

# A comparison of Charniak and Stanford parsers

Prashanth K

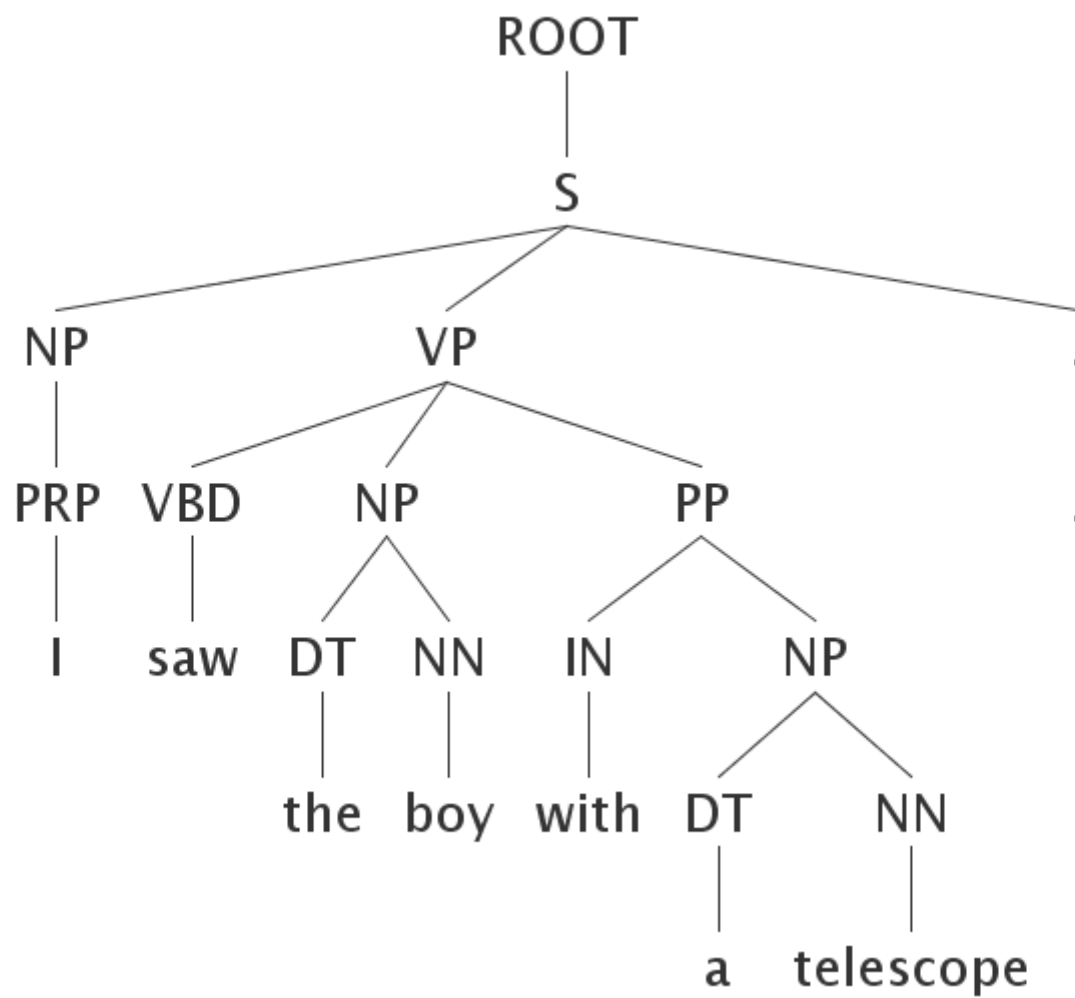
Under the guidance of Prof. Pushpak Bhattacharyya

