Assignment 7
L415 / L615
Due Wednesday, December 9

1. Based on your experience with NSP from a few assignments ago: Just for fun, POS tag some Enron email data using a POS tagger (preferably TreeTagger, but you’re welcome to try another), and then see what the significant collocations of POS tags are that NSP gives you. Be sure that you take the tagger output and operate strictly on the POS tags, i.e., no words involved.

   What are you doing when you do this? (i.e., what does it mean?) Do you find anything interesting?

2. Parse a small portion of the same Enron data using the Stanford CoreNLP tools. Based on the parse output you observe, propose an application for processing this data that would likely want to use such parses (and/or any other NLP components) and defend the proposal with example parses from your data.

3. Using the Penn Treebank corpus (/Volumes/Data/en/penntreebankv3/parsed/mrg/pos/) and Tregex, provide:
   (a) The number of relative clauses modifying subject NPs
   (b) The number of relative clauses modifying object NPs
   (c) Bonus: The number of headless relative clauses modifying subject NPs
   (d) Bonus: The number of headless relative clauses modifying object NPs

   For each case, also provide the Tregex query you used and any difficulties you encountered. You may restrict your search to a particular directory or set of directories (only section 00, only sections 02-10, etc.)

4. [nothing to turn in:] Work on your project and get ready to give a presentation!