Grammaticalization frozen in ink: Chinese character semantic radicals as proto-affixes

James Myers
National Chung Cheng University
Taiwan

Thanks!

2

Overview

- Orthographic grammar
- Radicals as affixes
  - Linguistic diagnostics
  - Corpus evidence
  - Experimental evidence
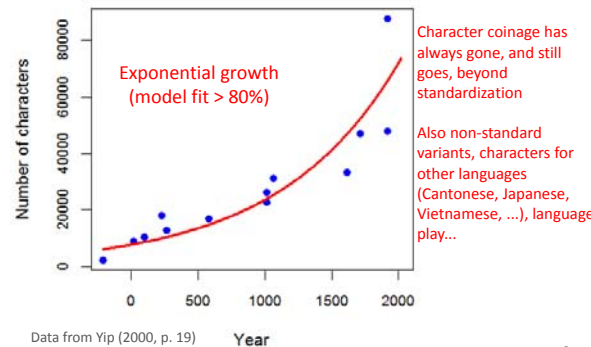- Grammaticalization
  - Radicals as proto-affixes

3

Homo grammaticus

- Humans have an innate drive for grammar
  - Productivity allows communication of new ideas
  - Predictability reduces burden on rote memory
- Grammar can arise in any modality
  - Spoken languages, sign languages
- Does orthography trigger the grammar drive?
  - Phonetic orthographies can piggyback on speech
  - But logographic systems need their own structure
  - Chinese is the sole survivor of this ancient type

4

Chinese characters are productive

Exponential growth (model fit > 80%)

Character coinage has always gone, and still goes, beyond standardization

Also non-standard variants, characters for other languages (Cantonese, Japanese, Vietnamese, …), language play…
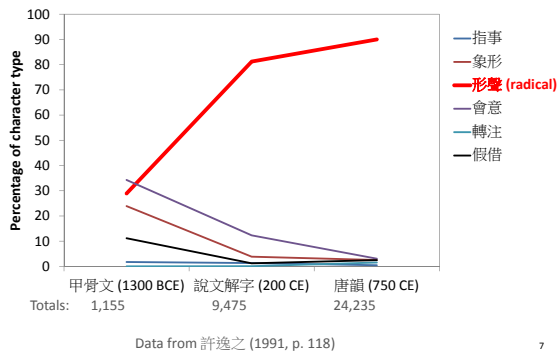
Data from Yip (2000, p. 19)

5

Formal character coinage operations

- Does this productivity depend on the same operations as in spoken/signed word coinage?
- Reduplication
  - E.g. 林森哥 (see Myers, 2016)
- Compounding
  - More common in other systems (as we'll see)
- Affixation
  - "Semantic radicals" (部首) **My focus**
  - A misnomer (Behr, 2017): They aren't "roots"

6

Radicals have been key for millennia

指事
象形
形聲 (radical)
會意
轉注
假借

甲骨文 (1300 BCE)　說文解字 (200 CE)　唐韻 (750 CE)
Totals:　1,155　　9,475　　24,235

Data from 許逸之 (1991, p. 118)

7

Some diagnostics of affixation

- Closed-class:
  farm**er** (*farm**ip**)　　桌子 人們 阿媽 门儿
- Bound
  farm**er** (*er)
- Semantically bleached:
  read**able** (cf. **able** to read)　院子 可能性
- Fixed locations (usually one word edge):
  farm**er** (***er**farm)　　**un**happy (*happy**un**)
- Phonologically reduced:
  read[ə]ble (cf. [ei]ble to read)　院子 (cf. 原子)

8

Radicals are closed-class, bound, and semantically bleached

- Inventory isn't fixed, but they're hard to coin
  - "Non-Chinese-ness" of other Sinographic systems
    - E.g. Vietnamese (Chữ Nôm) 吒 *ba* 'three' (Nguyen, 2015)
- There ain't no "free radicals"
  - 金 is "its own radical" as much as 一 is 丁's radical
  - Some radicals are never free: 宀 氵 癶 廴 …
  - Some radicals have bound forms: 忄 扌 氵 艹 …
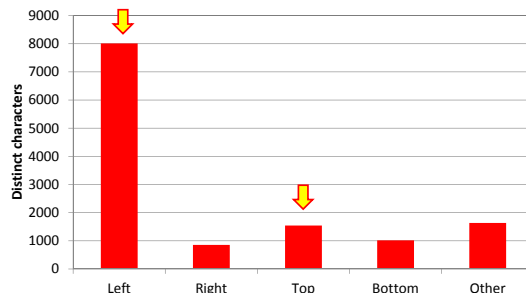- Radical meanings are vague or inconsistent
  - E.g. 媽 奸 好 姓　吃 叫 嗎 呢