Department of Computer Science and Engineering  
IIT Bombay

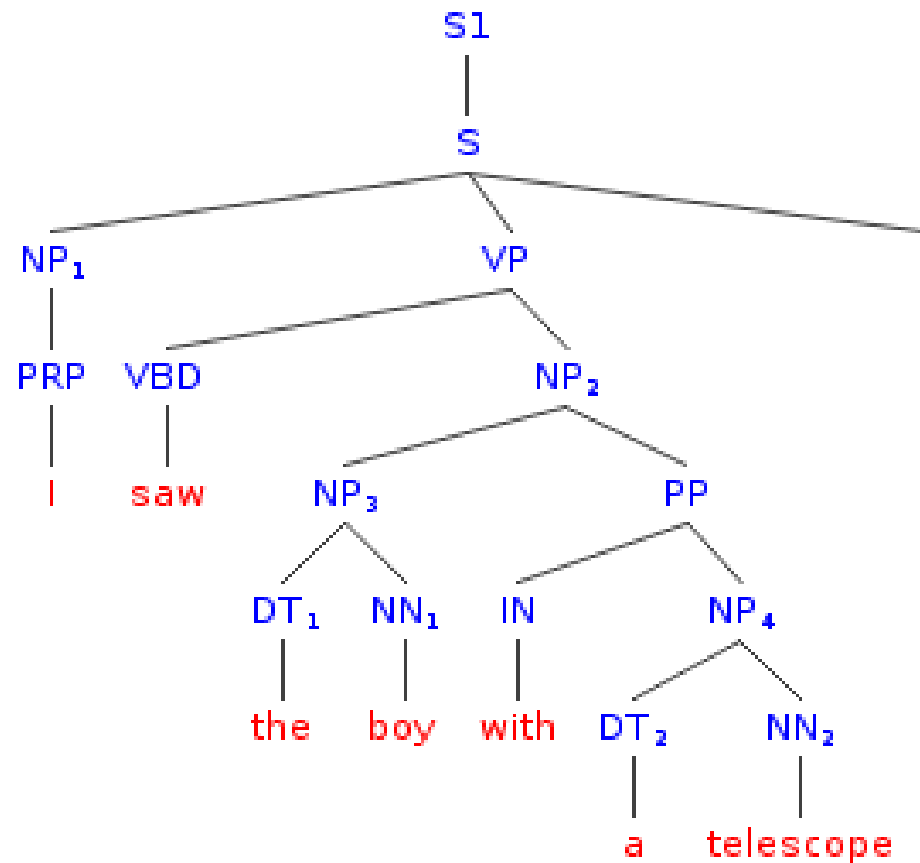
# Sentence with multiple parses

I saw the boy with a telescope

# Stanford



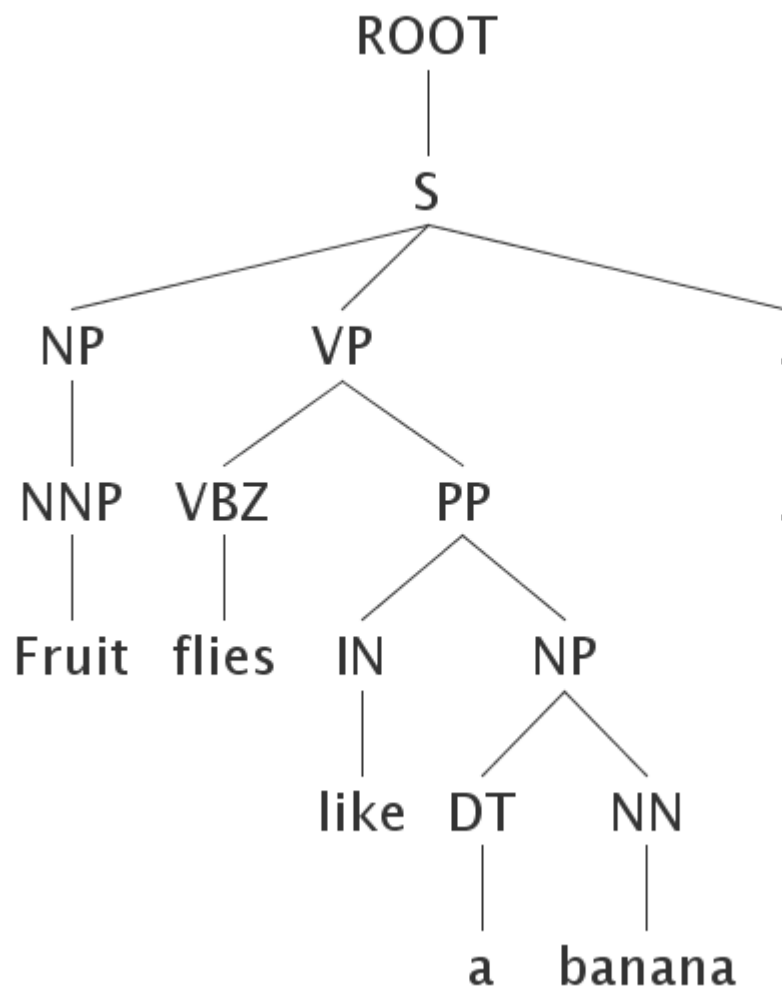
