

# OntoNotes English Treebanking – Y3

2009-02-24



# Treebanking – Year 3 (Ann Taylor)



<b>Newly Parsed (new NML &amp; hyphen styles)</b>	
<b>Eng-from-Ara Web</b>	55,279
<b>Eng-from-Chi Web</b>	41,694
<b>P2.5</b>	140,322
<b>TOTAL</b>	<b>237,295</b>

<b>Style Updated (Hyphens and/or NMLs)</b>			<b>TOTAL</b>
<b>BC</b>	151,267	Hyphens	<b>428,499</b>
<b>BN</b>	220,174	Hyphens	
<b>ECTB</b>	57,058	Hyphens	
<b>WSJ300</b>	352,957	Hyphens & NMLs	<b>889,409</b>
<b>WSJ400</b>	536,452	Hyphens & NMLs	

# Intra-*Annotator* Treebanking Experiment



- **Q: How consistent is human syntactic annotation?**
- **Experiment:**
  - Ann Taylor selected at random 23 files from BN she annotated in Y2 (6K words total).
  - The texts were re-parsed with a parser trained only on WSJ.
    - (7 sentences were skipped, due to retokenization during first annotation pass, yielding 223 re-parsed sentences, 4200 brackets.)
  - Ann re-corrected the texts.
  - *evalb* was run on new texts, given old texts as gold standard.
  - Separately, Ann used *sdiff* to find & examine all differences

# Results of *evalb* on all sentences

---



- **Bracketing Recall: 98.4**
- **Bracketing Precision: 98.5**
- **Tagging Accuracy: 98.7**
  - (Ann corrects wrong tags as she goes.)

# Mis-tagging seems to cause mis-parsing

---



- **No tagging inconsistencies: 182 sentences**
  - No bracketing inconsistencies of these: 161 (88%)
  
- **Tagging inconsistencies: 41 sentences**
  - No bracketing inconsistencies of these: 26 (63%)

# Error Analysis using *sdiff*



- 146 sentences categorized for inconsistencies
- 176 disagreements total
  - Many of these aren't *evalb* errors
- For 52 of these, Ann not sure which was “right”
- Primary cause of disagreements (some multiple)
  - POS Tag differences: 70\*
  - Dash-tag differences: 35\*
  - Attachment differences: 23
  - Coindexing differences: 11\*
  - Node label different: 9
  - Extra node: 8
  - Other: 8
  - Missing node: 7
  - Empty category: 5\*

\*: Not an *evalb* error