9

Radicals have semi-fixed locations

- Radicals have their own favored positions
  - Most radicals (90%) have a dominant position (in 13,060 traditional characters in Tsai, 2006)
  - Most characters (87%) have a radical on one edge: Left (詞), right (鵝), top (花), bottom (盒)
- The system as a whole favors the left
  - (Cf. suffixation > prefixation in English & Chinese)
  - Most characters (61%) have radicals on the left
  - More radicals (33%) favor left than other positions
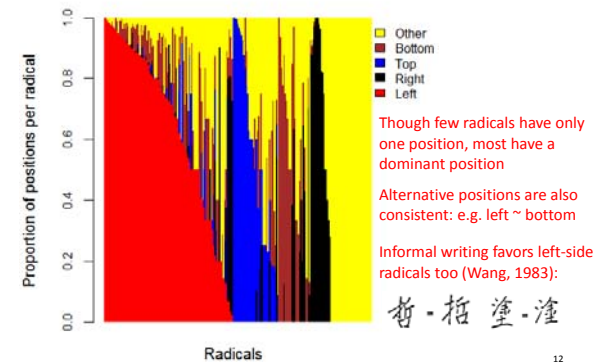 - The left is also the most productive in coinage

10

Where characters have their radicals

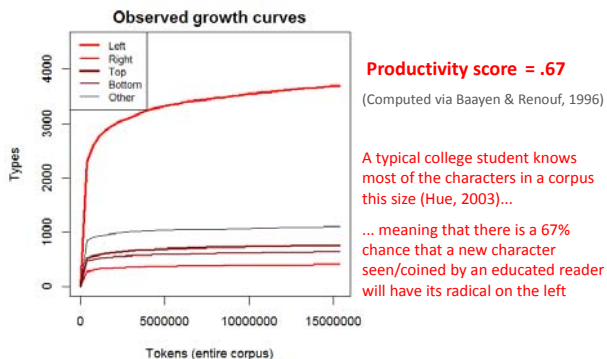Computed from 13,060 traditional characters in Tsai (2006)

11

Where radicals prefer to go

Though few radicals have only one position, most have a dominant position

Alternative positions are also consistent: e.g. left ~ bottom

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

哲‧栖 垔‧洼

12

Radical position productivity

Productivity score = .67
(Computed via Baayen & Renouf, 1996)

A typical college student knows most of the characters in a corpus this size (Hue, 2003)...

... meaning that there is a 67% chance that a new character seen/coined by an educated reader will have its radical on the left

Using Academia Sinica Balanced Corpus of Modern Chinese (Chen et al., 1996):
6,608 most common traditional characters

13

Radicals reduce in form

- Some radicals have suppletive forms at left/top

心~忙　人~位　　水~泊　　手~拾
艸~花　　竹~筆

- But not at right or bottom (generally)

忙~忘　泊~泉　拾~拿　獄　（分~刻）

- Other radicals reduce in a more regular way
  - Wang (1983): Diagonalization and stroke reduction

金：鉛~鑒　　土：地~型　　牛：物~牽
木：村~果　　米：精~粱　　衣：被~裝

14

Experiments on radical position

- Readers use character component position when recognizing characters (Taft et al., 1999)
  - Recognizing 另 is not affected by existence of 加
- Our questions:
  - Are readers aware that particular components tend to appear in particular positions?
 - Do they also know the system-wide preference for left-sided radicals?

15

Testing fake character acceptability

- Readers judge if they are/aren't like characters
- Lexicality: Component is/isn't a "real" radical
- Grammar: Do/don't appear in dominant position
  - "Radical" appears at left/top (vs. right/bottom)

	Lex		Nonlex	
	Gram	**Ungram**	**Gram**	**Ungram**
Left/Right	稞	顢	鰈	雗
Top/Bottom	圂	圗	齒	圇

(Myers, 2011, 2012)
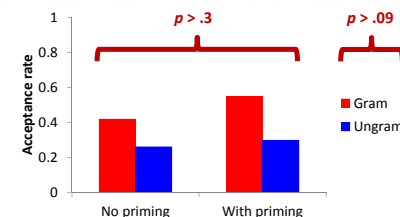
16

Readers favor left-sided "radicals"
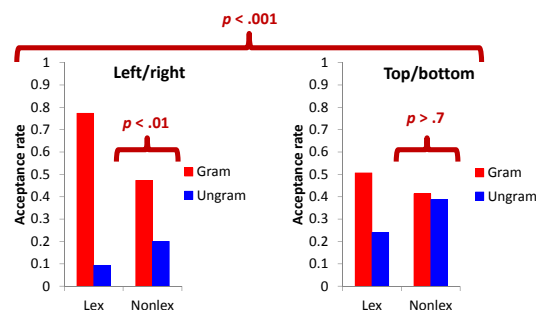
(From a mixed-effects logistic regression model by subject and item)

17

Does position itself have an effect?