# Charniak



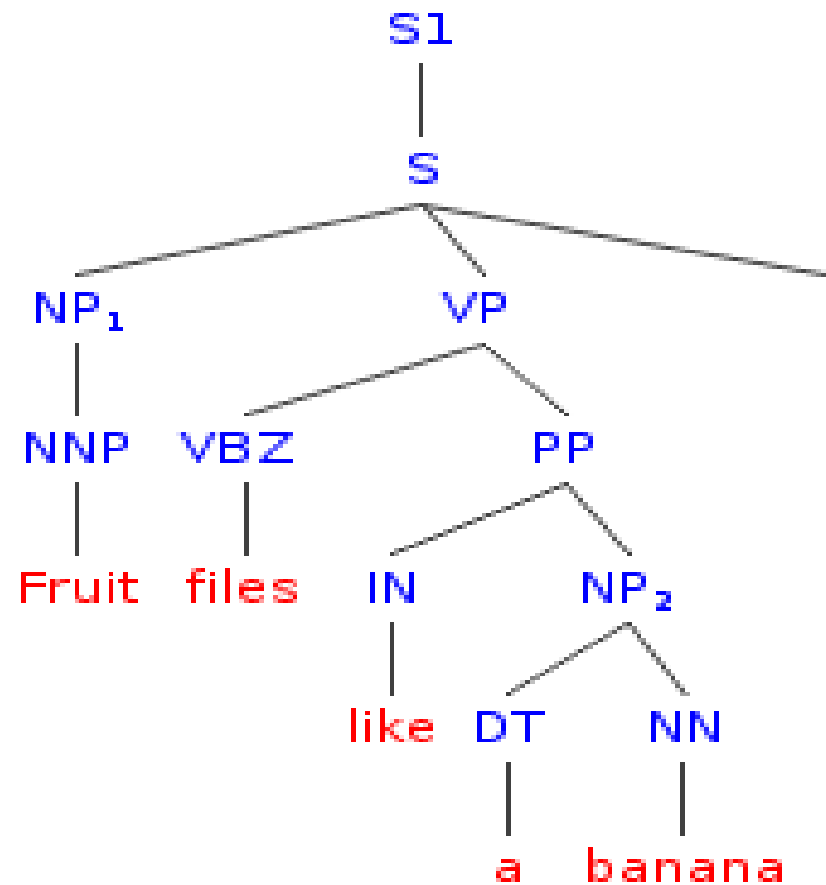
# Noun-verb confusing sentence

Fruit flies like a banana.

# Stanford



# Charniak

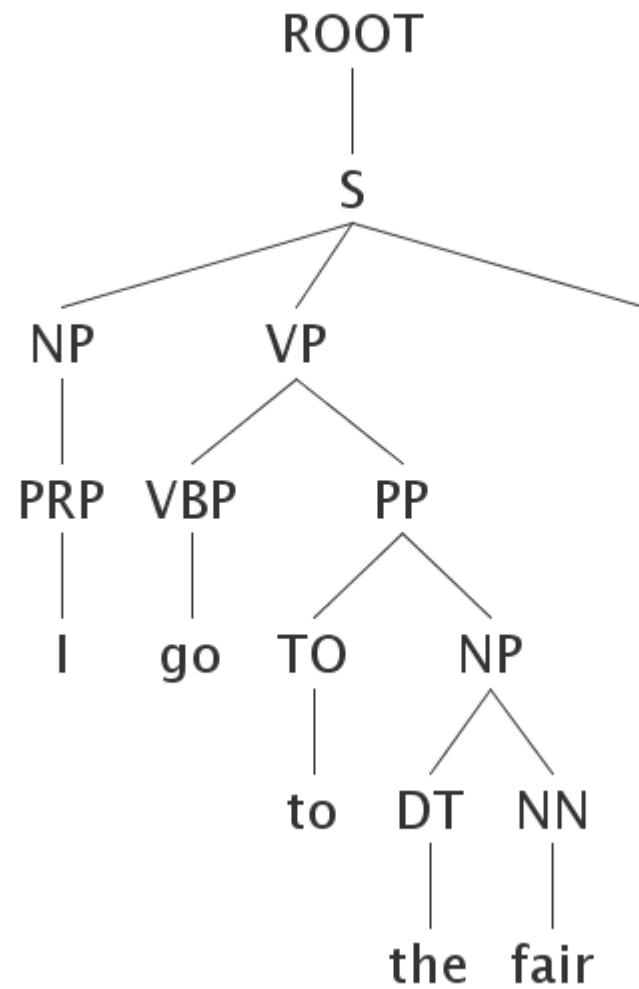


# Normal sentence

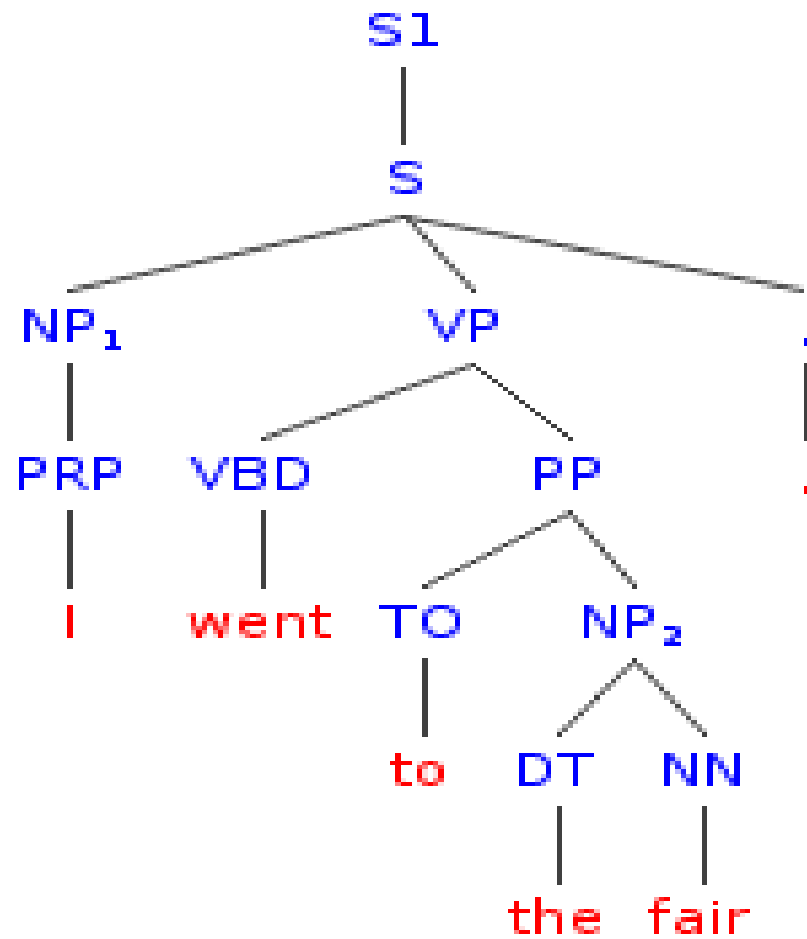
I go to the fair.



