Background & Motivation

- Part-of-speech tagging (PoS tagging) is the process of marking up a word in a text (corpus) as corresponding to a particular part of speech (e.g.: nouns, verbs, adjectives, adverbs).

- We are interested in the setting where the complexity of language system (i.e., the number of part-of-speech tags) increases by time.

  - In child language acquisition, Brown's Stages (Brown,1973) provide a framework for understanding and predicting the path that normal expressive language development in English usually takes, in terms of morphology and syntax.

  - In the historical development of language, a large number of new parts of speech emerged. 例如，在汉语史中，绝大部分介词都是从动词虚化而来的；量词（个/枚/匹）在东汉才大量产生；助词（的/地/得/着/过/了）基本都在唐代后产生。

Our Work Plan

- We have already found some code that implements HDP-HMM by sampling method.

- We will:
  1. verify the code
  2. implement the variational inference method
  3. compare the results of the two methods on different datasets, e.g.: child language acquisition dataset, the English/Chinese historical documents in different periods.