Assignment 5

L615

Due Wednesday, April 10

1. Using the Brown corpus (/Volumes/Data/en/brown/ and regular expression searching in Perl, try to find (i.e., approximate):

   (a) The number of relative clauses modifying subject NPs, as in:
       The guy *that you met* is an old friend of mine.
   (b) The number of relative clauses modifying object NPs, as in:
       That guy hates the author *that I was telling you about*.
   (c) The number of headless relative clauses modifying subject NPs, as in:
       The guy *you met* is an old friend of mine.
   (d) The number of headless relative clauses modifying object NPs, as in:
       That guy hates the author *I was telling you about*.

   For each case, provide the regular expression(s) you used and any difficulties you encountered.

2. Using the Penn Treebank corpus (/Volumes/Data/en/penntreebankv3/parsed/mrg/wsj/) and Tregex, provide:

   (a) The number of relative clauses modifying subject NPs
   (b) The number of relative clauses modifying object NPs
   (c) The number of headless relative clauses modifying subject NPs
   (d) 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.)

3. Final project: Work on it. All I’m going to ask for for this assignment is a status report, but I expect you to keep plugging away at it. For this brief report (a couple of paragraphs), ask yourself: “what do I need feedback on?” and write accordingly (e.g., reports of bottlenecks, puzzles, triumphs, ...).