

# Assignment 4

L615

Due Thursday, April 9

1. Work with one of the annotation tools we talked about (or that you find on [devoted.to/corpora](http://devoted.to/corpora))
  - Either use a text that you want to annotate or download a text from Project Gutenberg (<http://www.gutenberg.org>)
  - POS tag this text
  - Hand-correct the tags for the first 200 words of the text
  - Write up a 2 page summary of what you did, including:
    - A critique (positives and negatives) of the annotation tool you used
    - A description of how you POS tagged the text and how easy it was to get the POS annotation into the annotation tool
    - A discussion of what the challenges were in hand-correcting the POS tags
2. Write a Perl program that does the following:
  - Reads in the Brown corpus
  - Attempts to find the lines which have particular uses of *need* in them, namely:
    - NEED to be X-ED
    - NEED X-EDNote that NEED is a lemma, so it should match any form of *need*, and X-ED may not end in *ed* (e.g., *thrown*)
  - Outputs the different types of instances to 2 separate filesI highly recommend looking at the `help.pl` file from one of our classes a while back (February 11).
3. Final project stuff:
  - (a) Acquire the corpus that you need for your project, and/or describe to me at what stage you're at in the corpus acquisition process.
  - (b) Provide me a one-page (or so) outline of what your final paper will consist of. That is, what is the overall thesis? What will be your main arguments? Where does your corpus data fit in? etc.