

Searching the Penn Treebank with **Tregex**

Answer Sheet

Corpus-Based Computational Linguistics
(Chris Brew and Michael White)

2008 Linguistics Summer Mini-Institute

1. Query: `DT < the $ JJR`. The query matches 9 sentences, but no CCs.
2. Query: `__ < (DT < the) <- JJR`. There are 26 matching trees in total, many of which do indeed involve CCs.
3. Query: `/^X.* / < DT << JJR|RBR`. This turns up 22 matching sentences.
4. Query: `/^X.* / < DT << JJR|RBR !$ (__ << /VB.* /)`. The query turns up 7 trees, 4 of which do not involve parallel copula deletion!