Adminstrivia

- Homework 7 out today
  - due Saturday by midnight
Another Grammar for \{a^n b^n c^n \mid n > 0\}

- \{ \ldots \} embeds Prolog code inside grammar rules
- \text{nonvar}(X) true if X is not a variable, false otherwise
- \text{var}(X) true if X is a variable, false otherwise

\begin{verbatim}
s  -->  a(N), b(N), c(N).
a(1)  -->  [a].
a(N)  -->  [a], a(M), \{N is M+1\}.
b(1)  -->  [b].
b(N)  -->  \{\text{nonvar}(N), N>1, M is N-1\}, [b], b(M).
b(N)  -->  \{\text{var}(N)\}, [b], b(M), \{N is M+1\}.
c(1)  -->  [c].
c(N)  -->  \{\text{nonvar}(N), N>1, M is N-1\}, [c], c(M).
c(N)  -->  \{\text{var}(N)\}, [c], c(M), \{N is M+1\}.
\end{verbatim}
Another Grammar for \( \{ a^n b^n c^n | n > 0 \} \)

- **Set membership:**
  ```prolog
  ?- s([a,a,b,b,c,c],[[]]).
  true ;
  false.
  ```

- **Set membership:**
  ```prolog
  ?- s([a,a,b,b,b,c,c],[[]]).
  false.
  ```

- **Enumeration:**
  ```prolog
  ?- s(String,[]).
  String = [a, b, c] ;
  String = [a, a, b, b, c, c] ;
  String = [a, a, a, b, b, b, c, c, c] ;
  String = [a, a, a, a, b, b, b, b, c, c, c, c] ;
  String = [a, a, a, a, a, b, b, b, b, b, c, c, c, c, c, c] .
  ```
Homework 7

• Question 1:
  – Give a **regular** DCG grammar for the language with strings containing:
    • an odd number of a's
    • and an even number of b's (even: 0,2,4,...)
    • assume alphabet is \{a,b\}
  – Examples:
    • aaa, ab, babaabb
    • *aab, *ab, *ε
  – Show your grammar working, i.e. it accepts and rejects positive and negative examples, respectively.
  – Allow rules of the form:
    • x --> [].
Homework 7

• Question 2:
  – Order your grammar rules so it enumerates strings in the language.
  – Show it working.
  – Is your grammar capable of enumerating all the strings of the language? Explain why or why not.
Question 3:

(a) Determine how many strings of size 3 are in the language? List them.

(b) Compute how many strings of size < 10 are there in the language.

Examples:

