An online exploration of cross-linguistic variation in wordlikeness judgments

James Myers
National Chung Cheng University

Tsung-Ying Chen
National Tsinghua University

Funded by the Ministry of Science and Technology, Taiwan
(grant 106-2410-H-194-055-MY3)

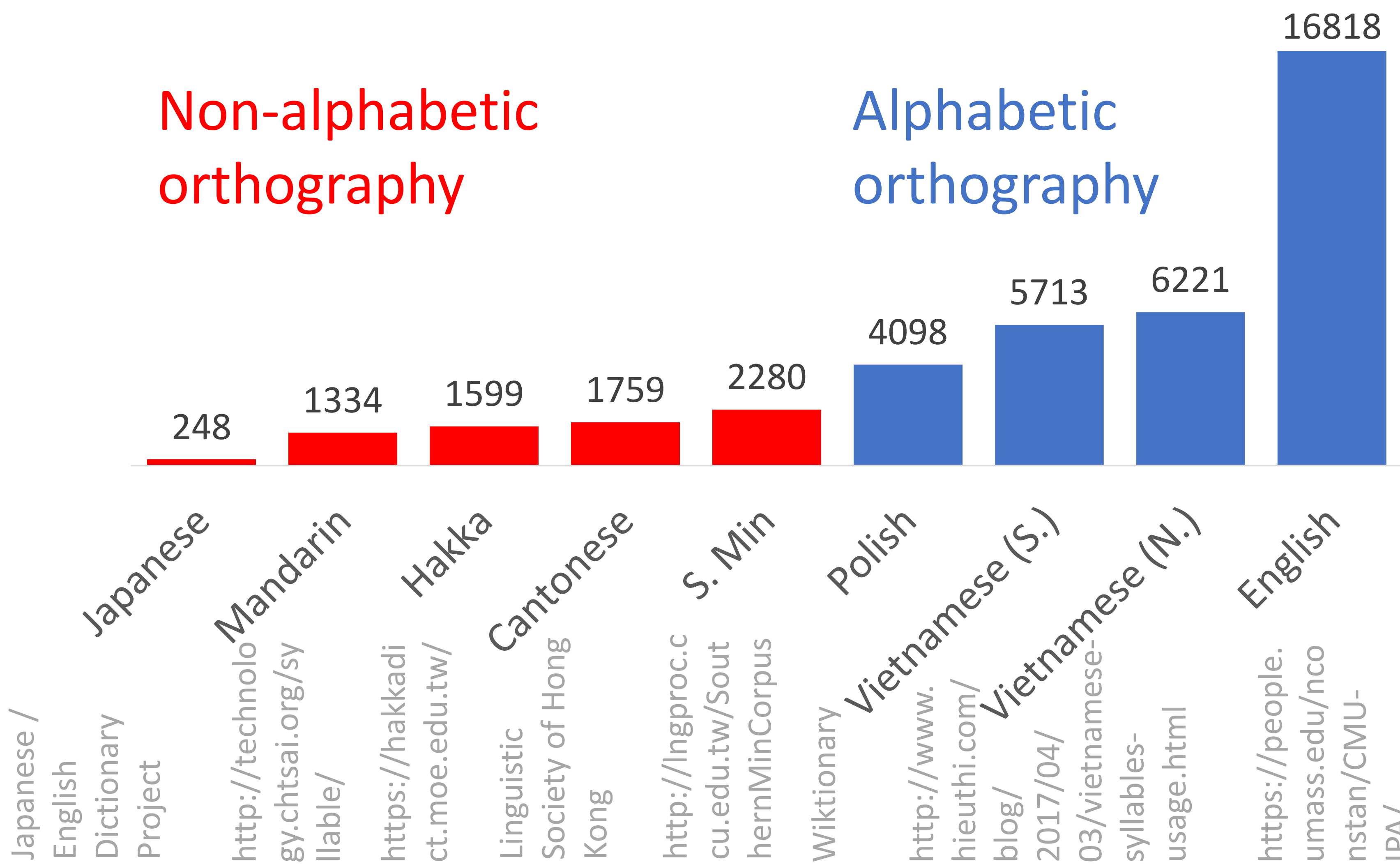
Thanks

- Sinitic languages:
  - 蘇思綺、蔡溥泰、陳宗穎、林冠佑、劉美君、張佑竹, Chi Shing Tse
- Japanese:
  - 岩野萬里子, Shigeto Kawahara, Kazuhiko Fukushima
- Polish:
  - Aleksandra Smolka, Szymon Grzelak, Norbert Kordek
- Vietnamese:
  - Võ Hồ Long An, Đào Ngọc Sơn, James Kirby (including https://github.com/kirbyj/vPhon)
