**Lecture 2 – Smoothing (Continued)**

1. Methods for smoothing
   1. Add (see lecture 2)
   2. Linear Interpolation (see lecture 2)
   3. Discounting (Kneser-Ney Smoothing)
      1. Very successful in NLP
      2. Idea: Take some probability mass from all existing bigrams, and redistribute (in some way, the probability among unseen bigrams)
         1. Create a new distribution
      3. Promiscuity
         1. Redistributing based on unigram probability not always good
         2. Instead, let

**Lecture 3 – Topic Models and EM**

1. Syntax trees shown to help statistical NLP, but not neural NLP
2. Higher level models – semantics trees? Not convincingly helpful as well
3. How to model whole documents of text?
   1. Hierarchical Segmentation?
   2. Centering?
   3. RST?
4. Topic Models
   1. Choose topics in a document, and generate text from those topics
      1. Learned in an unsupervised manner
   2. Blend = topic distribution (specific to individual documents)
   3. Topic = distribution over words (shared across the collection of documents)
      1. ,
   4. Generally, some topics found make sense…but many (most?) do not
   5. Example
      1. Topic 1: ,
      2. Topic 2: ,
   6. How to generate documents, given ?
      1. Model 1 (Not a topic model)
         1. For ,
      2. Model 2 (word order oblivious)
         1. For ,
   7. How can we compute ?
      1. Model 1
      2. Model 2
   8. How to estimate and ?
      1. Observed case
         1. Given
      2. Unobserved case
         1. Use stochastic gradient descent
         2. Use EM algorithm (will discuss next time)