Organisational

- Read Section 5.1 to 5.6
- Do Exercises 1-3 in section 5.9
- Familiarise yourselves with PTB Guidelines
- Submit Assignment 1 (POS-tagging) on Monday 12noon
- Work through logic worksheet
- Today: Assignment 2 (Phrase structure analysis of NPs)
• Tokenization
• Ambiguity
• Idioms
• Multi Word Units
• Finite vs. non-finite forms of the verb
• Use vs Mention: quoted material, titles, ...
• Discuss cases of uncertainty
Particles vs. Prepositions

- The man up the ladder fell.
- Kim ran up the stairs.
- Kim ran up a large bill.
- Kim slipped up.
- Kim washed up the dishes.
- Kim washes the dishes up.
## Lexical Features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Values</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num(ber)</td>
<td>Sg / Pl</td>
<td>boy(+s)</td>
</tr>
<tr>
<td>N-type</td>
<td>Mass, Count, Name</td>
<td>boy, research *research+s, Fred</td>
</tr>
<tr>
<td>Per(s)on</td>
<td>1,2,3</td>
<td>I (1sg), you (2sg) (s)he (3sg)</td>
</tr>
<tr>
<td>Case</td>
<td>Nom, Acc</td>
<td>he (nom), him (acc)</td>
</tr>
<tr>
<td>Valence</td>
<td>Intrans, Trans, Ditrans, Scomp,...</td>
<td>smile, kiss, give, believe,...</td>
</tr>
<tr>
<td>A-type</td>
<td>base / comparative / superlative</td>
<td>– old, older, oldest</td>
</tr>
<tr>
<td>Tag</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>CC</td>
<td>coordinating</td>
<td><em>and, but, or</em></td>
</tr>
<tr>
<td>CD</td>
<td>cardinal number</td>
<td><em>one, two</em></td>
</tr>
<tr>
<td>DT</td>
<td>determiner</td>
<td><em>a, the</em></td>
</tr>
<tr>
<td>EX</td>
<td>existential ‘there’</td>
<td><em>there</em></td>
</tr>
<tr>
<td>FW</td>
<td>foreign word</td>
<td><em>mea culpa</em></td>
</tr>
<tr>
<td>IN</td>
<td>preposition/</td>
<td><em>of, in, by</em></td>
</tr>
<tr>
<td></td>
<td>subordin-conj</td>
<td></td>
</tr>
<tr>
<td>JJ</td>
<td>adjective</td>
<td><em>yellow</em></td>
</tr>
<tr>
<td>JJR</td>
<td>comparative adj</td>
<td><em>bigger</em></td>
</tr>
<tr>
<td>JJS</td>
<td>superlative adj</td>
<td><em>wildest</em></td>
</tr>
<tr>
<td>LS</td>
<td>list item marker</td>
<td><em>1, 2, One</em></td>
</tr>
<tr>
<td>MD</td>
<td>modal</td>
<td><em>can, should</em></td>
</tr>
<tr>
<td>NN</td>
<td>sing or mass noun</td>
<td><em>llama</em></td>
</tr>
<tr>
<td>NNS</td>
<td>noun, plural</td>
<td><em>llamas</em></td>
</tr>
<tr>
<td>NNP</td>
<td>proper noun, sing.</td>
<td><em>IBM</em></td>
</tr>
<tr>
<td>NNPS</td>
<td>proper noun, plu.</td>
<td><em>Carolinas</em></td>
</tr>
</tbody>
</table>

**Figure 8.1** Penn Treebank part-of-speech tags (including punctuation).
Eh?

From the film “The Martian”.

I'm gonna have to science the shit out of this.

(An instance of productive derivational morphology (zero-derivation).)
Moving on...