- A ongoing experiment (with Chen Tsung-Ying) tests "radical" position, regardless of reduction:

▮▮ → 眠　▮▮ → 䀏　▮▮ → 䡇　▮▮ → 䢃

(From a mixed-effects logistic regression model on horizontal, nonlexical, nonreduced items)

18

Affixes, or just proto-affixes?

- Affix diagnostics don't work perfectly here:
  - Closed-class, bound: Mostly true…?
  - Semantic bleaching: True, but not very consistent
  - Fixed locations: Also more variable than affixes
  - Reduction: Inconsistent: 火 vs. 灬 in 燙 vs. 熱
- Grammaticalization frozen in ink?
  - The journey from free word to affix is gradual
    - Quasi-affixes in English and Chinese: **pro-** -好
  - Semantic bleaching before reduction (Booij, 2005)
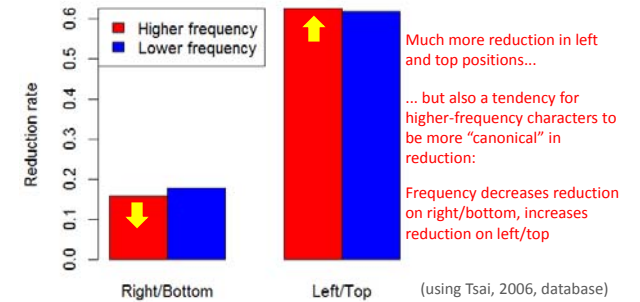 - Frequency speeds grammaticalization (Bybee, 2011)

Other steps along the continuum

- Vietnamese (Chữ Nôm)
  - Its radicals are more root-like than affix-like
    - Often "synonymic", not "taxonymic" (Handel, 2016)
    - Also often lack formal reduction: 秚 *tay* 'arm'
    - Cf. early logographic systems (Rude, 1986)
- Standard simplified characters of the PRC
  - Left-side reduction applies to more radicals:
    词~警 铅~鉴 红~累   (cf. 詞~警 鉛~鑒 紅~累)
  - But also degrammaticalization…?
    广 is no longer fully bound (cf. traditional 廣)

Frequency affects grammaticalization

Much more reduction in left and top positions…

… but also a tendency for higher-frequency characters to be more "canonical" in reduction:

Frequency decreases reduction on right/bottom, increases reduction on left/top

(using Tsai, 2006, database)

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

Summary and conclusions

- Chinese character semantic radicals are:
  - Closed-class, bound, bleached, consistent in position (mostly on one edge), regularly reduced
  - May be even more affix-like if not standardized
- So what?
  - Written Chinese (affixing) is typologically quite different from spoken Chinese (compounding)
  - Chinese characters seem to invoke the same human capacity for grammar-driven word formation as in speech and sign

References (1/3)

Baayen, R.H., & Renouf, A. (1996). Chronicling the *Times*: Productive lexical innovations in an English newspaper. *Language, 72*, 69-96.

Behr, W. (2017, April). Radical misconceptions: On the background of European ideas about *bùshǒu* 部首. *ECLL* Chinese Linguistics Day, University of Zürich. <http://www.brill.com/products/reference-work/encyclopedia-chinese-language-and-linguistics-5-volumes>

Booij, G. (2005). Compounding and derivation: Evidence for Construction Morphology. In W. U. Dressler (Ed.) *Morphology and its demarcations* (pp. 109-132). John Benjamins.

Bybee, J. (2011). Usage-based theory and grammaticalization. In B. Heine & H. Narrog (Eds.), *The Oxford handbook of grammaticalization* (pp. 69-78). Oxford: Oxford University Press.

Chen, K.-J., Huang, C.-R., Chang, L.-P. & Hsu, H.-L. (1996). Sinica Corpus: Design methodology for balanced corpora. *PACLIC, 11*, 167-176.

References (2/3)

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

Hue, C.-W. (2003). Number of characters a college student knows. *Journal of Chinese Linguistics, 31*(2), 300-339.

Myers, J. (2011, September). The psychological reality of formal regularities in Chinese characters. Presented at the 7th Conference of the European Association of Chinese Linguistics, Venice, Italy.

Myers, J. (2012, August). Levels of analysis in the generalization of Chinese character regularities. Presented at CogSci 2012: The 34th Annual Conference of the Cognitive Science Society, Sapporo, Japan.

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

Nguyen, T. (2015). http://www.chunom.org.

References (3/3)

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

Taft, M., Zhu, X., & Peng, D. (1999). Positional specificity of radicals in Chinese character recognition. *Journal of Memory and Language, 40*(4), 498-519.

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 Ccharacter structure and stroke order*. University of Wisconsin-Madison, Ph.D. dissertation.

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

許逸之 (1991).《中國文字結構說彙》.臺灣商務印書館。