LING/C SC/PSYC 438/538

Lecture 20
Sandiway Fong
Writing natural language grammars

• Need:
  • Ability to program with grammars
  • Prolog grammar rules
  +
  • Knowledge of language
    • You have the knowledge but it’s not completely accessible to you…
    • Textbook:
      • chapter 5, sections 1 and 2
      • chapter 12
Natural Language Parsing

• Syntax trees are a big deal in NLP

• Stanford Parser (see also Berkeley Parser)
  • http://nlp.stanford.edu:8080/parser/index.jsp
  • Uses probabilistic rules learnt from the Penn Treebank corpus
  • Produces parses in the same format
  • (modulo empty categories and subtags)
  • Also produces dependency diagrams

We do a lot with Treebanks in the follow-on course to this one (LING 581, Spring semester)
Natural Language Parsing

- Penn Treebank (WSJ section):
  - parsed by human annotators
  - Efforts by the Hong Kong Futures Exchange to introduce a new interest-rate futures contract continue to hit snags despite the support the proposed instrument enjoys in the colony’s financial community.

```plaintext
1 (S
2   (NP-SBJ-2
3     (NP (NN: Efforts)
4       ($) (-NONE- $20M-3)))
5     (VP (IN: by)
6       (NP (DT: the) (NP: Hong Kong) (NN: Futures) (NNP: Exchange)))
7     ($) (-NONE- $2))
8     (VP (TO: to)
9       (VP (VB: introduce)
11       (VP (VB: continue)
12         (NP (NN: enjoy)
13           ($) (-NONE- $2))
14           (VP (TO: to)
15             (VP (VB: hit)
16               (NP (NN: snags)))))
17              (,,)
18              (PP (IN: SN despite)
19                (NP (DT: the) (NN: support)
20                  (SBAR
21                    (NNP-1 (-NONE- 0)
22                      ($) (-NONE- $2)
23                      (VP (VBP: enjoy)
24                        ($) (-NONE- $2))
25                        (PP (IN: SN in)
26                          (NP (DT: the) (NN: Futures) (IN: instrument)))
27                          (VP (VBP: enjoy)
28                            (NP (NONE- $2))
29                            (PP (IN: SN in)
30                              (NP (DT: the) (NN: colony) (POS: 's'))
31                              (JJ: financial) (NN: community)))))))
32 (,,))
33)
```
Natural Language Parsing

• Penn Treebank (WSJ section):
  • parsed by human annotators
  • Efforts by the Hong Kong Futures Exchange to introduce a new interest-rate futures contract continue to hit snags despite the support the proposed instrument enjoys in the colony’s financial community.

Tree display software: tregex (from Stanford University)
Natural Language Parsing

**Stanford Parser**

Please enter a sentence to be parsed:

Efforts by the Hong Kong Futures Exchange to introduce a new interest-rate futures contract continue to hit snags despite the support the proposed instrument enjoys in the colony’s financial community.

Language: [English]  [Sample Sentence]  [Parse]

**Tagging**

Efforts/NNS by/IN the/OQ Hong/OQ Kong/OQ Futures/OQ Exchange/OQ introduce/VB a/OQ new/OQ interest-rate/NNS futures/NNS contract/NNS continue/VB hit/VB snags/NNS despite/IN the/OQ support/NN the/OQ proposed/VBN instrument/NNS enjoy/VB in/IN the/OQ colony/NN ’s/OQ financial/OQ community/OQ ./

**Parse**

(ROOT
  (S
    (NP (NN (NN (NN (NNP (NNP (NNP (NNP (NNP (NNP (NNP (NNP Efforts)))))))))))
    (PP (IN by))
    (NP (DT the) (NNP Hong) (NNP Hong) (NNP Futures) (NNP Exchange))))
  (S
    (VP (TO to))
    (VP (VB introduce))
    (VP (VB continue))
    (NP (DT a) (JJ new) (NN (NNP interest-rate) (NN (NNP Futures) (NN (NNP contract))))))
Natural Language Parsing

• Comparison between human parse and machine parse:
  • empty categories not recovered by parsing, otherwise a good match
Natural Language Parsing
Part of Speech (POS)

JM Chapter 5

• Parts of speech
  • Classic eight parts of speech:
    • e.g. englishclub.com =>
    • traced back to Latin scholars, back further to ancient Greek (Thrax)
    • not everyone agrees on what they are..

  The textbook lists:
  • open class 4 (noun, verbs, adjectives, adverbs)
  • closed class 7 (prepositions, determiners, pronouns, conjunctions, auxiliary verbs, particles, numerals)

• or what the subclasses are
  • e.g. what is a Proper Noun?
  • Saturday, April
  • Textbook answer below ...