# Stanford



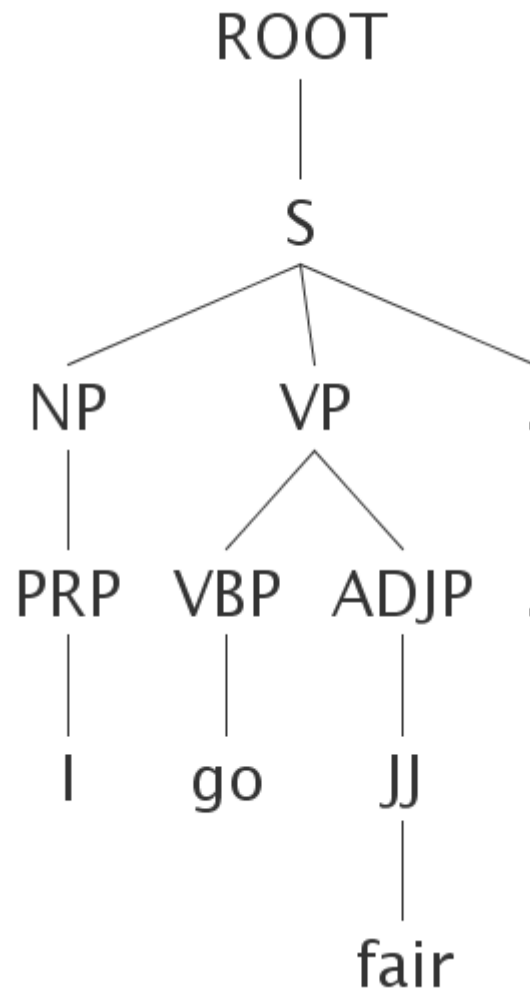
# Charniak



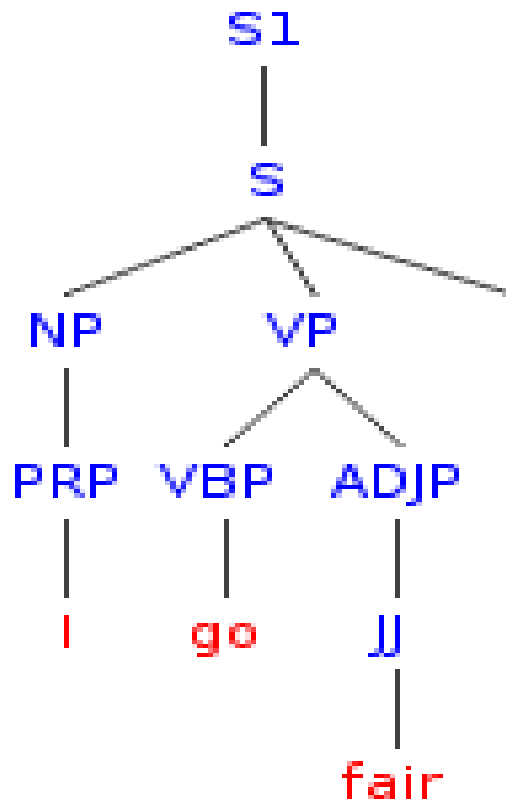
# Grammatically incorrect sentence

I go fair.

