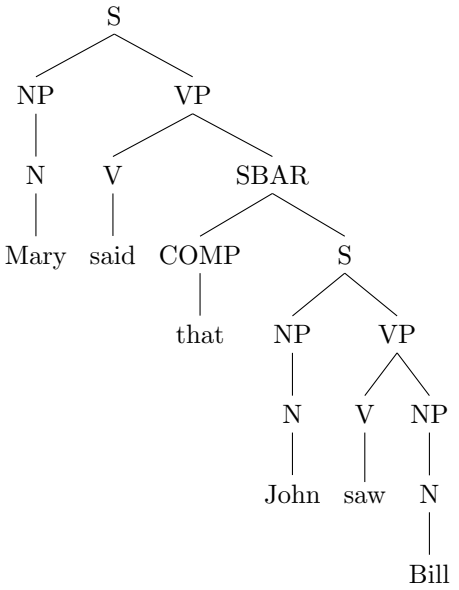
1 Lexicalization of a Treebank

1.1 Question (time: 10:44, slide: 8)

Say we have the sentence "Mary said that John saw Bill" with the parse tree



We are also given the head rules (where $*$ indicates the head)

S $\rightarrow$ NP **VP*** NP $\rightarrow$ **N*** VP $\rightarrow$ **V*** NP

VP $\rightarrow$ **V*** SBAR SBAR $\rightarrow$ **COMP*** S

List the head words (separated by a space) of the following non-terminals

1. the "SBAR"
2. the "S" spanning "Mary ... Bill"
3. the "VP" spanning "said... Bill"

2 Lexicalized PCFGs

2.1 Question (time: 4:54, slide: 11)

Say we are constructing a lexicalized PCFG with $|N| = 10$ and $|\Sigma| = 1000$. How many possible rule pairs are there of the form

$$X(h) \rightarrow_1 Y_1(h)Y_2(w)$$

$$X(h) \rightarrow_2 Y_1(w)Y_2(h)$$

where $X, Y_1, Y_2 \in N$ and $h, w \in \Sigma$?

(a) $10^3 \times 1000^2$

(b) $10^2 \times 1000^3$

(c) 10×1000

(d) 10^3

2.2 Question (time: 5:59, slide: 12)

We are given a lexicalized grammar with some valid rules and some invalid rules. Which of the following rules are **valid**?

(a) DT(the) $\rightarrow$ a

(b) DT(a) $\rightarrow$ a

(c) SBAR(that) $\rightarrow_1$ COMP(that) S(was)

(d) SBAR(was) $\rightarrow_2$ COMP(that) S(was)

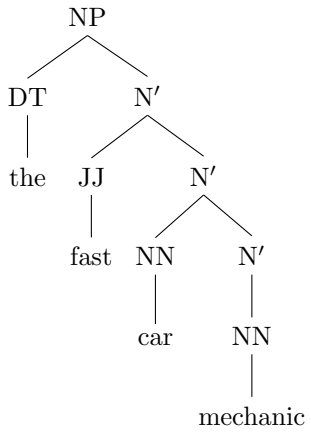
(e) SBAR(was) $\rightarrow_1$ COMP(that) S(was)

(f) PP(in) $\rightarrow_1$ IN(of) NP(company)

3 Parameter Estimation in Lexicalized PCFGs (Part 1)

3.1 Question (time: 2:57, slide: 16)

Say we have the sentence "the fast car mechanic" and the parse tree



We are also given the following head rules (where $*$ indicates the head)
$\text{NP} \rightarrow \text{DT } \mathbf{N'}^*$ $\text{N}' \rightarrow \text{JJ } \mathbf{N'}^*$ $\text{N}' \rightarrow \text{NN } \mathbf{N'}^*$ $\text{N}' \rightarrow \mathbf{NN}^*$

Which of the following parameters are used in calculating the probability of this parse tree?

