Midterm

• NLU vs. NLG
• Ambiguity in NL – at different levels of analysis
• Language modeling
  – Generative models: what are they? Why important?
• Issues for counting words
• Zipf’s Law
• Given a sample text/document: derive n-gram probabilities using MLE; use them to compute the probability of (the next word in) a sequence
• Smoothing/interpolation
• Evaluating language models
• **What is a lexeme? A lexicon?**

• **Lexical semantic relations**

• **WSD**
  – Lesk’s dictionary-based approach
  – Supervised ML approach using collocational and co-occurrence features
  – Yarowsky’s bootstrapping approach
  – Evaluation using precision and recall

• **POS tagsets, tagging**

• **Lectures through Tues 3/4**