What’s in the Penn Treebank

OSU Mini-Institute
(slides adapted from Erhard Hinrichs)
1. CC Coord. conjunc.  
2. CD Cardinal number  
3. DT Determiner  
4. EX Existential there  
5. FW Foreign word  
6. IN Prep./subord. conj.  
7. JJ Adject.  
8. JJR Adject., comp.  
9. JJS Adject., superl.  
10. LS List item marker  
11. MD Modal  
25. TO to  
26. UH Interjection  
27. VB V, base form  
28. VBD V, past tense  
29. VBG V, gerund/pres. part.  
30. VBN V, past part.  
31. VBP V, non-3rd ps. sing. pres.  
32. VBZ V, 3rd ps. sing. pres.  
33. WDT wh-det.  
34. WP wh-pronoun  
35. WP$ Poss. wh-pronoun
<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.</td>
<td>NN Noun, sing. or mass</td>
<td></td>
<td>36.</td>
</tr>
<tr>
<td>13.</td>
<td>NNS Noun, plural</td>
<td></td>
<td>37.</td>
</tr>
<tr>
<td>14.</td>
<td>NNP Proper noun, sing.</td>
<td></td>
<td>38.</td>
</tr>
<tr>
<td>16.</td>
<td>PDT Predeterminer</td>
<td></td>
<td>40.</td>
</tr>
<tr>
<td>17.</td>
<td>POS Possessive ending</td>
<td></td>
<td>41.</td>
</tr>
<tr>
<td>18.</td>
<td>PRP Personal pronoun</td>
<td></td>
<td>42.</td>
</tr>
<tr>
<td>20.</td>
<td>RB Adverb</td>
<td></td>
<td>44.</td>
</tr>
<tr>
<td>21.</td>
<td>RBR Adverb, comp.</td>
<td></td>
<td>45.</td>
</tr>
<tr>
<td>22.</td>
<td>RBS Adverb, superl.</td>
<td></td>
<td>46.</td>
</tr>
<tr>
<td>23.</td>
<td>RP Particle</td>
<td></td>
<td>47.</td>
</tr>
<tr>
<td>24.</td>
<td>SYM Symbol</td>
<td></td>
<td>48.</td>
</tr>
</tbody>
</table>
Declarative sentences

(S (NP-SBJ I)
  (VP consider
    (S (NP-SBJ Kris)
      (NP-PRD a fool))))

- SBJ and PRD mark grammatical functions
- “Kris a fool” is a constituent, labeled with S.
(SQ Was
(NP–SBJ he)
(ADVP–TMP ever)
(ADJP–PRD successful)
?)

• SQ marks sentence type

• TMP is “temporal”
WH-question

(SBARQ
  (WHNP-1 What
   (SQ is
    (NP-SBJ Tim)
    (VP eating
     (NP *T*-1)))))

• SBARQ marks sentence type
• WHNP marks question word
• WHNP-1 coindexes it with trace.
Passives

(S
  (NP-SBJ-1 The ball
   (VP was
    (VP thrown
     (NP -*1)
     (PP by
      (NP-LGS Chris))))))

• Surface subject marked by SBJ
• Logical subject by LGS
• Trace after verb
Control verbs

(S
  (NP-SBJ-1 Chris)
  (VP wants
    (S
      (NP-SBJ *-1)
      (VP to
        (VP throw
          (NP the ball)))))))

• Similar idea to passive.

• Uses same tools
Sentence types

- **S** — Simple declarative clause, i.e. one that is not introduced by a subordinating conjunction or *wh*-word and that does not exhibit subject-verb inversion.

- **SBAR** — Clause introduced by a subordinating conjunction.

- **SBARQ** — Direct question introduced by a *wh*-word or *wh*-phrase.

- **SINV** — Inverted declarative sentence, i.e. one in which the subject follows the tensed verb or modal.

- **SQ** — Inverted yes/no question, or main clause of a *wh*-question, following the *wh*-phrase in SBARQ.
Grammatical roles

- BNF (benefactive)
- DTV (dative)
- LGS (logical subject)
- PRD (predicate)
- PUT (locative PP of put)
- SBJ (surface subject)
- TPC ("topicalized")
- VOC (vocative)
- DIR (direction)
- EXT (extent)
- LOC (locative)
- MNR (manner)
- PRP (purpose or reason)
- TMP (temporal)
- CLR (closely related)
- CLF (cleft)
- HLN (headline)
- TTL (title)
DTV

(S
  (NP-SBJ Aristotle)
  (VP dedicated
    (NP the book)
    (PP-DTV to
      (NP Plato)))))

- Marks the argument relation in dative double object construction.
Extraposition

(S (NP-SBJ Plato)
  (VP knew
    (SBAR *ICH*-1)
    (NP-TMP yesterday)
    (SBAR-1 that
      (S (NP-SBJ Terry)
        (VP would
          (VP accept
            (NP the honor)))))))

- Proliferation of empty elements
Attachment ambiguity

(S
(NP-SBJ I)
  (VP saw
    (NP (NP the man)
      (PP *PPA*-1))
    (PP-CLR-1 with
      (NP the telescope))))

• *PPA* to mark “Permanent predictable ambiguity”

• CLR to mark “closely related” (Hard to decide whether is an argument or an adjunct, so don’t)
Empty elements

*T* (trace of A’-movement, including parasitic gaps)
(NP *) (arbitrary PRO, controlled PRO, and trace of A-movement)
0 (null complementizer, including null *wh*-operator)
*U* (unit)
*?* (placeholder for ellipsed material)
*NOT* (anti-placeholder in template gapping)
*RNR* (pseudo-attach: right node raising)
*ICH* (pseudo-attach: interpret constituent here)
*EXP* (pseudo-attach: expletive)
*PPA* (pseudo-attach: permanent predictable ambiguity)