097215 - Natural Language Processing

Homework Assignment 2

Question 1

Say we have a PCFG with start symbol S, and the following rules with associated probabilities:

- q(S $\rightarrow$ NP VP) = 1.0
- q(VP $\rightarrow$ Vt NP) = 1.0
- q(Vt $\rightarrow$ saw) = 1.0
- q(NP $\rightarrow$ John) = 0.25
- q(NP $\rightarrow$ DT NN) = 0.25
- q(NP $\rightarrow$ NP CC NP) = 0.3
- q(NP $\rightarrow$ NP PP) = 0.2
- q(DT $\rightarrow$ the) = 1.0
- q(NN $\rightarrow$ dog) = 0.25
- q(NN $\rightarrow$ cat) = 0.25
- q(NN $\rightarrow$ house) = 0.25
- q(NN $\rightarrow$ mouse) = 0.25
- q(CC $\rightarrow$ and) = 1.0
- q(PP $\rightarrow$ IN NP) = 1.0
- q(IN $\rightarrow$ with) = 0.5
- q(IN $\rightarrow$ in) = 0.5

Denote $S_1 =$ "John saw the cat and the dog with the mouse".

Claim:

"All parse trees for S_1 have either probability C under the given PCFG (for some $C > 0$)"

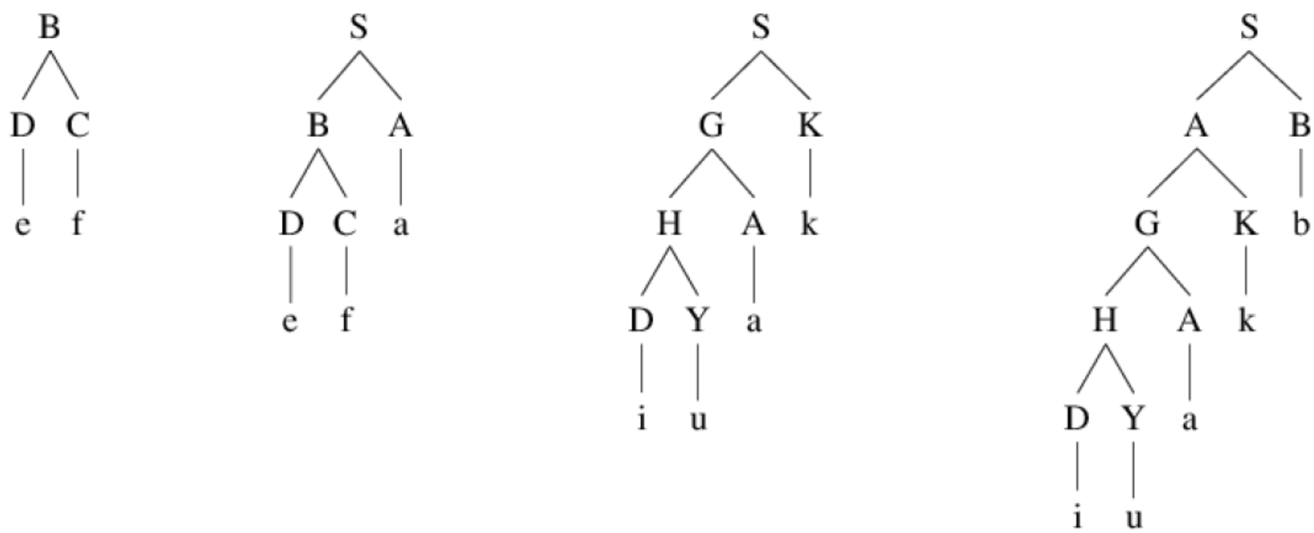
Prove the claim and find C.

