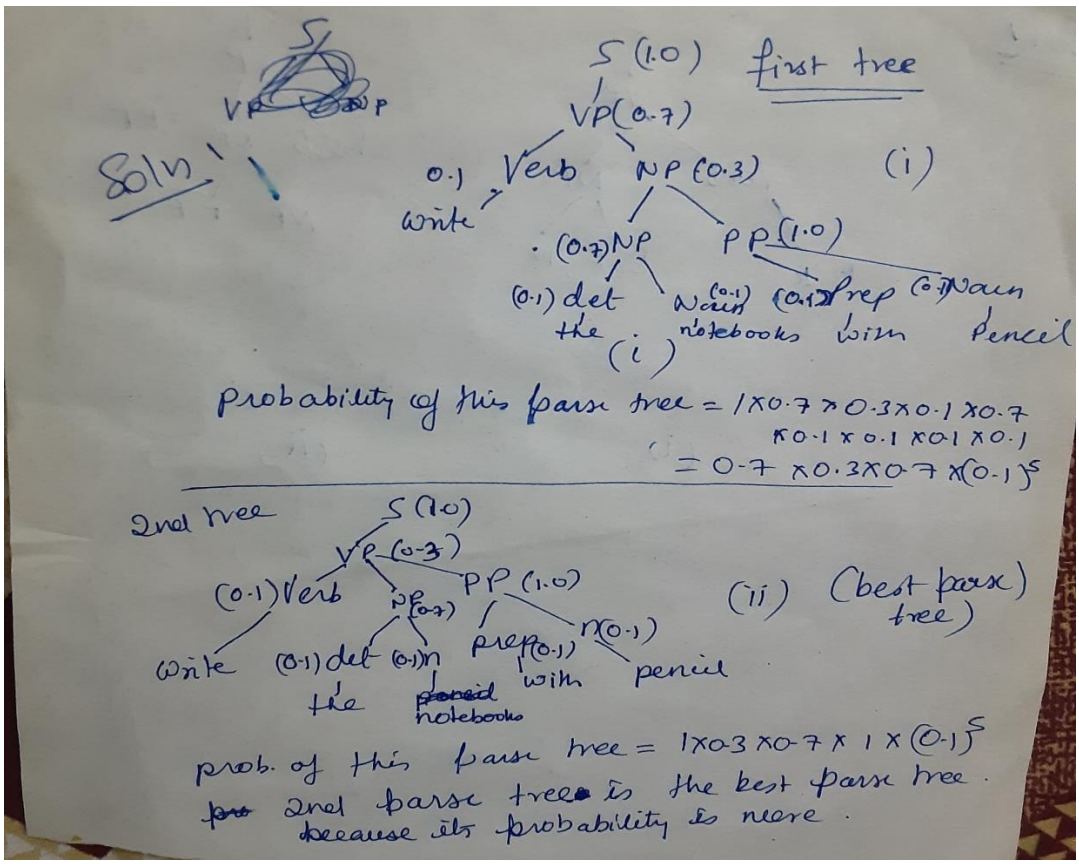
Q.1 (5+1=6 marks)

a) Draw the top-ranked parse tree for the sentence below by applying the PCFG given in below table. Does the results are good? Provide your comments.

Sentence: Write the notebooks with pencil.

Consider the following PCFG

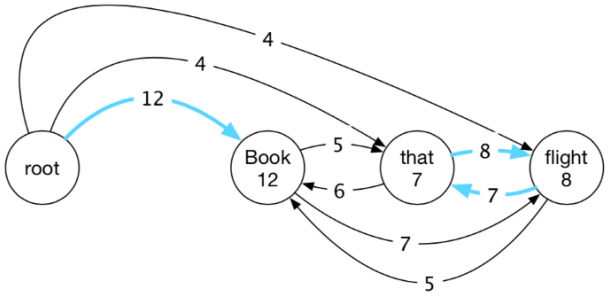
S → VP	1.0
VP → Verb NP	0.7
VP → Verb NP PP	0.3
NP → NP PP	0.3
NP → Det Noun	0.7
PP → Prep Noun	1.0
Det → the	0.1
Verb → Write | Ask | Find| ...	0.1
Prep → with | in | ...	0.1
Noun → notebooks | teacher | pencil | college | bike | summer |.....	0.1

