Constituency Parsing
Syntax

- Study of word order and how words form sentences
- Why do we care about syntax?
  - Multiple interpretations of words (noun or verb? Fed raises... example)
  - Recognize verb-argument structures (who is doing what to whom?)
  - Higher level of abstraction beyond words: some languages are SVO, some are VSO, some are SOV, parsing can canonicalize
Constituency Parsing

- Tree-structured syntactic analyses of sentences

- *Constituents*: (S)entence, (N)oun (P)hrases, (V)erb (P)hrases, (P)repositional (P)hrases, and more

- Bottom layer is POS tags

- Examples will be in English. Constituency makes sense for a lot of languages but not all
ternary rule

unary rule

sentential complement

whole embedded sentence

adverbial phrase
a refund that the court estimated *-1
If we do not annotation, these trees differ only in one rule:

$\text{VP} \rightarrow \text{VP} \text{PP}$

$\text{NP} \rightarrow \text{NP} \text{PP}$

Parse will go one way or the other, regardless of words.

Lexicalization allows us to be sensitive to specific words.

**Challenges**

**PP attachment**

same parse as “the cake with some icing”
Challenges

Modifier scope:

```
  NP
   NP
    JJ   NN   NN
    plastic cup holder
```

```
  NP
   NP
    JJ   NN   NN
    plastic cup holder
```

Complement structure:

The students complained to the professor that they didn’t understand

Coordination scope:

The man picked up his hammer and saw

compare: The man picked up his hammer and swung

[Eisenstein book]
How do we know what the constituents are?

Constituency tests:

- Substitution by *proform* (e.g., pronoun, *did so*)
- Clefting (*It was with a spoon that...*)
- Answer ellipsis (*What did they eat? the cake*)
  (How? *with a spoon*)
- Sometimes constituency is not clear, e.g., coordination: *she went to and bought food at the store*