Sequence Tagging

• Today
  – Part-of-speech tagging
    • Introduction
Part of speech tagging

“There are 10 parts of speech, and they are all troublesome.”
-Mark Twain

• POS tags are also known as word classes, morphological classes, or lexical tags.

• Typically much larger than Twain’s 10:
  – Penn Treebank: 45
  – Brown corpus: 87
  – C7 tagset: 146
Part of speech tagging

- Assign the correct part of speech (word class) to each word/token in a document
  “The/DT planet/NN Jupiter/NNP and/CC its/PPS moons/NNS are/VBP in/IN effect/NN a/DT mini-solar/JJ system/NN ,/, and/CC Jupiter/NNP itself/PRP is/VBZ often/RB called/VBN a/DT star/NN that/IN never/RB caught/VBN fire/NN ./.”

- Needed as an initial processing step for a number of language technology applications
  – Answer extraction in Question Answering systems
  – Base step in identifying syntactic phrases for IR systems
  – Critical for word-sense disambiguation
  – Information extraction
  – …
Why is p-o-s tagging hard?

• Ambiguity
  – He will race/VB the car.
  – When will the race/NOUN end?
  – The boat floated/VBD.
  – The boat floated/VBD down Fall Creek.
  – The boat floated/VBN down Fall Creek sank.

• Average of ~2 parts of speech for each word

• The number of tags used by different systems varies a lot. Some systems use < 20 tags, while others use > 400.
Hard for Humans

• particle vs. preposition
  – He talked over the deal.
  – He talked over the telephone.

• past tense vs. past participle
  – The horse walked past the barn.
  – The horse walked past the barn fell.

• noun vs. adjective?
  – The executive decision.

• noun vs. present participle
  – Fishing can be fun.

To obtain gold standards for evaluation, annotators rely on a set of tagging guidelines.

From Ralph Grishman, NYU
# Penn Treebank Tagset

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Example</th>
<th>Tag</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>Coordin. Conjunction</td>
<td><em>and, but, or</em></td>
<td>SYM</td>
<td>Symbol</td>
<td><em>+%, &amp;</em></td>
</tr>
<tr>
<td>CD</td>
<td>Cardinal number</td>
<td><em>one, two, three</em></td>
<td>TO</td>
<td>“to”</td>
<td><em>to</em></td>
</tr>
<tr>
<td>DT</td>
<td>Determiner</td>
<td><em>a, the</em></td>
<td>UH</td>
<td>Interjection</td>
<td><em>ah, oops</em></td>
</tr>
<tr>
<td>EX</td>
<td>Existential ‘there’</td>
<td><em>there</em></td>
<td>VB</td>
<td>Verb, base form</td>
<td><em>eat</em></td>
</tr>
<tr>
<td>FW</td>
<td>Foreign word</td>
<td><em>mea culpa</em></td>
<td>VBD</td>
<td>Verb, past tense</td>
<td><em>ate</em></td>
</tr>
<tr>
<td>IN</td>
<td>Preposition/sub-conj</td>
<td><em>of, in, by</em></td>
<td>VBG</td>
<td>Verb, gerund</td>
<td><em>eating</em></td>
</tr>
<tr>
<td>JJ</td>
<td>Adjective</td>
<td><em>yellow</em></td>
<td>VBN</td>
<td>Verb, past participle</td>
<td><em>eaten</em></td>
</tr>
<tr>
<td>JJR</td>
<td>Adj., comparative</td>
<td><em>bigger</em></td>
<td>VBP</td>
<td>Verb, non-3sg pres</td>
<td><em>eats</em></td>
</tr>
<tr>
<td>JJS</td>
<td>Adj., superlative</td>
<td><em>wildest</em></td>
<td>VBZ</td>
<td>Verb, 3sg pres</td>
<td><em>eats</em></td>
</tr>
<tr>
<td>LS</td>
<td>List item marker</td>
<td><em>1, 2, One</em></td>
<td>WDT</td>
<td>Wh-determiner</td>
<td><em>which, that</em></td>
</tr>
<tr>
<td>MD</td>
<td>Modal</td>
<td><em>can, should</em></td>
<td>WP</td>
<td>Wh-pronoun</td>
<td><em>what, who</em></td>
</tr>
<tr>
<td>NN</td>
<td>Noun, sing. or mass</td>
<td><em>llama</em></td>
<td>WPS</td>
<td>Possessive wh-</td>
<td><em>whose</em></td>
</tr>
<tr>
<td>NNS</td>
<td>Noun, plural</td>
<td><em>llamas</em></td>
<td>WRB</td>
<td>Wh-adverb</td>
<td><em>how, where</em></td>
</tr>
<tr>
<td>NNP</td>
<td>Proper noun, singular</td>
<td><em>IBM</em></td>
<td>$</td>
<td>Dollar sign</td>
<td>$</td>
</tr>
<tr>
<td>NNPS</td>
<td>Proper noun, plural</td>
<td><em>Carolinias</em></td>
<td>#</td>
<td>Pound sign</td>
<td>#</td>
</tr>
<tr>
<td>PDT</td>
<td>Predeterminer</td>
<td><em>all, both</em></td>
<td>“</td>
<td>Left quote</td>
<td>(* or ““*</td>
</tr>
<tr>
<td>POS</td>
<td>Possessive ending</td>
<td><em>‘s</em></td>
<td>”</td>
<td>Right quote</td>
<td>(* or ””*</td>
</tr>
</tbody>
</table>
| PP   | Personal pronoun       | *I, you, he*  | (    | Left parenthesis       | ([, (, {, <)
| PPS  | Possessive pronoun     | *your, one’s* | )    | Right parenthesis      | ([], ), }, >) |
| RB   | Adverb                 | *quickly, never* | ,    | Comma                  | ,             |
| RBR  | Adverb, comparative    | *faster*      | .    | Sentence-final punct    | (., !, ?)     |
| RBS  | Adverb, superlative    | *fastest*     | :    | Mid-sentence punct      | (:, ;, ..., --) |
Let’s give it a try...
P-o-s tagging exercise
1. It is a nice night.