# Stanford

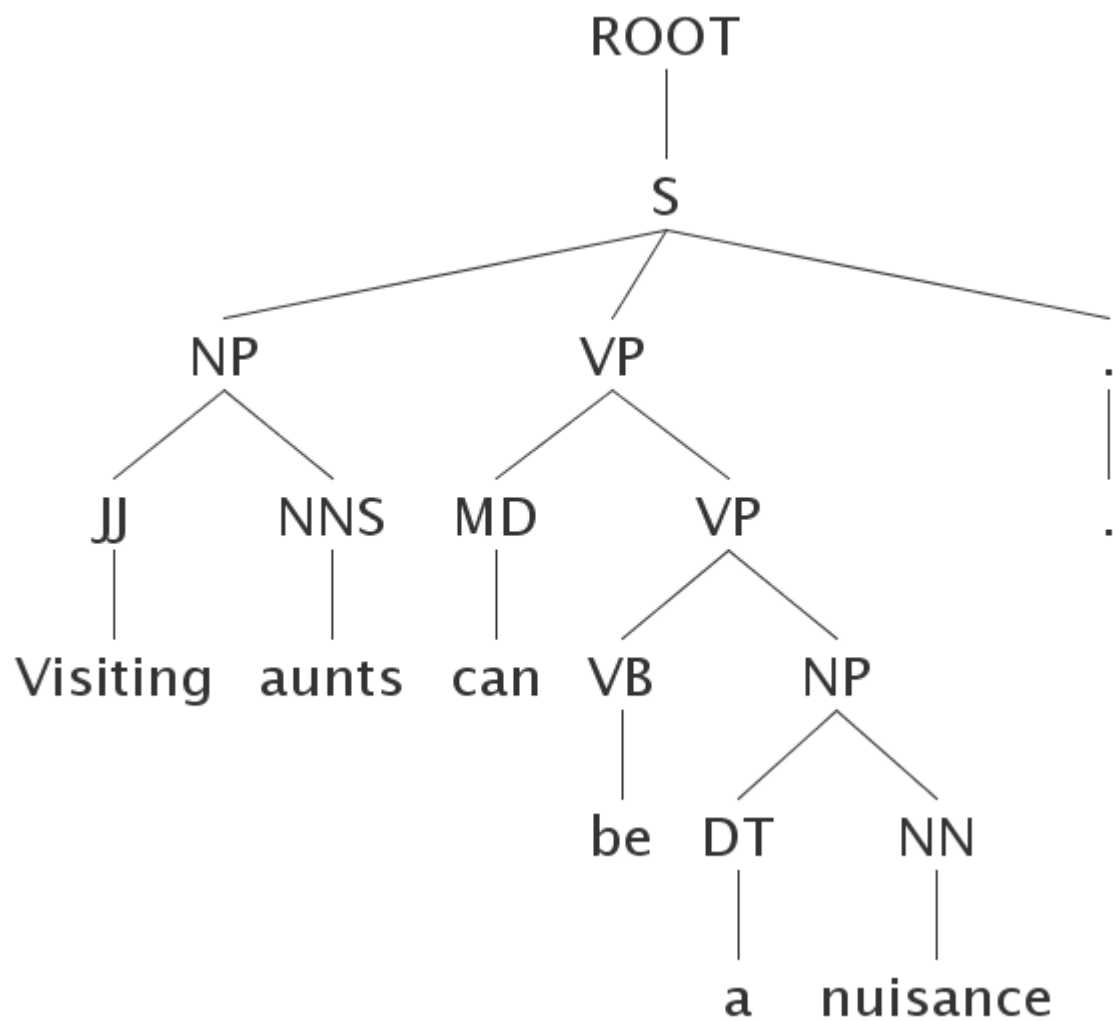


# Charniak

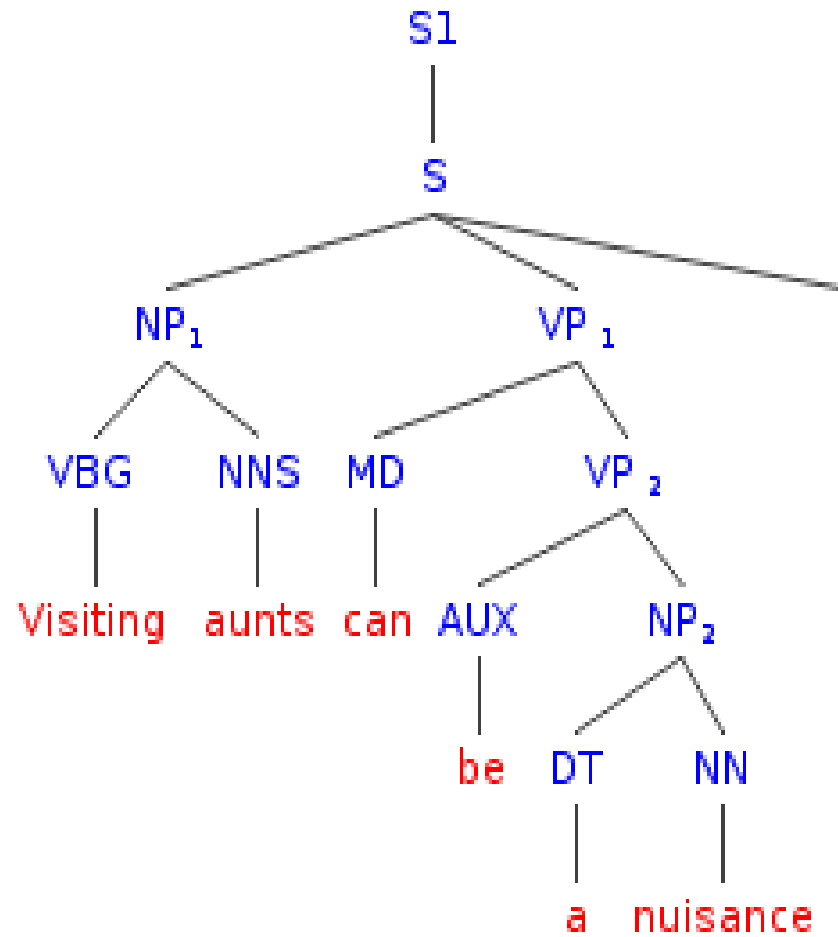


Visiting aunts can be a nuisance.

