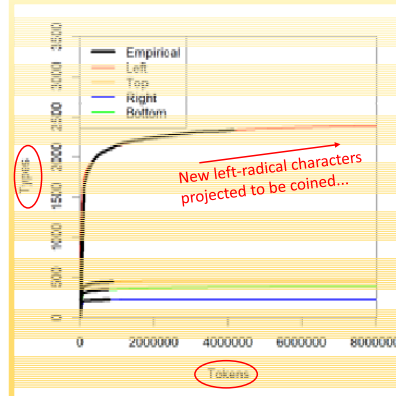
氏 > 氏 氏 > 民 長 > 長 長 > 長 衣 > 衣

- Leftward hooking has two historical sources

了 > 了 可 > 可 事 > 事 才 > 才 丁 > 丁 寸 > 寸

22

Productivity in character coinage



Left edge the most productive position for semantic radicals

Based on 4,671 semantic-phonetic characters (形聲) in Sinica Corpus (<http://asbc.iis.sinica.edu.tw/>)

Growth curves estimated using zipfR (<https://cran.r-project.org/web/packages/zipfR/index.html>)

23