Assignment 1

L415 / L615

Due Wednesday, September 9

1. Go to the interface at http://corpus.byu.edu/ and search for different orderings of adverbs and adverbial/prepositional phrases in either COCA (American English) or the BNC (British English). Find a page that hypothesizes the order of adverbial phrases when occurring together (e.g., http://grammar.ccc.commnet.edu/grammar/adverbs.htm posits manner < place < frequency < time < purpose), or propose one of your own.

After figuring out how you are going to approximate a search for various kinds of adverbial phrases, think about questions such as the following: do the frequencies of the orderings you find more or less match the hypothesis? Are there notable exceptions? Is this consistent throughout different registers of English (within the same variety)? Do you see different trends in different parts of the sentence? Do you find an avoidance of multiple adverbials together?

Feel free to tackle the questions which seem most pertinent and most interesting to you. The importance now is to search, analyze, and explore.

2. [Moved to next assignment] You want to build a corpus of all the (handwritten) assignments you've turned in during any portion of your schooling. In approximately two pages:

   (a) Discuss potential applications for this corpus. Keep these in mind as you address the next points.
   (b) Propose a corpus structure.
   (c) Propose a set of meta-data tags to use for your files.
   (d) Take a page or two of your writing and scan it, performing OCR in the process. Describe the challenges, referencing specific examples. Are there systematic/automatic solutions for your challenges? What is the potential impact on downstream processing of the corpus?

3. Write a perl script that does the following (and make sure it works):
   - Has two variables: $firstname and $lastname
   - Prints out the concatenation of those variables, with a space in between and a newline at the end.

   For this question, be sure to send me the electronic file, so I can verify that it works.

4. I want you to start thinking about a question to investigate for your final project. This could be something you have already discussed in other classes (perhaps intuition-based), something you always wanted to find out about a particular language, some technology you’d like to see developed for handling corpus data, some corpus you want to collect for various reasons, or whatever it is that interests you. You can change your topic/question later in the semester, so do not worry about overcommitting here.

   Write up three paragraphs describing the issue, breaking it down as follows:

   (a) Describe the question and give examples based on intuition or previous research.
   (b) Describe how the question can be approached from a corpus-based perspective. What kind of corpora do you need? Are there specific requirements with regard to the corpus (e.g., does it need to be balanced, specialized, etc.?) What types of information do you need to extract from the corpora? (The specific questions you address here depend upon the exact nature of your proposal.)
   (c) Describe for a linguist why your project will be of interest to them and why using corpus data is a valid way to approach the question or provides insight you cannot get elsewhere.