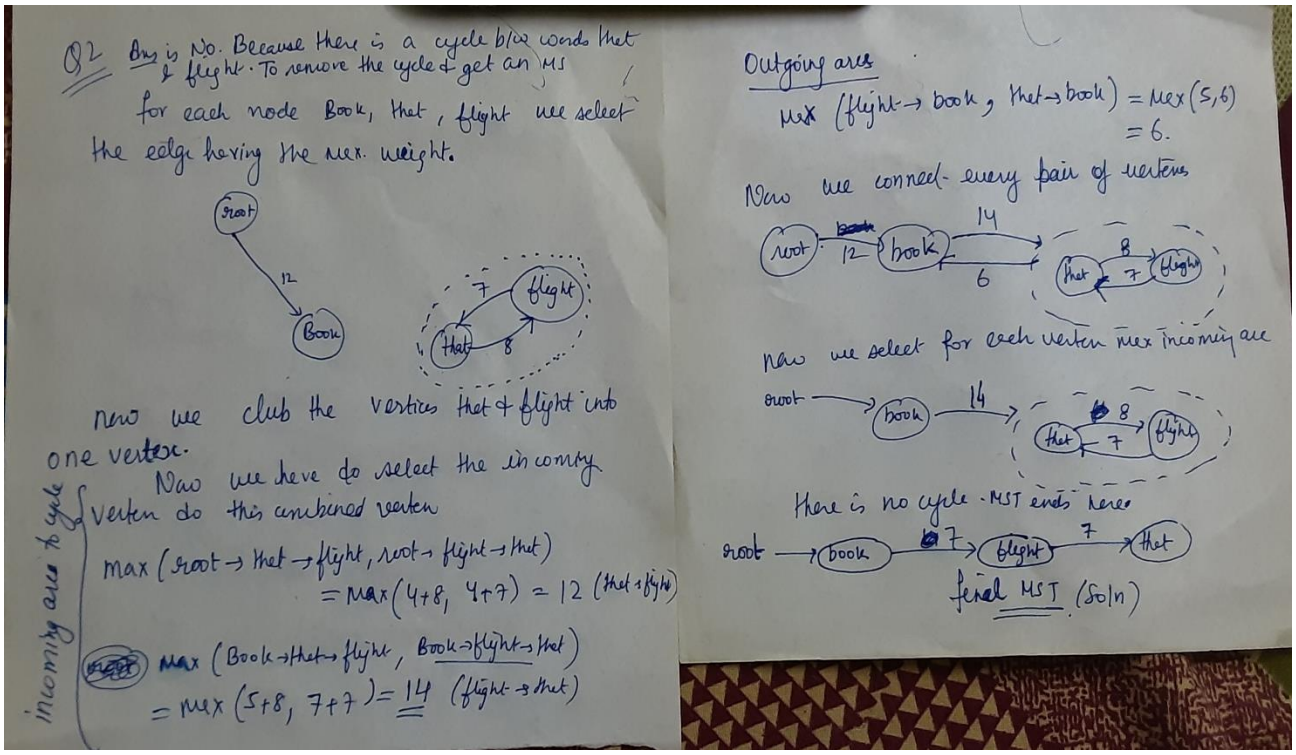


b) Which of the three Noun Phrases (1. Pronoun, 2. Proper Noun, 3. Common Noun) to be the most difficult to handle computationally while performing top-down parsing. Explain why?

Q2. (6+2=8 marks)

a). Does the following stage of a Edmond algorithm parsing has an MST ? If not, continue the algorithm for one more step with an Explanation. Obtain MST.





b) What are the basic differences between syntactic parsing and dependency parsing.

Ans: See from the slides

Q3 (6+2=8 marks)

a) **Given, the following training corpus, Using a bigram language model with and without add-one smoothing, what is P(Delhi is beautiful)?**

<s> Delhi is the capital of India </s>

<s> Delhi is cold </s>

<s> Delhi has beautiful gardens </s>

Solution:

P(Delhi is beautiful) = P(Delhi | <s>) * P(is|Delhi) * P(beautiful | is) * P(</s> | beautiful)

P(wn|wn−1) = C(wn−1 wn)/ C(wn−1)

Without Smoothing
P(Delhi | <s>) = 3/3 = 1
P(is|Delhi) = ⅔ = 0.676
P(beautiful | is) = 0/2 = 0
P(</s> | beautiful) = 0/1 = 0
Unique words = 10
With Smoothing
P(Delhi | <s>) = (3+1)/(3+10) = 0.31
P(is|Delhi) = (2+1)/(3+10) = 0.23
P(beautiful | is) = (0+1)/(2+10) = 0.08
P(</s> | beautiful) = (0+1)/(1+10) = 0.09
P(Delhi is beautiful) = 0.31*0.23*0.08*0.09 = 5.13 *10^-4

b) Suppose the sentence consists of random alphabets (A, a, B, b, ….,Z, z) and each of the 26 letters in upper and lower case occurs with equal probability. What is the perplexity of this sentence?

PP(W) = P(w1w2 ...wN) ^*(− 1 /N)

Since both upper and lower cases are considered 52 letters have equal probability.

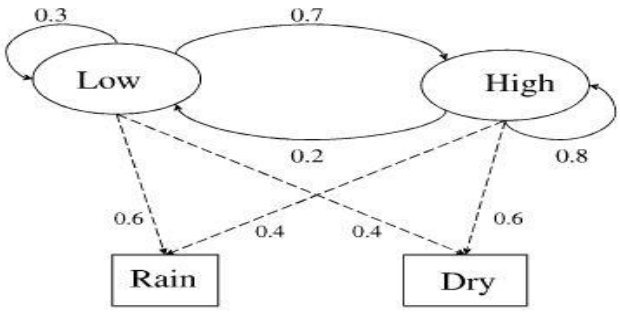
Perplexity is $\left(\left(\frac{1}{52}\right)^{52}\right)^{-1/52}$

= 52

Q4. (3+5=8 marks)

The following diagram describes HMM model with two hidden states: Low and High

and the observations are rainy and dry. Both the states are equally probable to be initial states



a) **Construct transition state matrix and emission matrix.**

b) **Let the observation sequence be given as Dry, Rain. Give the corresponding Hidden state sequence.**