This lecture is about **part-of-speech (POS) tags**. POS tags can be defined in a casual way based on syntactic coherence, but linguists tend to use distributional arguments for defining the parts of speech.

The textbook gives an extensive discussion of what the different kinds of tags standardly considered are. It's important to remember that a tag set is a convention, and what's useful or works well for one problem may be different than what's useful or works well for another problem. You should know the main and common POS tags, and which ones are "open" and "closed" class. You should recognize that the tags can be defined in coarse ways (e.g., "verb") or more fine-grained ways (e.g., "singular present tense transitive verb in 3rd person"). In languages that are morphologically rich, this interacts with morphology.

For a couple of examples of tag sets:

- Brown corpus, `http://www.comp.leeds.ac.uk/ccalas/tagsets/brown.html`
- "A Universal Part-of-Speech Tagset" (2011). `http://www.petrovi.de/data/universal.pdf` `http://code.google.com/p/universal-pos-tags/`

The most commonly used one for English is the Penn Treebank tagset.

The two difficulties in POS tagging are (1) knowing what the tag(s) for a given word are (dictionary knowledge), and coping with unknown words, and (2) disambiguating words that have more than one possible POS.

We discussed a very simple baseline that gets around 90% on the Penn Treebank tagset and data: pick the most likely tag for each word.

We can frame POS tagging as a classification problem, and we discussed a couple of variants of the noisy channel model as applied to this problem.

When we want to take into account the interaction among POS tagging decisions of words that occur in sequence, things become more challenging, because your classification decisions are now *interdependent*. In the next lecture we will see an elegant solution to this problem.