**Effects of lexical frequency and compositionality on phonological reduction in English compounds**

Forrest Davis and Abigail C Cohn
Cornell University

25th Architectures and Mechanisms of Language Processing Conference, September 2019

---

**Research Questions**

- Are more opaque compounds (e.g., *cupboard*) phonologically different from more transparent compounds (e.g., *blueberry*)?
- Are effects of compositionality distinct from those of lexical frequency and degree of conventionalization?

---

**Introduction**

- **Compositional** Degree that the meaning of a compound is the sum of its parts (e.g., *humbag* vs. *blueberry*).
- **Opaque** less compositional
- **Transparent** more compositional
- **Reduced** neither opaque nor transparent
- **Not Reduced** very opaque or very transparent

- Is this relationship between phonological reduction and compositional property (Libben and Jarema, 2006) robust? Is it distinct from those of lexical frequency (Jurafsky et al., 2001; Boll et al., 2009)?

---

**Data**

- Buckeye Corpus (Pitt et al., 2007)
- Conversational speech from 40 American English speakers
- Each word labeled with both a phonemic (citation form) and phonetic transcription
- 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.

---

**Main Results**

- **Rating Results**
  - Ratings of compositional transparency are distinct from lexical frequency:
  - 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.
  - Stepwise linear regression with independent variables lemma, rime, rating, frequencies, PMI, and duration of the same rime in monosyllabic non-compounds.
  - Rating and PMI statistically significant predictors (p < 0.005).

**Selected References**


---

**Discussion**

- Results provide evidence of semantic opacity and reflexes in phonological form.
- Cast doubts on categorical notions of compositionality assumed in theoretical aspects of composition representation.

---

**Future Directions**

- Collect additional compounds from other corpora like BNC and Boston Radio Corpus.
- Broaden the number of compounds rated for compositionality.

---

**Acknowledgements**

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

---

**Contact Information**

- [http://conf.ling.cornell.edu/forrestdavis/](http://conf.ling.cornell.edu/forrestdavis/)
- fd252@cornell.edu