Nouns are traditionally grouped into proper nouns and common nouns. Proper nouns, like *Regina, Colorado,* and *IBM,* are names of specific persons or entities. In English, they generally aren’t preceded by articles (e.g., *the book is upstairs,* but *Regina is upstairs*). In written English, proper nouns are usually capitalized.

---

<table>
<thead>
<tr>
<th>part of speech</th>
<th>function or &quot;job&quot;</th>
<th>example words</th>
<th>example sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Verb</em></td>
<td>action or state</td>
<td>(to) be, have, do, like, work, sing, can, must</td>
<td>EnglishClub.com is a web site. I like EnglishClub.com.</td>
</tr>
<tr>
<td><em>Noun</em></td>
<td>thing or person</td>
<td>pen, dog, work, music, town, London, teacher, John</td>
<td>This is my dog. He lives in my house. We live in London.</td>
</tr>
<tr>
<td><em>Adjective</em></td>
<td>describes a noun</td>
<td>a, an, the, 69, some, good, big, red, well, interesting</td>
<td>My dog is big. I like big dogs.</td>
</tr>
<tr>
<td><em>Adverb</em></td>
<td>describes a verb, adjective or adverb</td>
<td>quickly, silently, well, badly, very, really</td>
<td>My dog eats quickly. When he is very hungry, he eats really quickly.</td>
</tr>
<tr>
<td><em>Pronoun</em></td>
<td>replaces a noun</td>
<td>I, you, he, she, some</td>
<td>Tara is Indian. She is beautiful.</td>
</tr>
<tr>
<td><em>Preposition</em></td>
<td>links a noun to another word</td>
<td>to, at, after, on, but</td>
<td>We went to school on Monday.</td>
</tr>
<tr>
<td><em>Conjunction</em></td>
<td>joins clauses or sentences or words</td>
<td>and, but, when</td>
<td>I like dogs and I like cats. I like cats and dogs. I like dogs but I don’t like cats.</td>
</tr>
<tr>
<td><em>Interjection</em></td>
<td>short exclamation, sometimes inserted into a sentence</td>
<td>ooh, augh! Huh, well...</td>
<td>Ouch! That hurts! Hi! How are you? Well, I don’t know.</td>
</tr>
</tbody>
</table>

* Some grammar sources categorize English into 9 or 10 parts of speech. At EnglishClub.com, we use the traditional categorization of 8 parts of speech. Examples of other categorizations are:
  • Verbs may be treated as two different parts of speech:
    - *Lexical Verbs* (work, like, run)
    - *Auxiliary Verbs* (be, have, must)
  • Determiners may be treated as a separate part of speech, instead of being categorized under *Adjectives*
Part of Speech (POS)

- Getting POS information about a word
  1. dictionary
  2. pronunciation: e.g. are you content with the content of the slide?
  3. possible n-gram sequences
    e.g. *pronoun << common noun the << common noun
  4. structure of the sentence/phrase (Syntax)
  5. possible inflectional endings:
    e.g. V-s/-ed/-en/-ing
    e.g. N-s

Task: POS tagging

45 tags listed in textbook
36 POS + 10 punctuation

In computational linguistics, the Penn Treebank tagset is the most commonly used tagset (reprinted inside the front cover of your textbook). 45 tags listed in textbook 36 POS + 10 punctuation
Part of Speech (POS)

- [http://faculty.washington.edu/dillon/GramResources/penntable.html](http://faculty.washington.edu/dillon/GramResources/penntable.html)

<table>
<thead>
<tr>
<th>POS Tag</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>coordinating conjunction</td>
<td>and</td>
</tr>
<tr>
<td>CD</td>
<td>cardinal number</td>
<td>1, three</td>
</tr>
<tr>
<td>DT</td>
<td>determiner</td>
<td>the</td>
</tr>
<tr>
<td>EX</td>
<td>existential there</td>
<td>there is</td>
</tr>
<tr>
<td>FW</td>
<td>foreign word</td>
<td>d’hoevre</td>
</tr>
<tr>
<td>IN</td>
<td>preposition/subordinating conjunction</td>
<td>in, of, like, after, that</td>
</tr>
<tr>
<td>JJ</td>
<td>adjective</td>
<td>green</td>
</tr>
<tr>
<td>JJR</td>
<td>adjective, comparative</td>
<td>greener</td>
</tr>
<tr>
<td>JJS</td>
<td>adjective, superlative</td>
<td>greenest</td>
</tr>
<tr>
<td>LS</td>
<td>list marker</td>
<td>1)</td>
</tr>
<tr>
<td>MD</td>
<td>modal</td>
<td>could, will</td>
</tr>
<tr>
<td>NN</td>
<td>noun, singular or mass</td>
<td>table</td>
</tr>
<tr>
<td>NNS</td>
<td>noun plural</td>
<td>tables</td>
</tr>
<tr>
<td>NP</td>
<td>proper noun, singular</td>
<td>John</td>
</tr>
<tr>
<td>NPS</td>
<td>proper noun, plural</td>
<td>Vikings</td>
</tr>
</tbody>
</table>