Question 2

Consider the CKY algorithm for parsing with PCFGs. The usual recursive definition in this algorithm is as follows:

$$\pi(i,j,X) = \max_{\substack{X\to Y\ Z\in R,\\ s\in\{i...(j-1)\}}} \left(q(X \to Y\ Z) \times \pi(i,s,Y) \times \pi(s+1,j,Z)\right)$$

Now we would like to modify the CKY parsing algorithm to that it returns the maximum probability for any “left-branching” tree for an input sentence. Here are some example left-branching trees:



It can be seen that in left-branching trees, whenever a rule of the form X-> Y Z is seen in the tree, then the non-terminal Z must directly dominate a terminal symbol.
Assuming that our goal is to find the highest probability left-branching tree.

Write all recursive definitions and explain it (make sure you cover all necessary details)

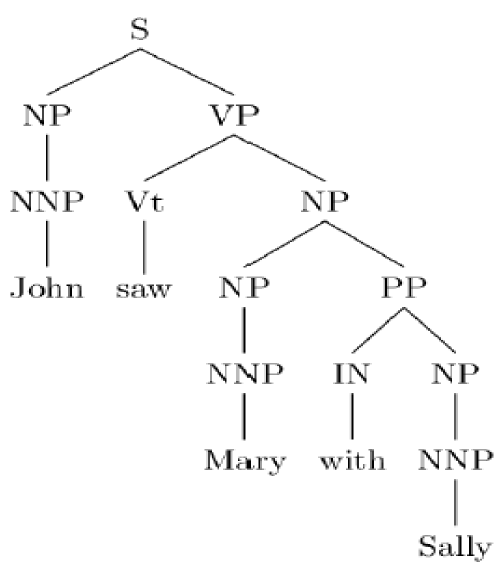
Question 3

Say we have a PCFG with the following rules and probabilities:

- q(S $\rightarrow$ NP VP) = 1.0
- q(VP $\rightarrow$ Vt NP) = 0.2
- q(VP $\rightarrow$ VP PP) = 0.8
- q(NP $\rightarrow$ NNP) = 0.8
- q(NP $\rightarrow$ NP PP) = 0.2
- q(NNP $\rightarrow$ John) = 0.2
- q(NNP $\rightarrow$ Mary) = 0.3
- q(NNP $\rightarrow$ Sally) = 0.5
- q(PP $\rightarrow$ IN NP) = 1.0
- q(IN $\rightarrow$ with) = 1.0
- q(Vt $\rightarrow$ saw) = 1.0

Consider the sentence: "John saw Mary with Sally".

The gold-standard parse tree for this sentence is:



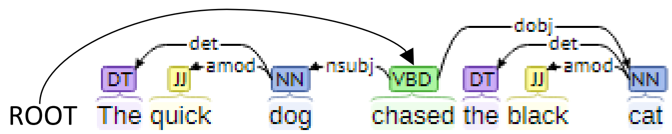
Run the CKY algorithm for the sentence "John saw Mary with Sally" in detail and produce the resulting parse tree under the given grammar.

Compute the F-score for the resulting parse tree.

(Note: the grammar is not in Chomsky normal form, so you must add a few improvements to the CKY algorithm. Describe these improvements in detail.)

Question 4

Given the following sentence and parse tree, show the states of the Transition-Based (MALT) Parser, as presented in tutorial 8:



Question 5

In tutorial 8 we saw the Chu-Liu-Edmonds algorithm, employed as an inference algorithm in a graph-based dependency parser.

For a given sentence: $\boldsymbol{x} = x_1 \ldots x_n$

- The directed graph $G_x = (V_x, E_x)$ given by:

$$V_{\boldsymbol{x}} = \{x_0 = \text{root}, x_1, \ldots, x_n\}$$
$$E_{\boldsymbol{x}} = \{(i,j) : i \neq j, (i,j) \in [0:n] \times [1:n]\}$$

Prove:

a. The algorithm terminates after a finite number of steps.
b. The algorithm's overall time complexity is $O(V_x \cdot E_x)$, given that:
   i. $|V_x| \leq |E_x|$.
   ii. The complexity of finding a cycle in the graph is $O(V_x + E_x)$.