(a) $q(\text{N '(mechanic)} \rightarrow_2 \text{JJ(fast) N '(mechanic)})$

(b) $q(\text{N '(fast)} \rightarrow_1 \text{JJ(fast) N '(mechanic)})$

(c) $q(\text{N '(mechanic)} \rightarrow_2 \text{NN(car) N '(mechanic)})$

(d) $q(\text{N '(car)} \rightarrow_1 \text{NN(car) N '(mechanic)})$

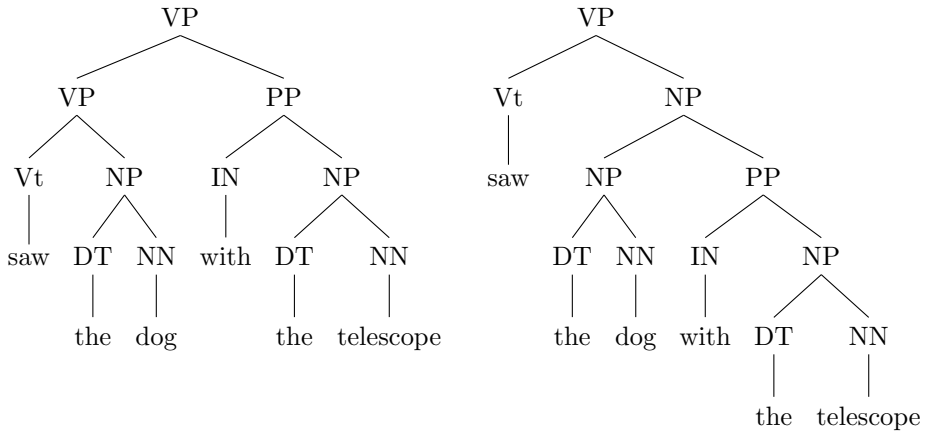
(e) $q(\text{NP(the)} \rightarrow_1 \text{DT(the) N '(mechanic)})$

(f) $q(\text{NP(mechanic)} \rightarrow_2 \text{DT(the) N '(mechanic)})$

4 Evaluation of Lexicalized PCFGs (Part 1)

4.1 Question (time: 5:46, slide: 22)

Say we have the phrase "saw the dog with the telescope" and we are given the gold parse tree (left) and a test parse tree (right)

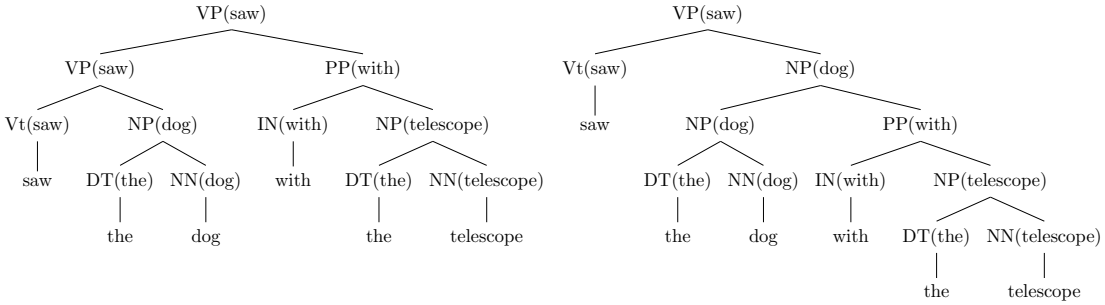


What is the **precision** of this test parse tree?

5 Evaluation of Lexicalized PCFGs (Part 2)

5.1 Question (time: 6:35, slide: 25)

Say we have the fragment "saw the dog with the telescope" and we are given the gold parse tree (left) and a test parse tree (right)



What is the **dependency accuracy** of this test parse tree (to three decimal places)?

A Answers

- that said said

  Trace the head word up from the bottom of the tree to the indicated non-terminal.

- (a)

  There are 10 non-terminals, so there are 10^3 possible unlexicalized rules of the form $X \rightarrow Y_1 Y_2$. Each one of these rules can have any word as its head h and any other word as w yielding 1000^2 lexicalized variants. The final grammar has $10^3 \times 1000^2$ rules of this form.

- (b) (c) (d)

  Invalid rules have a head word that (1) does not match any right-hand side head word, (2) matches the wrong right-hand side word, i.e. SBAR(was) $\rightarrow_1$ COMP(that) S(was).

- (a) (c) (f)

  Note that these rules always have the right-most noun as the head word. The incorrect parameters have a non-noun as head or the left noun in the case of car.

- 0.8

  The answer is 0.8. There are 5 constituents in both parse trees, and 4 of them are in both trees. The incorrect constituent is (NP, 2, 6), and the missing constituent is (VP, 2, 4).

- 0.833

  The answer is 0.833. There are 6 words in this sentence and the test parse gets 5 of the dependencies correct. The one incorrect dependency is "with" modifying "dog" instead of "with" modifying "saw".