   It/PRP is/VBZ a/DT nice/JJ night/NN ./. 
5. . . . I am sitting in Mindy’s restaurant putting on the gefillte fish, which is a dish I am very fond of, . . .

. . . I/PRP am/VBP sitting/VBG in/IN Mindy/NNP ’s/POS restaurant/NN putting/VBG on/RP the/DT gefillte/NN fish/NN ,/, which/WDT is/VBZ a/DT dish/NN I/PRP am/VBP very/RB fond/JJ of/RP ,/, . . .
Think buffalo

buffalo buffalo buffalo buffalo buffalo buffalo buffalo buffalo buffalo.

Buffalo buffalo Buffalo buffalo buffalo buffalo buffalo Buffalo buffalo.

Buffalo buffalo, Buffalo buffalo buffalo, buffalo Buffalo buffalo.
Think buffalo

n1. the city of Buffalo, NY
n2. an animal…the American bison
v. to bully, confuse, deceive, or intimidate

Buffalo\textsuperscript{n1} buffalo\textsuperscript{n2} Buffalo\textsuperscript{n1} buffalo\textsuperscript{n2} buffalo\textsuperscript{v} buffalo\textsuperscript{v} Buffalo\textsuperscript{n1} buffalo\textsuperscript{n2}.

[Those] (Buffalo buffalo) [whom] (Buffalo buffalo) buffalo, buffalo (Buffalo buffalo).
[Those] buffalo(es) from Buffalo [that are intimidated by] buffalo(es) from Buffalo intimidate buffalo(es) from Buffalo.
Bison from Buffalo, New York, who are intimidated by other bison in their community, also happen to intimidate other bison in their community.
THE buffalo FROM Buffalo WHO ARE buffaloed BY buffalo FROM Buffalo, buffalo (verb) OTHER buffalo FROM Buffalo.
Among easiest of NLP problems

• State-of-the-art methods achieve ~97% accuracy.

• Simple heuristics can go a long way.
  – ~90% accuracy just by choosing the most frequent tag for a word (MLE)
  – To improve reliability: need to use some of the local context.

• But defining the rules for special cases can be time-consuming, difficult, and prone to errors and omissions
Approaches

1. **rule-based**: involve a large database of hand-written disambiguation rules, e.g. that specify that an ambiguous word is a noun rather than a verb if it follows a determiner.

2. **learning-based**: resolve tagging ambiguities by using a training corpus to compute the probability of a given word having a given tag in a given context.
   - HMM tagger

3. **hybrid ML-/rule-based**: E.g. transformation-based tagger (Brill tagger); learns symbolic rules based on a corpus.

4. **ensemble methods**: combine the results of multiple taggers.
• Today
  – Part-of-speech tagging
    • HMM’s for p-o-s tagging
HMM p-o-s Tagger

Given $W = w_1, \ldots, w_n$, find $T = t_1, \ldots, t_n$ that maximizes

$$P(t_1, \ldots, t_n|w_1, \ldots, w_n)$$

Restate using Bayes’ rule:

$$(P(t_1, \ldots, t_n) \ast P(w_1, \ldots, w_n|t_1, \ldots, t_n))/P(w_1, \ldots, w_n)$$

Ignore denominator...
Make independence assumptions...
$P(t_1, \ldots, t_n)$: approximate using **n-gram model**

**bigram** $\prod_{i=1,n} P(t_i | t_{i-1})$

**trigram** $\prod_{i=1,n} P(t_i | t_{i-2}t_{i-1})$
\[ P(w_1, \ldots, w_n \mid t_1, \ldots, t_n): \text{approximate by assuming that a word appears in a category independent of its neighbors} \]

\[ \prod_{i=1,n} P(w_i \mid t_i) \]

Assuming bigram model:

\[ P(t_1, \ldots, t_n) \ast P(w_1, \ldots, w_n \mid t_1, \ldots, t_n) \approx \]

\[ \prod_{i=1,n} P(t_i \mid t_{i-1}) \ast P(w_i \mid t_i) \]

transition probabilities  
lexicographic generation probabilities