

Homework #4

1. Using the first 100,000 words of the tagged Brown corpus from the website, build a bigram tagger. Test this on the next 10,000 words and report percent correct for words and sentences.
2. Exercise 6.2 in Manning & Schütze.
3. **539 students only:** Using the first 100,000 words of the tagged Brown corpus from the website, build a *trigram* tagger. Test this on the next 10,000 words and report percent correct for words and sentences.

**Things to remember:**

1. This is due by email by the **beginning** of class on **Oct. 28**. (You can email it to me before then if you prefer.)
2. Code must be in the form of a working/runnable m-file that you email to me.
3. Prose questions can be answered as a commented section in an m-file or in a pdf file. Microsoft Word documents are *not* acceptable.
4. The prose can be no more than *three* double-spaced pages.
5. Keep in mind that there may be funny symbols here that you've never used before. Leave time to make sure you have them right.
6. Remember: nothing late. Do not wait until the last minute to do this.
7. You may certainly talk to each other about this and other assignments, but everyone must turn in their own work. (If you do talk to others, I need to see evidence that you are doing more than listening and writing down what others say.)