- English:
  - Mike Hammond, Benjamin Tucker

Goals

- Phonemes seem to be more active in English than in Mandarin [O'Seaghdha *et al.* (2010). *Cognition*]
- Due to differences in **syllabary size**?
  - Syllables are perceptually more salient than phonemes
  - But accessing the lexicon solely via syllables becomes more difficult the more syllables need to be memorized
- Ideal test: **meta-megastudy** [Myers (2016). *Mental Lexicon*]
  - Language-level variables also included in regression
- This study
  - "Convenience" sample of nine languages or dialects
  - Online testing needed for English, Japanese, Polish

Syllabary size (including tone)



Methods

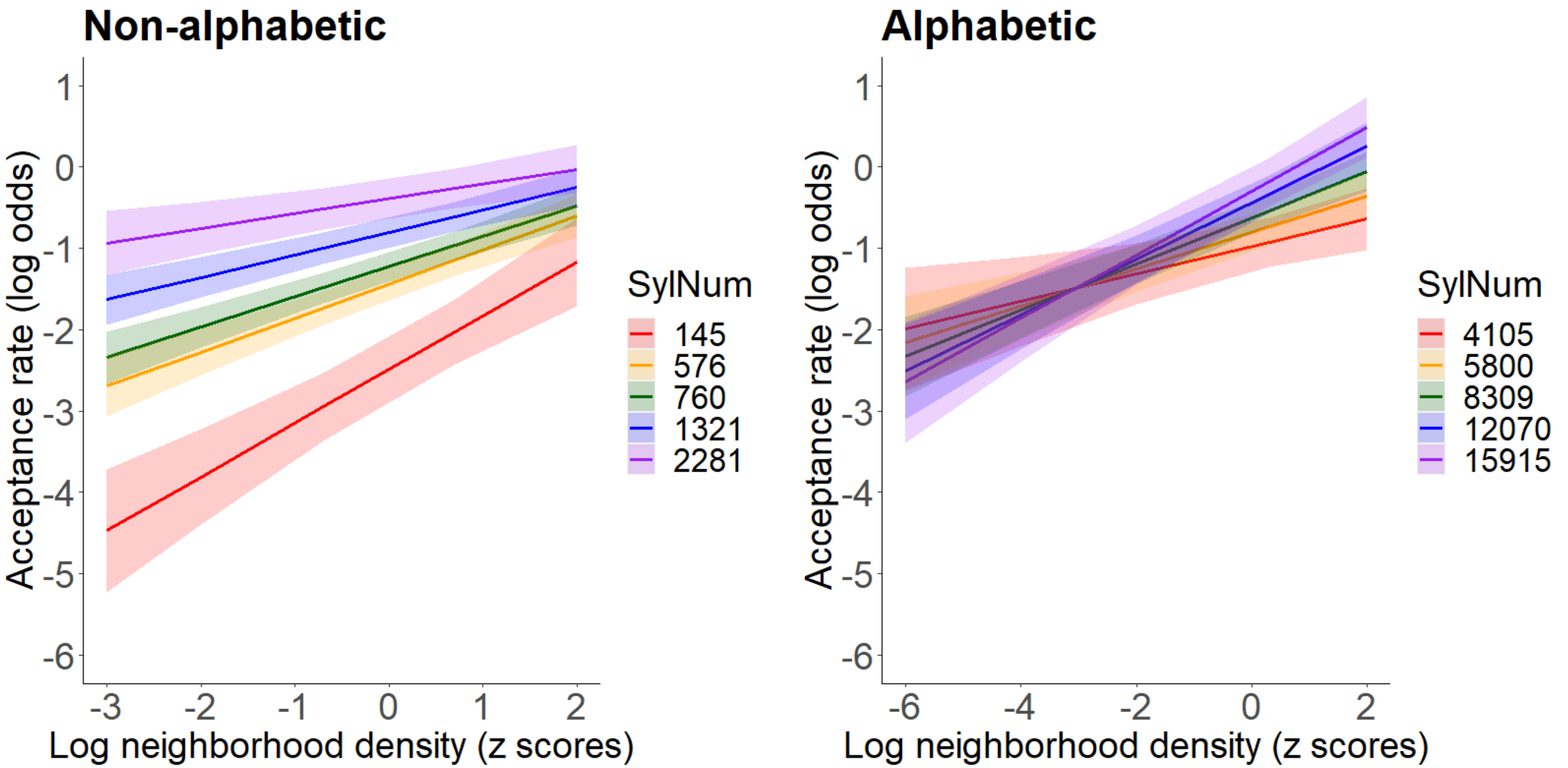
- **Wordlikeness**
  - Most direct test of productive linguistic knowledge
  - Binary responses: like vs. unlike a word in the language
- Stimuli
  - ≈ 200 (range: 193-214) **prosodically minimal nonwords**
    - Japanese: CV(X).Ca; others: CV(X)(T) (X = C or glide, T = tone)
  - Randomly generated from language's segments & tones
  - Spoken stimuli produced by native speakers
  - Presented in different random order to each listener
- Participants
  - ≈ 31 per language (range: 22-64)