# Stanford



# Charniak

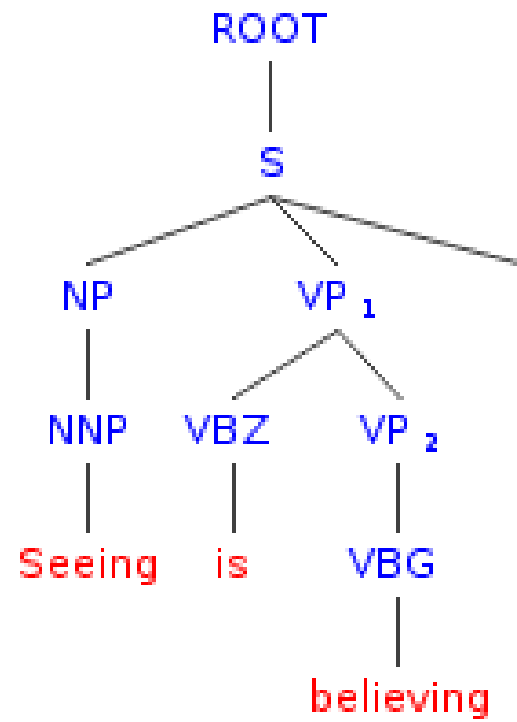




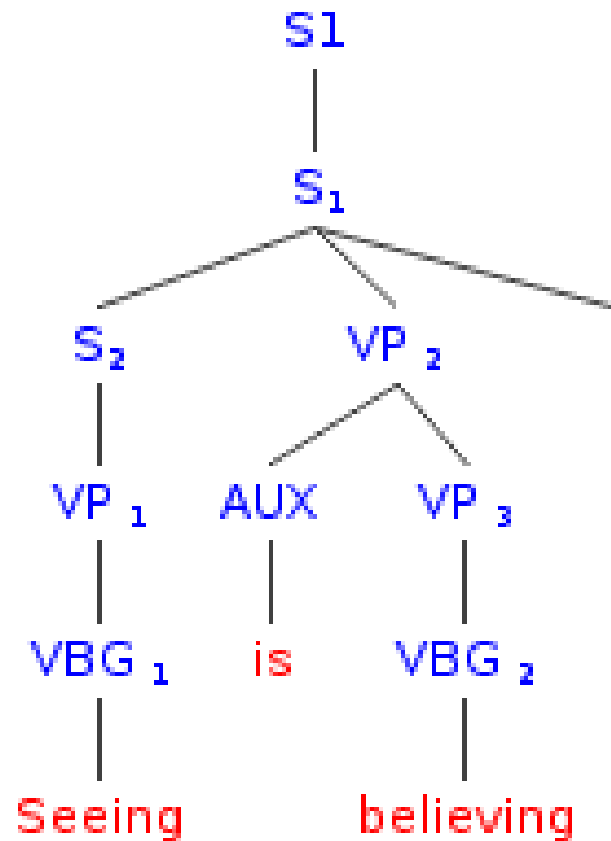
# Multiple POS

Seeing is believing.

