Assignment 5

L545

Due Wednesday, March 6

1. Choose Your Own Adventure: I want you to select an implementation of a language modeling toolkit **or** of a Brill tagger and report on its:

   - Usability/ease of use
   - Accuracy (perhaps reporting with different options)
   - Availability of different options
   - Internal details—i.e., not an exhaustive search of the code, but rather an understanding of what this implementation does that may be different from other implementations: how many templates are used, how smoothing is done, etc.

   I'm going to make various converted versions of the SUSANNE Corpus (`http://www.grsampson.net/Resources.html`) available via oncourse. If you can't find a format that fits your needs, we can try to rework the data.

   Here are some lists of tools, but this is by no means exhaustive.

   - Language model toolkits:
     - `http://www.speech.sri.com/projects/srilm/`
     - `http://code.google.com/p/mitlm/`
     - `http://www.speech.cs.cmu.edu/SLM_info.html`
   - Brill tagger implementations:
     - `http://www.cs.jhu.edu/~rflorian/fntbl/`
     - `http://www.ling.gu.se/~lager/Mutbl/course/index.html`
     - `http://gposttl.sourceforge.net`
     - `http://en.wikipedia.org/wiki/Brill_tagger#Code`
     - `http://nltk.org/_modules/nltk/tag/brill.html` (see also general NLTK documentation: `http://nltk.org/book/ch05.html`) Note: you can use the Brown corpus or other pre-installed NLTK corpora if you use the NLTK Brill tagger.

2. Determine the set of properties of the following relations on the set of all people. In each case, make the strongest possible statement, e.g., call a relation irreflexive whenever possible instead of nonreflexive.[1]

  (a) is a child of

  (b) is a brother of

  (c) is a descendant of

  (d) is an uncle of

3. Using section 12.3.4 as a help and using what you know about English pronouns, write out all the CFG rules you need to capture the following sentences, where * indicates an ungrammatical sentence:

(1) a. I eat.   h. *Him/Her eats.

      b. *I/Me eats.   i. We eat.

      c. *Me eat.   j. *We/Us eats.

      d. You eat.   k. *Us eat.

      e. *You eats.   l. They eat.

      f. He/She eats.   m. *They/Them eats.

      g. *He/She eat.   n. *Them eat.