Two levels of regularity in writing systems

Evidence from frequency effects

Congrès de l'ACL 2026 | 2026 CLA Conference

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

Overview

- **Morphology** shows **negative** frequency effects, as memory lapses regularize **rarer** words (Pinker 1999)
  - ↘ Regular past tense: ***flow ~ flowed*** vs. *grow ~ grew*
- **Phonology** shows **positive** frequency effects, as more **common** words regularize through use (Bybee 2001)
  - ↗ Trochaic foot: *hómily* vs. ***fám'ly***
- **Writing systems** show the same contrast, suggesting written analogues to morphology and phonology

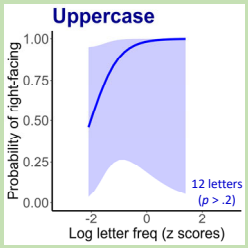
Two levels in writing systems (Gelb 1963, Meletis & Dürscheid 2022)

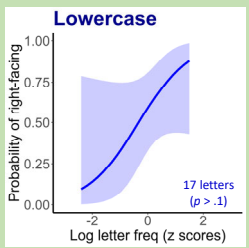
- **Inner structure** (graphematics) describes the mapping between written unit combinations and spoken language
  morphemes (logographies), syllables (syllabaries), subsyllabic groups (abugidas), phonemes (alphabets), ...
- **Outer form** (graphetics) describes the shapes of the written units themselves

English (alphabet) (theory: Watt 1975; regularity: Seidenberg & McClelland 1989; corpora: Brysbaert & New 2009, Carnegie Mellon University 2015; Constant 2012)

- **Letter combinations**
  Regular pronunciations *decrease* as frequency increases
  <cove> /ow/ vs. <love> /ʌ/
  <braid> /ej/ vs. <said> /ɛ/

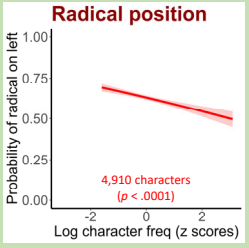
- **Letter shapes**
  - Asymmetrical letters tend to **face right** <B, b>
  - This regularity *increases* as freq. increases
    <J, j> vs. **<B, b>**



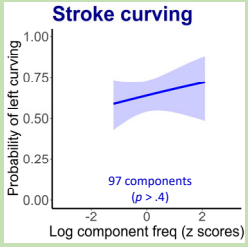


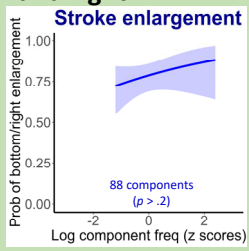
Chinese (logography) (theory: Myers 2019; corpora: Chuang & Teng 2009, Tsai, 2006)

- **Component combinations**
  - Most characters are **phono-semantic**, with the **semantic radical on the left:** 媽 *mā* 'mother' = 女 'woman' + 馬 *mǎ*
  - Both regularities *decrease* as freq. increases
    姝 *shū* 'pretty' vs.
    好 *hǎo* 'good'*
    娟 *juān* (name) vs.
    婆 *pó* 'old woman'
    *(女 'woman' + 子 'child')



- **Component shapes**
  - Character components favour **stroke curving on the left** and **stroke enlargement on the bottom and right**
  - Both regularities *increase* as frequency increases
    丙 *bǐng* 'third' vs.
    月 *yuè* 'month'
    士 *shì* 'scholar' vs.
    土 *tǔ* 'earth'

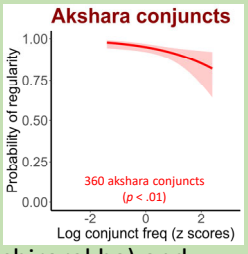




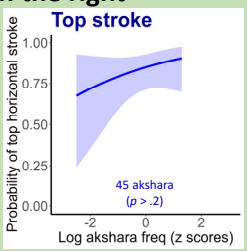
Thanks! • National Science and Technology Council (Taiwan) grant 112-2410-H-194-030-MY3 • Yangru Chen • Amalia Gnanadesikan • Paul Saka • Lian Hee Wee • YOU!

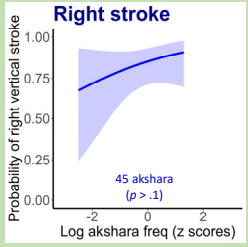
Sanskrit Devanagari (abugida) (theory: Gnanadesikan 2022; corpora: *Sanskrit frequency list* n.d., Stiehl 2007)

- **Akshara combinations**
  Consonant conjuncts *decrease* in regularity as frequency increases
  त्म *tma* = त *ta* + म *ma* vs.
  त्र *tra* = त *ta* + र *ra*



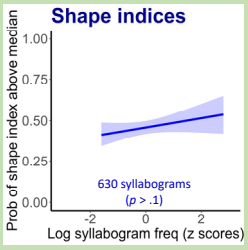
- **Akshara shapes**
  - Akshara favour **top horizontal strokes** (shirorekha) and **vertical strokes on the right**
  - Both regularities *increase* as frequency increases
    थ *tha* vs. ऊ *ū*
    ठ *ṭha* vs. त *ta*





Yi (syllabary) (corpora: Walters et al. 2015, *Yi script index book* 1984)

- **Syllabogram combinations** are fully regular
- **Syllabogram shapes** are grouped by conventional indices (e.g., index ꑍ for ꑍ *njyt* and ꑎ *jip*)
  - Index group size *increases* as token frequency increases
    ꒉ *siep* [ꒉ: index size 6] vs.
    ꓑ ***la*** **[ꓑ: index size 70]**



References • **Brysbaert & New (2009)** Moving beyond Kučera and Francis. *Behavior Research Methods*. • **Bybee (2002)** *Phonology and language use*. Cambridge. • **Carnegie Mellon University (2015)** The CMU pronouncing dictionary. github.com/cmusphinx/cmudict. • **Chuang & Teng (2009)** 漢字構形資料庫的研發與應用. Academia Sinica. • **Constant (2012)** The CMU pronouncing dictionary in IPA. people.umass.edu/nconstan/CMU-IPA/ • **Gelb (1963)** *A study of writing*. Chicago. • **Gnanadesikan (2022)** Amodal morphology. In Haralambous (Ed.), *Grapholinguistics in the 21st Century*. Fluxus. • **Meletis & Dürscheid (2022)** *Writing systems and their use*. De Gruyter Brill. • **Myers (2019)** *The grammar of Chinese characters*. Routledge. • **Pinker (1999)** *Words and rules*. HarperCollins. • ***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*. • **Stiehl (2008)** *Sanskrit-Kompendium*. Economica Verlag. • **Tsai (2006)** Frequency and stroke counts of Chinese characters. technology.chtsai.org/charfreq/. • **Walters et al. (2015)** *A preliminary study of Nuosu Yi syllable frequency in text*. SIL International ms. www.sil.org/resources/archives/61553. • **Watt (1975)** What is the proper characterization of the alphabet? *Visible Language*. • ***Yi script index book*** **(1984)** 彝文检字本. Sichuan Nationalities Publishing House.