- NNP
- NNPS
Part of Speech (POS)

<table>
<thead>
<tr>
<th>POS Tag</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDT</td>
<td>predeterminer</td>
<td><em>both</em> the boys</td>
</tr>
<tr>
<td>POS</td>
<td>possessive ending</td>
<td><em>friend's</em></td>
</tr>
<tr>
<td>PP</td>
<td>personal pronoun</td>
<td><em>I, he, it</em></td>
</tr>
<tr>
<td>PP$</td>
<td>possessive pronoun</td>
<td><em>my, his</em></td>
</tr>
<tr>
<td>RB</td>
<td>adverb</td>
<td><em>however, usually, naturally, here, good</em></td>
</tr>
<tr>
<td>RBR</td>
<td>adverb, comparative</td>
<td><em>better</em></td>
</tr>
<tr>
<td>RBS</td>
<td>adverb, superlative</td>
<td><em>best</em></td>
</tr>
<tr>
<td>RP</td>
<td>particle</td>
<td><em>give up</em></td>
</tr>
<tr>
<td>SYM</td>
<td>Symbol</td>
<td><em>$, %</em></td>
</tr>
<tr>
<td>TO</td>
<td><em>to</em></td>
<td><em>to go, to him</em></td>
</tr>
</tbody>
</table>
**Part of Speech (POS)**

<table>
<thead>
<tr>
<th>POS Tag</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>UH</td>
<td>interjection</td>
<td>uhhuhhhuhh</td>
</tr>
<tr>
<td>VB</td>
<td>verb, base form</td>
<td>take</td>
</tr>
<tr>
<td>VBD</td>
<td>verb, past tense</td>
<td>took</td>
</tr>
<tr>
<td>VBG</td>
<td>verb, gerund/present participle</td>
<td>taking</td>
</tr>
<tr>
<td>VBN</td>
<td>verb, past participle</td>
<td>taken</td>
</tr>
<tr>
<td>VBP</td>
<td>verb, sing. present, non-3d</td>
<td>take</td>
</tr>
<tr>
<td>VBZ</td>
<td>verb, 3rd person sing. present</td>
<td>takes</td>
</tr>
<tr>
<td>WDT</td>
<td>wh-determiner</td>
<td>which</td>
</tr>
<tr>
<td>WP</td>
<td>wh-pronoun</td>
<td>who, what</td>
</tr>
<tr>
<td>WPS</td>
<td>possessive wh-pronoun</td>
<td>whose</td>
</tr>
<tr>
<td>WRB</td>
<td>wh-abverb</td>
<td>where, when</td>
</tr>
</tbody>
</table>
Part of Speech (POS)

• Stanford parser: walk noun/verb
Part of Speech (POS)

- Stanford parser: *walk* noun/verb
Part of Speech (POS)

• Word sense disambiguation is more than POS tagging:

```
(ROOT
  (S
    (NP (PRP I))
    (VP (VBP walk)
      (PP (IN along)
        (NP (DT the) (NN river) (NN bank))))
    (.
.

(Root
  (S
    (NP (PRP I))
    (VP (VBP walk)
      (PP (IN along)
        (NP (DT the) (NN river) (NN bank))))
    (PP (TO to)
      (NP (PRP$ my) (NN bank)))))
    (.
.

```

different sense of the word bank
Syntax

• Words combine recursively with one another into phrases (aka constituents)
  • usually when two words combine, one word will head the phrase
    • e.g [VB/VBP eat] [NN chocolate]
    • e.g [VB/VBP eat] [DT some][NN chocolate]

Warning: terminology and parses in computational linguistics not necessarily the same as that used in linguistics
Syntax

• Words combine recursively with one another into phrases (aka constituents)
  • e.g. \([_{PRP} \text{we}]_{_{VB/VBP}} \text{eat} \] \([_{NN} \text{chocolate}]\)
  • e.g. \([_{TO} \text{to}]_{_{VB/VBP}} \text{eat} \] \([_{NN} \text{chocolate}]\)
Syntax

• Words combine recursively with one another into phrases (aka constituents)
  
e.g. [NNP John][VBD noticed][IN/DT/WDT that][PRP we][VBP eat] [NN chocolate]
Syntax

- Words combine recursively with one another into phrases (aka constituents)

How about SBAR?

*cf. John wanted me to eat chocolate*
Syntax

• Words combine recursively with one another into phrases (aka constituents)
  1. *John noticed that we eat chocolate*
  2. *John noticed we eat chocolate*