- Constituency
- Phrase structure Grammar
- Phrase structure Trees
- Tests for Constituency
Linguistic Methodology

• Descriptive, not prescriptive
• Example: should we (or should we not) split infinitives?
  • Captain Kirk has gone beyond our galaxy
  • Captain Kirk’s mission is to go beyond our galaxy
  • Captain Kirk has been travelling the universe for 30 years
Linguistic Methodology

- Descriptive, not prescriptive
- Example: should we (or should we not) split infinitives?
  - Captain Kirk (boldly) has (boldly) gone beyond our galaxy
  - Captain Kirk’s mission is (boldly) to (boldly) go beyond our galaxy
  - Captain Kirk (boldly) has (boldly) been (boldly) travelling (*boldly) the universe for 30 years
- Which rule can we derive from this?
Distributional analysis

- Algorithm:
  - Create a template
  - Perform substitutions
  - Test for grammaticality
  - Ungrammaticality, semantic oddness/implausibility

- The _____ can run
- _____ can run
- collect an equivalence class of strings that can go in the slot.

- They are called Constituents

- This is an entirely bottom-up approach championed in the 1930s

- Today, two different dominant approaches:
  - Generative grammar (phrase structure)
  - Headedness (relational)
Generative Methodology

- Noam Chomsky (1957): Syntactic Structures
- Finite sets of rules predict all and only grammatical sentences
- Generative grammars: mappings between sentences and meaning
A context-free grammar

Rules:

S → NP VP
VP → VP PP
VP → V
VP → V NP
VP → V VP
NP → NP PP
PP → P NP

Lexicon:

V → can
V → fish
NP → fish
NP → they
NP → rivers
NP → pools
NP → December
NP → Scotland
NP → it
P → in
Headedness

- The heads of NPs are nouns.
- Heads are the only constituent of a constituent that cannot be dropped:
  - The castle is old
  - *The is old
  - The big castle is old
  - *The big is old
  - The castle by the hill is old
  - *The by the hill is old
  - Castles are interesting
Phrase Marker Trees

```
S[decl]
  /   \
NP    VP
  |    |
Det   V[trans,pl]
 |     |
the   love
 
  /   \
NP    NP
  |    |
N[pl]  N[name]
 |     |
people Sandy
```
Constituency Tests

Substitution test

- use “proform” (e.g. “do so” stands in for a VP; e.g. “that” stands in for an NP)
- If substitution is felicitous, then phrase is a constituent (of same category as the proform).
- What are other NPs (like “the people”)?
- What are other transitive verbs (like “love”)?
Constituency Tests

Movement test
- Constituents can be moved around in the sentence.
  - The old man has come to dinner.
  - Has the old man come to dinner?
  - *The has old man come to dinner
Constituency Tests

Insertion test

- Appositions are parentheticals.
- They cannot be inserted into constituents, only at the end of constituents.
  - The President of America, Ronald Reagan, is over 70.
  - *The President, Ronald Reagan, of America is over 70.
  - *The President of America is, Ronald Reagan, over 70.
Constituency Tests

Omissibility test (only suitable for some constituent types)

- Some constituents can be omitted
- Non-constituents cannot be omitted
  - Some friends of the old man came to dinner.
  - Some friends came to dinner.
  - *Some friends man came to dinner.
Constituency Tests

**Coordination test** (well-known exceptions)

- Constituents of the same type can be coordinated
  - Kim and Sandy kissed each other
  - The old men and women came to dinner
  - The old man and his young nephew came to dinner
  - Kim and Sandy divorced and remarried each other
  - Kim kissed Sandy and remarried her
  - That rather old and very unreliable car belongs to Kim
  - Kim washed up and Sandy watched the TV
Problems for Coordination Test

- Kim is a conservative and proud of it
- Kim became a conservative and arrogant
- Kim enjoys chess and watching football
- Kim gave Sandy a pen and Fido a bone
- "To hell with them and be dammed", he said.
Assignment 2

- Perform a phrase structure analysis of all noun phrases in your chosen sentences (the same ones from assignment 1)
- First bracket all NPs
- Recursively embedded
- Draw a Phrase Structure tree for each NP
- Reuse your Tokenisation and POS analysis from assignment 1
- Submit by Monday 28 October
My aunt's can opener
Noun compounds
Reading for next time

- Read 5.7 and 5.8
- We will discuss exercises in 5.9 (1-3), so if you haven't done them, another chance.
- Keep working on logic worksheet