

# Assignment 3

L615

Due Thursday, March 12

1. I'm curious about the argument structure of *need* in English, and for this assignment, we'll use a POS tagged corpus. Before searching, outline the different possibilities for *need*; then, discuss how you searched for these; and, finally, present and discuss the resulting frequencies. Be clear in how you determined how to do the searches (e.g., how do you define a noun phrase, a sentential clause, etc.?).
  - Use the POS-tagged Brown corpus available on `jones` at `/Volumes/Data/Corpora/en/brown/`
  - I'm providing a perl script from the course webpage which might help you get started. You could also try using any sort of concordancing tool you can find. I'm willing to help you work with Perl, but be sure you've clearly articulated what it is you're searching for first.
2. Write a Perl program that does the following:
  - Reads in a single short text file
  - Breaks the text into "words" (where a word is anything between whitespaces)
  - Calls a subroutine for each word to determine whether the word ends in a period or a comma: return 1 for yes, 0 for no.
  - Outputs each non-punctuation-ending word to a file, one word per line
3. Starting with the research question that you described in the first 2 assignments, or switching to another one that interests you a bit more ...
  - (a) Describe the question on one paragraph, as it stands now (this may be the same as the last assignment).
  - (b) Find a corpus which will be applicable in answering your research question, and describe *in detail* how it is appropriate. Note: If you have a corpus in mind, but cannot access it, please let me know immediately.
  - (c) Also in detail, describe the limitations of the corpus for answering your research question, and discuss briefly how you will attempt to overcome these limitations.