The relationship between lexical frequency, compositionality, and phonological reduction in English compounds

Forrest Davis and Abigail C Cohn
Cornell University
94th Annual Meeting of the Linguistic Society of America, January 2020

Research Questions

- Are more opaque compounds (cupboard) phonologically different from more transparent compounds (blueberry)?
- Are effects of compositionality distinct from effects of lexical frequency and degree of conventionalization?

Introduction

- Compositionality Degree that the meaning of a compound is the sum of its parts (e.g. humbug vs. blueberry)
- Is this relationship between phonological reduction and compositionality (Libben and Jarema, 2006) robust? Is it distinct from those of lexical frequency (Jurafsky et al., 2001; Bell et al., 2009)?

Data

- Buckeye Corpus (Pitt et al., 2007)
- Used 21 most frequent bisyllabic nominal compounds orthographically represented with no space (e.g. roommate, airline, freshman, football)

Hypothesis

- More opaque compounds are more phonologically reduced than transparent ones

Measure of Compositionality

- Goal to establish a gradient measure of compositionality (cf. Libben and Jarema, 2006)
- Survey of 24 native American English speakers using a 7 point Likert Scale

Measure of Lexical Frequency

- Frequency for compound and its constituents (e.g. homework: 6069, home: 196061, man: 216061)
- Counts with add-one smoothing Corpus of Contemporary American English (COCA) (Davies, 2008)
- Pointwise mutual information (PMI) calculated (e.g. homework: 4.16, freshman: 15.95): correlated with conventionalization (Evert, 2008; Ramsch et al., 2010)

PMI(xy) = \log_{2} \frac{p(x,y)}{p(x)p(y)}

Main Results

- Rating and PMI (degree of conventionalization) are significant predictors of final rime duration
- The less compositional a compound the shorter its final rime

Reduction Results

- Mean duration of the final rime shorter when ratings are low (opaque) than expected given rime duration in non-compounds (e.g. ware in software is half the duration of where)
- Rating and PMI statistically significant predictors (p < 0.005)
- Modifier frequency and duration in monosyllabic compounds also statistically significant (p < 0.05)

Selected References

- Alan Bell, Jason M Brenier, Michelle Crayen, Cynthia Ernest, and Dee Jardine. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60.

Contact Information

- http://conf.ling.cornell.edu/forestdavis/
- fd252@cornell.edu

Acknowledgements

Special thanks to friends and colleagues from Cornell Linguistics and the Cornell Phonetics Lab