Syntax: grammars, datasets, tree search

Linguistics 165, Professor Roger Levy
27 February 2015

1. Resources for today’s lecture:
   - Penn Treebank hand-annotated parsed dataset: [http://www.cis.upenn.edu/~treebank/](http://www.cis.upenn.edu/~treebank/) This is installed in /home/linux/ieng6/ln165w/public/stanford-tregex.

2. Goals for today’s lecture:
   - Use NLTK to write fragments of an English syntactic grammar, illustrating problems of coverage and ambiguity
   - Browse the Penn Treebank a bit, to get a sense of the diversity and richness of natural language syntax
   - Learn how to do tree searches with Tregex