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
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**

Complement structure:

\[ \text{The man picked up this hammer and saw} \]

Coordination scope:

\[ \text{The man picked up his hammer and saw} \]

Compare: The man picked up his hammer and swung

[Eisenstein book]

- **Modifier scope:**

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]