

Visualizing Tree Queries

CS 294-10 Final Project Proposal
Mason Smith
March 31, 2010

Background

- Penn Treebank
 - syntactically-annotated parse trees

```
( (S (CC But)
  (ADVP-TMP (RB so) (RB far) )
  (, ,)
  (S
    (NP-SBJ (DT the) (NN effort) )
    (VP (VBZ has)
      (VP (VBN failed) )))
  (, ,)
  (CC and)
  (S
    (NP-SBJ (NN traffic) )
    (VP (VBZ is)
      (ADVP-TMP (RB still) )
      (ADJP-PRD (JJ slow) )))
  (. .) ))
```


Motivation

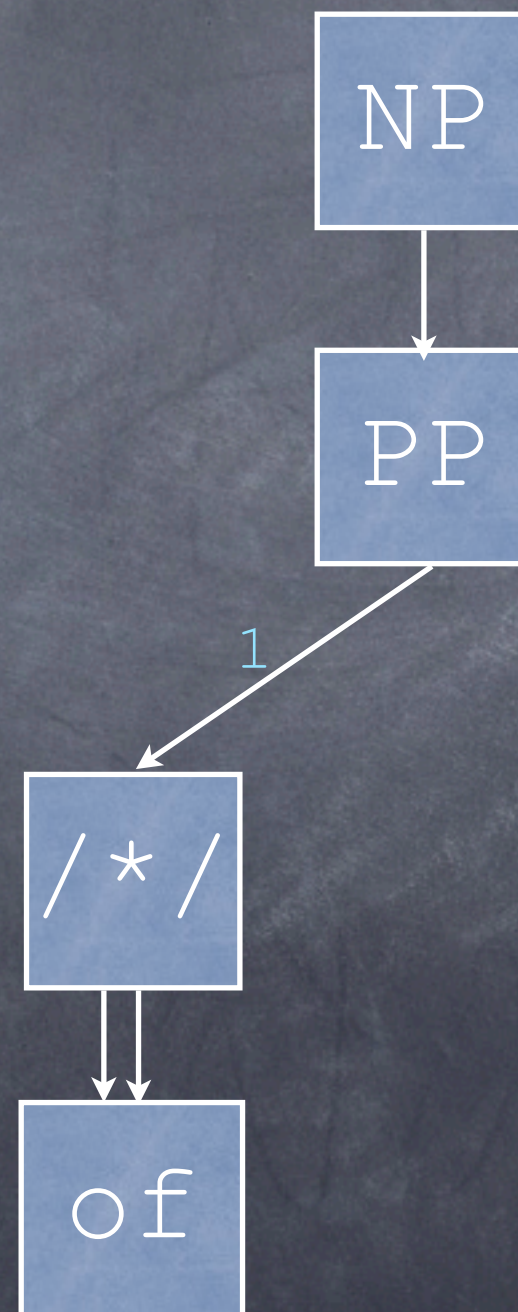
- Example query: Find all instances of an NP immediately dominating a PP starting with 'of'
- Difficult (impossible?) with regular expressions

Previous Work

- Current text-based tools (tgrep/tgrep2 and tregex) are specially designed for this task
- Previous example: NP << (PP <1 (/*/
<< /of/))
- Complicated queries get much harder to read/write

Proposal

- Graphical tool for constructing queries of parse trees
 - Point-and-click interface for query constructing
 - Translation to (and from?) text-based version



Milestones

- Interface with existing tools (Java-based tregex or C-based tgrep2)
- Display parse trees
- Flesh out visual representation of query language
 - possibly limit to a manageable subset
- Implement visual -> text translator

Questions?

Comments?