# Stanford



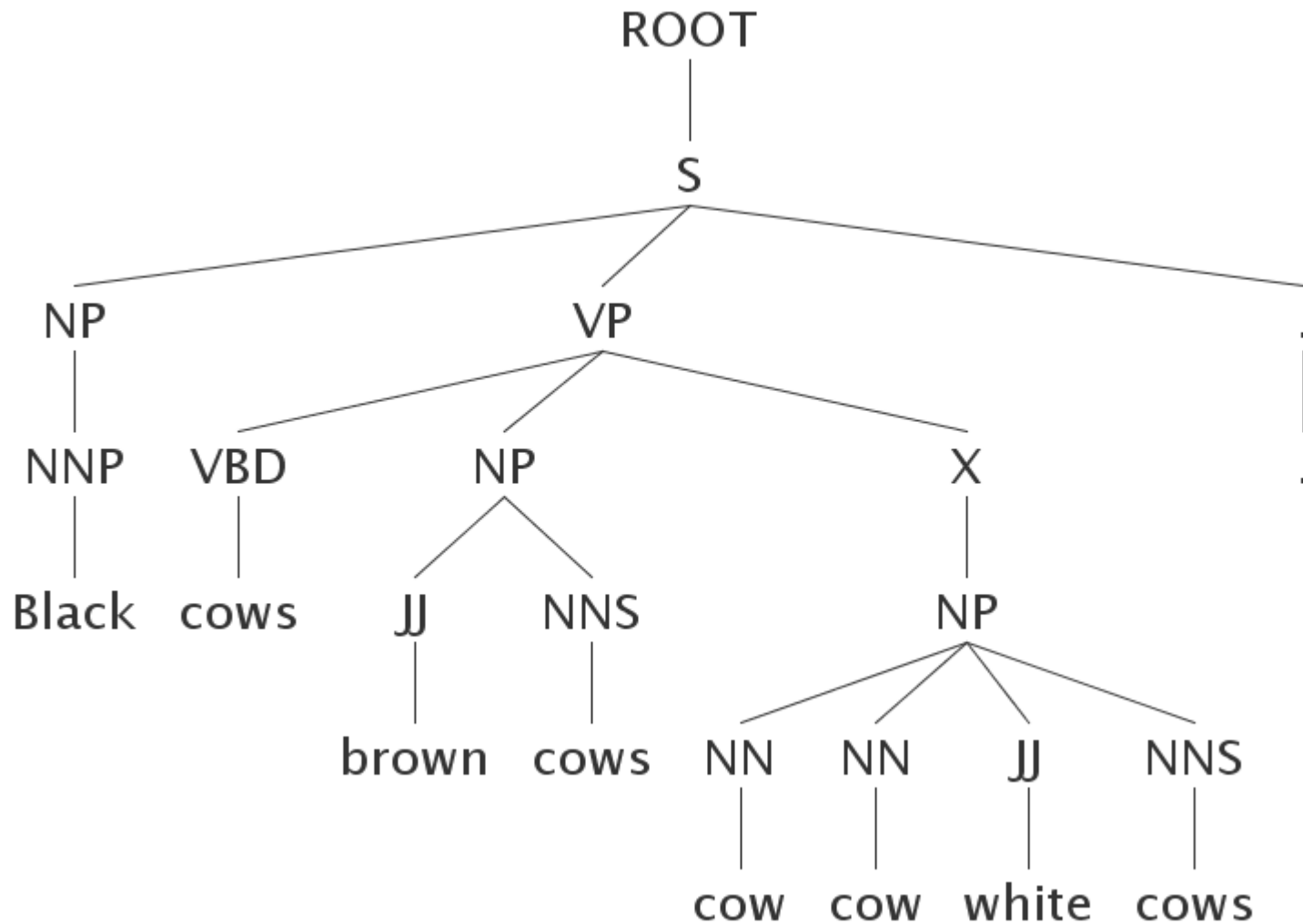
# Charniak



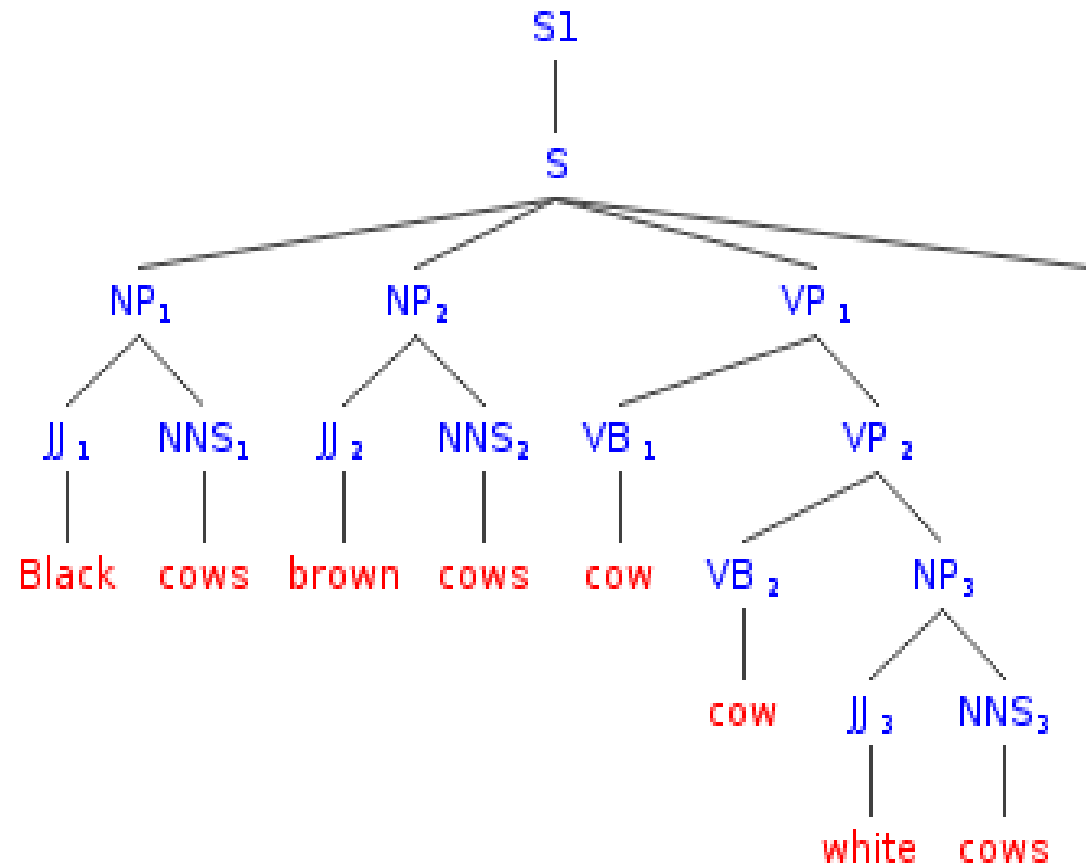
# Repeated words with changing POS

Black cows brown cows cow cow white cows.