- \([b, b, a]\) is of size 3
- \([a, a, a, b, b, b, b]\) is of size 7
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)*
  – 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
(S
  (NP-SBJ Efforts)
  (NP Hong Kong Futures Exchange)
  (VP (VB introduce)
    (VP (VB continue)
      (S
        (NP-SBJ support)
        (SBAR (WHNP-1 proposed)
          (S
            (NP-SBJ the )
            (VP (VBZ enjoys)
              (NP (NONE- =Tx-1)
                (PP-LOC in
                  (NP (NONE- =pos-'s)
                    (JJ financial)
                    (NN community))))))))))
```

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.

Your query
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.

Tagging
Efforts/NNS by/IN the/DT Hong Kong/NNP Kong/NNP Futures/NNS Exchange/NNP to/TO introduce/VB a/DT new/JJ interest-rate/NN futures/NNS contract/NN continue/VBP to/TO hit/VB snags/NNS despite/IN the/DT support/NN the/DT proposed/VBN instrument/NN enjoys/VBZ in/IN the/DT colony/NN 's/PSS financial/JJ community/NN ./

Parse
(ROOT
  (S
    (NP (NNS Efforts))
    (PP (IN by)
      (NP (DT the) (NNP Hong) (NNP Kong) (NNS Futures) (NNP Exchange)))
    (S
      (VP (VBP to)
        (VP (VBD introduce)
          (NP (DT a) (JJ new) (NN interest-rate) (NN futures) (NN contract)))))))
    (VP (VBD continue)
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 ...

<table>
<thead>
<tr>
<th>part of speech</th>
<th>function or “job”</th>
<th>example words</th>
<th>example sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb</td>
<td>action or state</td>
<td>(to) be, have, do, like, work, sing, can, must</td>
<td>EnglishClub.com is a website. I like EnglishClub.com.</td>
</tr>
<tr>
<td>Noun</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>Adjective</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>Adverb</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>Pronoun</td>
<td>replaces a noun</td>
<td>I, you, he, she, some</td>
<td>Tara is Indian. She is beautiful.</td>
</tr>
<tr>
<td>Preposition</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>Conjunction</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>Interjection</td>
<td>short exclamation, sometimes inserted into a sentence</td>
<td>ohh, ouche, hlo, well</td>
<td>Ouch! That hurts! Hlo! 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

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.
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

In computational linguistics, the Penn Treebank tagset is the most commonly used tagset (*reprinted inside the front cover of your textbook*).

Penn Treebank Part-of-Speech Tags

<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>coordinating conjunction</td>
<td>and, but, or</td>
<td>SYM</td>
<td>symbol</td>
<td>+, %, &amp;</td>
</tr>
<tr>
<td>CD</td>
<td>cardinal number</td>
<td>one, two, three</td>
<td>TO</td>
<td>“to”</td>
<td>to</td>
</tr>
<tr>
<td>DT</td>
<td>determiner</td>
<td>a, the</td>
<td>UH</td>
<td>interjection</td>
<td>ah, oops</td>
</tr>
<tr>
<td>EX</td>
<td>existential “there”</td>
<td>there</td>
<td>VB</td>
<td>verb, base form</td>
<td>eat</td>
</tr>
<tr>
<td>FW</td>
<td>foreign word</td>
<td>men culpa</td>
<td>VBD</td>
<td>verb, preterite</td>
<td>ate</td>
</tr>
<tr>
<td>IN</td>
<td>preposition or subordinating conjunction</td>
<td>of, in, by</td>
<td>(past tense)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JJ</td>
<td>adjective</td>
<td>yellow</td>
<td>VBG</td>
<td>verb, gerund</td>
<td>eating</td>
</tr>
<tr>
<td>JJR</td>
<td>adj., comparative</td>
<td>bigger</td>
<td>VBN</td>
<td>verb, past participle</td>
<td>eaten</td>
</tr>
<tr>
<td>LS</td>
<td>list item marker</td>
<td>1, 2, One</td>
<td>VBZ</td>
<td>verb, 3rd person singular</td>
<td>eats</td>
</tr>
<tr>
<td>MD</td>
<td>modal</td>
<td>can, should</td>
<td>WDT</td>
<td>wh-determiner</td>
<td>which, that</td>
</tr>
<tr>
<td>NN</td>
<td>noun, singular</td>
<td>llama, snow</td>
<td>WP</td>
<td>wh-pronoun</td>
<td>what, who</td>
</tr>
<tr>
<td>NNS</td>
<td>noun, plural</td>
<td>llamas</td>
<td>WPS</td>
<td>possessive wh-word</td>
<td>whose</td>
</tr>
<tr>
<td>NNP</td>
<td>proper noun, singular</td>
<td>IBM</td>
<td>WRB</td>
<td>wh-adverb</td>
<td>how, where</td>
</tr>
<tr>
<td>NPS</td>
<td>proper noun, plural</td>
<td>Carolinas</td>
<td>S</td>
<td>dollar sign</td>
<td>$</td>
</tr>
<tr>
<td>PDT</td>
<td>preposition</td>
<td>all, both</td>
<td>$</td>
<td>pound sign</td>
<td>¢</td>
</tr>
<tr>
<td>POS</td>
<td>possessive ending</td>
<td>’s</td>
<td>”</td>
<td>left quote</td>
<td>‘ or “</td>
</tr>
<tr>
<td>PRP</td>
<td>personal pronoun</td>
<td>I, you, he</td>
<td>”</td>
<td>right quote</td>
<td>’ or ”</td>
</tr>
<tr>
<td>PRPS</td>
<td>possessive pronoun</td>
<td>your, one’s</td>
<td>(</td>
<td>left parenthesis</td>
<td>[, {, &lt;</td>
</tr>
<tr>
<td>RB</td>
<td>adverb</td>
<td>quickly, never</td>
<td>)</td>
<td>right parenthesis</td>
<td>], }, &gt;</td>
</tr>
<tr>
<td>RBR</td>
<td>adverb, comparative</td>
<td>faster</td>
<td>:</td>
<td>comma</td>
<td>,</td>
</tr>
<tr>
<td>RBR</td>
<td>adverb, superlative</td>
<td>fastest</td>
<td>;</td>
<td>sentence-final punctuation</td>
<td>: , . ,</td>
</tr>
<tr>
<td>RP</td>
<td>particle</td>
<td>up, off</td>
<td>:</td>
<td>mid-sentence punctuation</td>
<td>: . ,</td>
</tr>
</tbody>
</table>

45 tags listed in textbook
36 POS + 10 punctuation

Task: POS tagging
Part of Speech (POS)

- [Link](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>NNP</td>
<td>proper noun, singular</td>
<td>John</td>
</tr>
<tr>
<td>NNPS</td>
<td>proper noun, plural</td>
<td>Vikings</td>
</tr>
</tbody>
</table>
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>PDT</td>
<td>predeterminer</td>
<td><em>both the boys</em></td>
</tr>
<tr>
<td>POS</td>
<td>possessive ending</td>
<td><em>friend's</em></td>
</tr>
<tr>
<td>I</td>
<td>personal pronoun</td>
<td><em>I, he, it</em></td>
</tr>
<tr>
<td>PRP</td>
<td>personal pronoun</td>
<td><em>my, his</em></td>
</tr>
<tr>
<td>PRP$</td>
<td>possessive pronoun</td>
<td><em>my, his</em></td>
</tr>
<tr>
<td>RB</td>
<td>adverb</td>
<td><em>however, usually, naturally, here, good</em></td>
</tr>
<tr>
<td>RBR</td>
<td>adverb, comparative</td>
<td><em>better</em></td>
</tr>
<tr>
<td>RBS</td>
<td>adverb, superlative</td>
<td><em>best</em></td>
</tr>
<tr>
<td>RP</td>
<td>particle</td>
<td><em>give up</em></td>
</tr>
<tr>
<td>SYM</td>
<td>Symbol</td>
<td><em>$ , %</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)

- [Link](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>UH</td>
<td>interjection</td>
<td>uhhuhuhhh</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>WP$</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 (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 \([\text{VB/VBP eat}] [\text{NN chocolate}]\)
    • e.g \([\text{VB/VBP eat}] [\text{DT some}] [\text{NN chocolate}]\)

**Warning:**
eterminology 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 we][VB/VBP eat] [NN chocolate]
  • e.g. [TO to][VB/VBP eat] [NN chocolate]
Syntax

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

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

```
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