Analysis

- Item- and trial-level variables
  - **Neighborhood density**
    - Stronger effects may imply more “holistic” syllable processing
  - **Phonotactic probability** (including tone)
    - Stronger effects may imply more phoneme-level processing
    - But no effects were found at all
  - **Cross-trial onset priming**
    - Primed = target shares first segment with preceding item
    - Stronger effects may imply more phoneme-level processing
- Language-level variables
  - **Syllabary size**, nested within **orthography** (alpha vs. not)
- Mixed-effects logistic regression

  Resp ~ Alpha / (log_nSyl.z * (Priming + PP.z + logND.z)) + (Priming + PP.z + logND.z|Subj) + (Priming|Item)

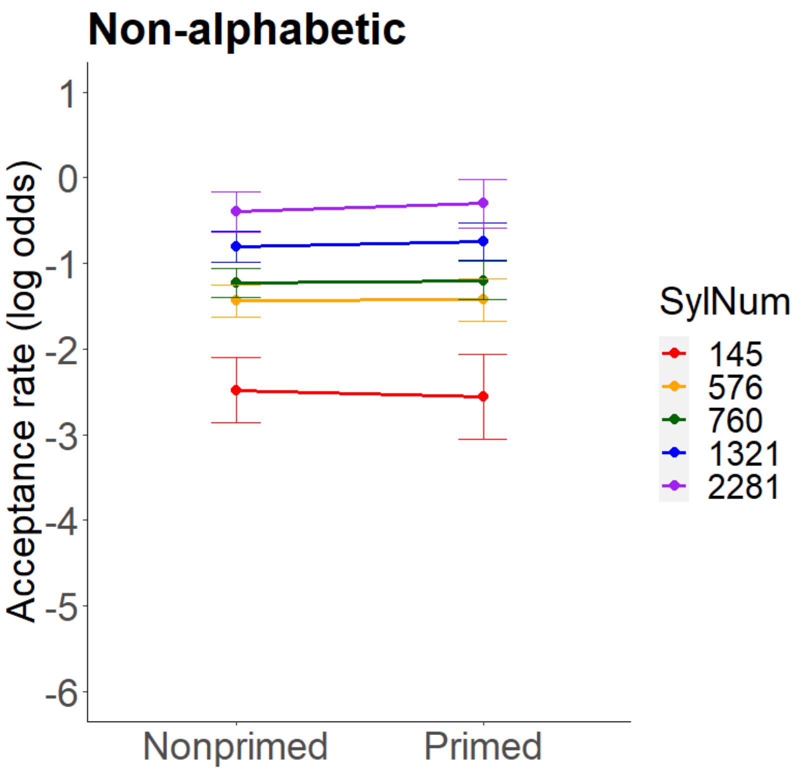
Neighborhood density

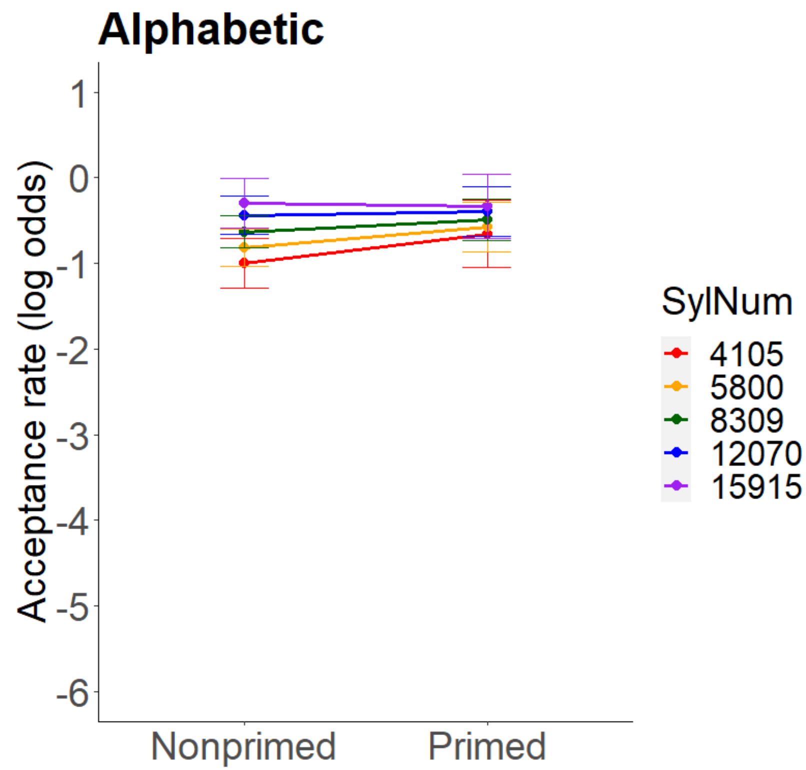


Bigger syllabary = weaker (as expected)

Bigger syllabary = stronger (orthographic neighbors?)

Cross-trial onset priming





No priming (as expected)
Also, bigger syllabary = more tolerance of novelty (as expected)

Priming is present (as expected), but bigger syllabary = weaker (orthographic inhibition?)

Take-aways

- Syllabary size helps explain cross-language variation in syllable/phoneme influences on wordlikeness
- Meta-megastudies
  - Simple regression-based designs can tell us a lot
  - Cross-trial priming can be tested even without overtly designing a priming experiment
  - Other easily conducted analyses include cross-trial perseveration (responding the same as in previous trial)
- **Worldlikeness** app [Chen & Myers (2021). *Linguistics Vanguard*]
  - Design options intentionally limited to encourage cross-study consistency, allowing meta-megastudy-ready databases to emerge even without a coordinated team