# Stanford



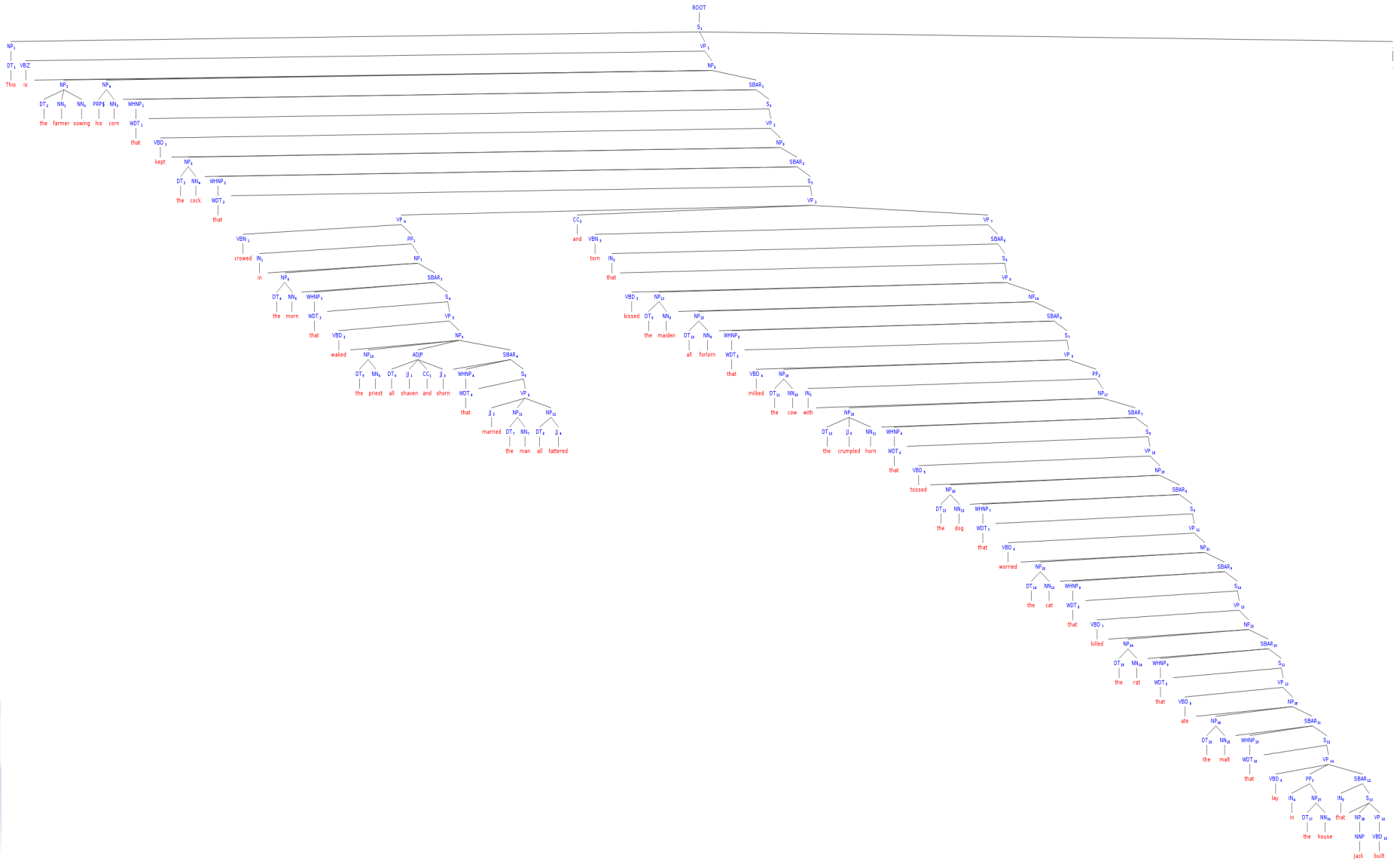
# Charniak



# Highly embedded sentence

This is the farmer sowing his corn that kept the cock that crowed in the morn that waked the priest all shaven and shorn that married the man all tattered and torn that kissed the maiden all forlorn that milked the cow with the crumpled horn that tossed the dog that worried the cat that killed the rat that ate the malt that lay in the house that Jack built.

# Stanford







# Long sentence (202 words)

- Stanford
  - Out of heap memory
- Charniak
  - No result

# Time

- 106 sentences, 1695 words
- Stanford
  - real 0m44.046s
  - user 0m48.525s
  - sys 0m0.600s
- Charniak
  - real 0m50.435s
  - user 0m52.943s
  - sys 0m0.505s

# Observations

- Charniak gives better parses but is slower
- Stanford parser gives dependency parse, and can also give top N parses of sentences
- Both parsers fail on long sentences
- Both parsers can handle highly embedded sentences
- Charniak's POS tagger seems to be more precise than Stanford's

# Observations (contd.)

- Where Stanford parser goes wrong
  - It assumes that every sentence is of the form  $S \rightarrow NP VP$  .
- Charniak parser
  - $S \rightarrow NP NP VP$  .
  - $S \rightarrow S VP$  .
  - etc.