# Semantic predictability and word sense disambiguation in spontaneous speech

Michael McAuliffe

April 15, 2012

# Brief recap

- Words can vary in their production according to how predictable they are from context
  - The cow gave birth to the calf
  - She is glad Jane called about the calf
  - I have an ache in my calf
- Words in predictable contexts are reliably shorter (Clopper & Pierrehumbert, 2008; Scarborough, 2010)
- Word sense disambiguation should provide more reliable estimates of predictability

# Semantic predictability

- Key words (Kalikow, Stevens, & Elliott, 1977)
  - Earlier words in the sentence can point to the identity of later words
  - Operationalized as semantically related words
  - Greater average relatedness of the context is more predictable
  - Word-context relatedness (Pucher, 2003)

$$rel_W(w|C) = \frac{1}{|C|} \sum_{w' \in C} rel(w, w')$$

# Semantic relatedness

- One of the best measures is Extended Lesk (Pedersen, Patwardhan, & Michelizzi, 2004)
    - Gloss overlaps
    - Stop words removed
    - N-word overlap scores $n^2$
    - Pairwise comparison of glosses and related senses' glosses
- Similar method used here
    - Bag of ngrams model
    - Same stop word list as Extended Lesk
    - Unigrams through five-grams extracted from glosses and related senses' glosses
    - Scored by summing their inverse document frequency in WordNet (Princeton University, 2010) and squaring

# Comparison of measures

- Sentences from previous studies (Kalikow et al., 1977; Scarborough, 2010)
  - Annotated as high or low predictability (Sentence completion task)
- Model using bag of ngrams with IDF performs best

| Measure | AIC | Probability of lowest AIC |
|---|---|---|
| Extended Lesk | 545.3 | 0.01 |
| Bag of words (IDF) | 540.3 | 0.13 |
| Bag of words | 547.4 | 0.003 |
| Bag of ngrams (IDF) | 536.3 | 1.0 |
| Bag of ngrams | 545.7 | 0.01 |

# Word sense disambiguation

- Simple Lesk (Kilgarriff, Rosenzweig, et al., 2000)
  - Overlap between words in context and WordNet gloss/example of all senses of a word
  - Weighted by IDF in WordNet
  - Window of +/-3 content words (Vasilescu, Langlais, & Lapalme, 2004)
    - Within 10 seconds on either side

# Data

- Buckeye Corpus of spontaneous speech corpus (Pitt et al., 2007)
- Words to be analyzed
  - Content words (Noun, adjective, verb, adverb) according to CELEX (Baayen, Piepenbrock, & Gulikers, 1995)
    - Part of speech tagging of spontaneous speech is difficult (Kübler, Scheutz, Baucom, & Israel, 2010)
  - No adjacent pauses or disfluencies

# Evaluation

- Linear mixed effects model
  - Random effects for Word and Speaker
    - Allows for Word Duration to vary between Words and Speakers
  - Random slopes for contextual factors
    - Allows for relationship between Word Duration and contextual factors to vary between Words and Speakers
- Three models
  - Control model (Gahl, Yao, & Johnson, 2012)
    - **Lexical factors** (frequency, neighbourhood density, phonotactic probability, length measures, POS)
    - **Contextual factors** (speaking rate before and after, conditional probability before and after, repetitions)
    - **Speaker factors** (age group, gender, average speaking rate)
  - Default sense model
    - Semantic predictability from relatedness using default senses
  - Disambiguated sense model
    - Semantic predictability from relatedness using disambiguated senses

# Results - semantic predictability

- Control model performs best
- Default performs better than Disambiguated

| Model | AIC | Pr |
|---|---|---|
| Control | 1538 | 1.0 |
| Default | 1627 | $< 0.001$ |
| Disambiguated | 1639 | $< 0.001$ |

# Results - word sense disambiguation

- Random slopes beneficial for Default model
- Random slopes detrimental for Disambiguated model

| Model | AIC | Pr |
|---|---|---|
| Default | 1627 | 1.0 |
| Disambiguated | 1639 | 0.002 |
| Default (no random slopes) | 1636 | 0.01 |
| Disambiguated (no random slopes) | 1635 | 0.01 |

# Discussion - semantic predictability

- Adding semantic predictability terms of any kind resulted in worse performance
- Possible reasons
  - Poor part of speech tagging
    - POS acts a filter on possible word senses (Vasilescu et al., 2004)
  - Unnormalized measure
    - Raw semantic relatedness score similar to count of word occurrences
    - Given a particular context, there could be many or few possible related senses
    - Different senses have different numbers of realizations
  - Difference between previous laboratory research and this study
    - Already a number of contextual variables in the control model
    - Conditional probabilities based on previous and following word could be accounting for predictability effects

# Discussion - word sense disambiguation

- Word sense disambiguation did have an interesting effect
  - Random slopes for predictability significantly improved default model but not disambiguated model
  - Disambiguation reduces variability of the relationship between word duration and predictability

# Future directions

- Improved spontaneous speech POS tagging
- Normalization of semantic predictability
- More general models for probability of word occurrence

Thank you!

# References I

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (version release 2)[CD-ROM]. *Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania*.

Clopper, C. G., & Pierrehumbert, J. B. (2008). Effects of semantic predictability and regional dialect on vowel space reduction. *Journal of the Acoustical Society of America*, *124*(3), 1682–1688.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*. (In press)

Kalikow, D., Stevens, K., & Elliott, L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America*, *61*(5), 1337–1351.

# References II


Kilgarriff, A., Rosenzweig, J., et al. (2000). English SENSEVAL: Report and results. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation, LREC.*

Kübler, S., Scheutz, M., Baucom, E., & Israel, R. (2010). Adding context information to part of speech tagging for dialogues. In *Proceedings of the Ninth International Workshop on Treebanks and Linguistic Theories (TLT)* (pp. 115–26).

Pedersen, T., Patwardhan, S., & Michelizzi, J. (2004). WordNet:: Similarity: measuring the relatedness of concepts. In *Demonstration Papers at HLT-NAACL 2004* (pp. 38–41).

Pitt, M. A., Dilley, L., Johnson, K., Kieslin, S., Raymond, W., Hume, E., et al. (2007). *Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu].* Columbus, OH: Department of Psychology, Ohio State University (Distributor).

# References III


Princeton University. (2010). *WordNet.* (http://wordnet.princeton.edu)

Pucher, M. (2003). Performance evaluation of WordNet-based semantic relatedness measures for word prediction in conversational speech. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL 2007)* (Vol. 18, pp. 129–132).

Scarborough, R. (2010). Lexical and contextual predictability: Confluent effects on the production of vowels. *Laboratory Phonology*, *10*, 557–586.

Vasilescu, F., Langlais, P., & Lapalme, G. (2004). Evaluating variants of the Lesk approach for disambiguating words. In *Proceedings of the Conference of Language Resources and Evaluations (LREC 2004)* (pp. 633–636).