Word size in spoken and written Mandarin Chinese

James Myers (National Chung Cheng University, Taiwan)
49th Annual Meeting of the Societas Linguistica Europaea, 31 August – 3 September 2016, Naples, Italy

Summary
• Mandarin words tend to be disyllabic
• Do disyllables play a special role in spontaneous speech?
• To find out, we analyzed word sizes in spoken and written corpora

Productivity
• Disyllabic/two-character words predominate in both speech and writing

Productivity was quantified as growth curves (Evert & Baroni 2007)
  - The number of distinct word types as a function of the number of sampled word tokens
  - Word lengths compared by extrapolating to the same sample size via LNRE modeling
  - The growth curves showed a dramatic effect of modality (see plots)

Primming
• Thousands of Mandarin lemmas (syntactic/semantic lexical entries) are elastic
  - “Freely” vary in word length as monosyllabic or disyllabic (Duanmu & Dong 2016)

- Optional empty suffix: zhuō ‘table’
- Optional reduplication: dì ‘younger brother’
- Redundant modifier: gē ‘elder brother’
- Redundant head: dōng ‘east’
- Superordinate category: guā ‘melon’

An experimental priming effect (Perry & Zhuang 2005)
  - In a picture-naming task, speakers were more likely to choose the disyllabic variant of elastic words when there also were pictures with fixed disyllabic names in the test set
  - A corpus-based analysis
    - Predict disyllabic variant of elastic lemmas from log ratio of disyllabic to monosyllabic in adjacent ten words, with or without repetition of the target lemma
    - Elastic word size was primed by context, both preceding (see plots) or following (same pattern)
  - Prosodic effects in writing? Or an indirect effect of shifting degrees of formality?

What next?
• Productivity
  - Are affixed and compound words different?
• Priming
  - What about trisyllabic or longer words?
  - Is there cross-speaker priming?
• Other languages
  - Are there similar patterns in other languages with disyllabic feet?
  - What happens in languages with bimoraic or unbounded feet?

Effect of prosody and lemma repetition on probability of producing disyllabic elastic variant
(mixed-effects logistic regression on 146 spoken and 990 written elastic nouns)

References


Worldlikeness
http://lingproc-4083.nitrouspro.com:3000/