# Tagged Text and Tagging

* In tagged text each token is assigned a part of speech tag (POS)
* A tagger is a program which automatically ascribes tags to words in text
* From the context we are (most often) able to determine the tag, but some sentences are genuinely ambiguous and hence so are the tags

# Various POS tag sets

* A tagged text is tagged according to a fixed small set of tags
* There are various tagsets:
  + Brown Tagset: 87 tags, Versions with extended tags <original>-<more>
  + Penn treebank tags: 35+9 punctuation tags
  + Universal POS Tagset: 12 tags

# Tagging as Sequence Classification

* Classification (earlier):
  + A well-defined set of observations O
  + A given set of classes, S = {s1, s2, …, sk}
  + Goal: a classifier, γ, a mapping from sequences of elements from O to sequences of elements from S: γ (o1, o2, …, on) = (sk1, sk2, …, skn)

# Baseline Tagger