Remaining topics

1. Predict C = f([Not bad]), where C ∈ {Positive, Negative} and f is a Bernoulli Naive-Bayes classifier. The train set is given as follows: (Show the computation) [3]

   good → positive
   not good → negative
   bad → negative
   can not praise enough → positive

Ans.

Vocab = {good, not, bad, can, praise, enough}

word	P(word|positive)	P(word|negative)
good	$\frac{1+1}{2+2}$	$\frac{1+1}{2+2}$
not	$\frac{1+1}{2+2}$	$\frac{1+1}{2+2}$
bad	$\frac{0+1}{2+2}$	$\frac{1+1}{2+2}$
can	$\frac{1+1}{2+2}$	$\frac{0+1}{2+2}$
praise	$\frac{1+1}{2+2}$	$\frac{0+1}{2+2}$
enough	$\frac{1+1}{2+2}$	$\frac{0+1}{2+2}$

$$P(Positive|\ Not\ Bad) = \frac{2}{4} * \frac{2}{4} * \frac{1}{4} * (1-\frac{2}{4})(1-\frac{2}{4})(1-\frac{2}{4})$$

$$= (\frac{2}{4})^2 * \frac{1}{4} * (1-\frac{2}{4})^4$$

$$= (\frac{1}{2})^2 * \frac{1}{4} * (\frac{1}{2})^4 = \frac{1}{2^{2+2+4}}$$ (1.25 mark)

$$\Rightarrow P(Positive|\ Not\ Bad) = \frac{1}{2^8}$$ ....(i)

$$P(Negative|Not\ Bad) = \frac{2}{4} * \frac{2}{4} * \frac{2}{4} * (1-\frac{2}{4})(1-\frac{1}{4})^3$$

$$= \frac{1}{2^3} * \frac{1}{2} * (\frac{3}{4})^3$$

$$= \frac{1}{2^3} * \frac{1}{2} * \frac{3^3}{2^6}$$ (1.25 mark)

$$\Rightarrow P(Negative|\ Not\ Bad) = \frac{3^3}{2^{10}} = \frac{3^3}{2^2} * \frac{1}{2^8}$$ ....(ii)

From (i) and (ii),

As, P(Negative|Not Bad) > P(Positive|Not Bad)

C = $f([Not\ Bad])$ **= Negative** (0.5 mark)

2. In multi-class classification, mention the scenarios when macro-average and weighted-average F1 scores are the preferred choices, respectively? [1]

   Ans. In the multi-class classification, the macro-average is generally used when the data is balanced.
   In the multi-class classification, the weighted-average is generally used when the data is imbalanced.

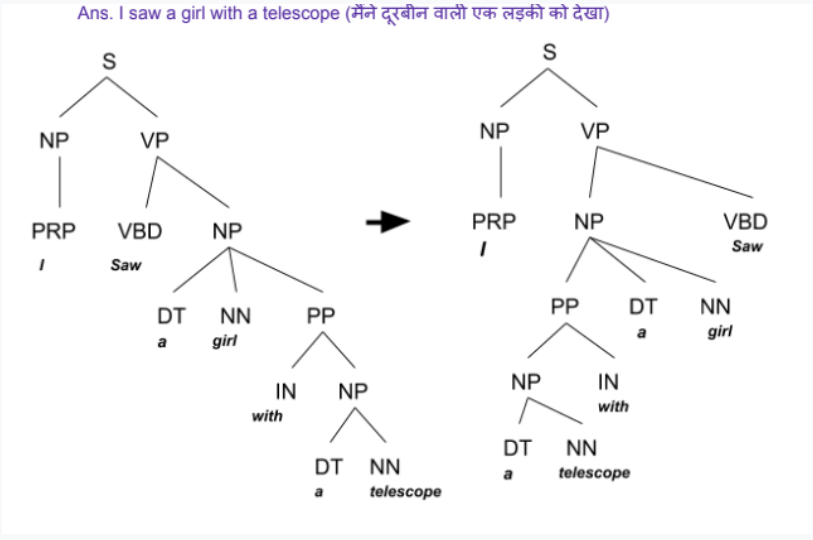
3. You have to choose one machine out of the following two to predict a hailstorm in a real scenario. Which machine are you going to choose and why? (You are required to provide proper reasoning for your answer) [1]
   a. Machine 1: Having 70% precision but 85% recall for predicting whether there will be a hailstorm or not.
   b. Machine 2: Having 85% precision but 70% recall for predicting whether there will be a hailstorm or not.

   Ans. a) Machine 1
   The high recall relates to a low false-negative rate whereas high precision relates to a low false-positive rate. For predicting the hailstorm, false-negative should be less, i.e., there should be less number of instances when the model predicts no hailstorm when there is actually a hailstorm. Therefore, high recall.

---

1. In CNN, if the input has a dimension of m X n and we expect the convolution output to be h X d, what should be the dimension of convolution filter? [1]
   Ans. Dimension of filter = (m-h+1) X (n-d+1)

2. Perform syntax transfer on the following sentences in English-Hindi MT. First give appropriate translations for both interpretations of the sentence in target language. [3*2]
   a. I saw a girl with a telescope. (telescope is with me)
   b. I saw a girl with a telescope. (telescope is with the girl.)



3. Given the following word-level alignment, compute the translation probability of the following. [1*2]
   a. P(girl | लड़की)
   b. P (एक | a)

Sentence 1	Sentence 2	Sentence 3
I (मैंने) saw (को) saw (देखा) a (एक) girl (लड़की) with (से) a (दूरबीन) telescope (दूरबीन) .(।)	I (मैंने) saw (देखी) a (लड़की) girl (लड़की) with (वाली) a (साइकिल) bicycle (साइकिल) . (।)	I (मैंने) saw (को) saw (देखा) a (एक) girl (लड़की) with (के) with (साथ) a (एक) cat (बिल्ली) . (।)

[Note: If you're not comfortable with Hindi, imagine the hindi words as some special symbols. E.g., imagine दूरबीन as X, लड़की as Y, etc.]

Ans.
a. P(girl | लड़की) = #(girl, लड़की) / #(लड़की) = 3/4
b. P (एक | a) = # (एक, a) / #(a) = 3/6 = 1/2

4. Create a phrase table (with max length of any phrase <= 4) using the above alignment. [6]

	मैंने	दूरबीन	से	एक	लड़की	को	देखा
I	✔						
Saw						✔	✔
a				✔			
girl					✔		
with			✔				
a		✔					
telescope		✔					

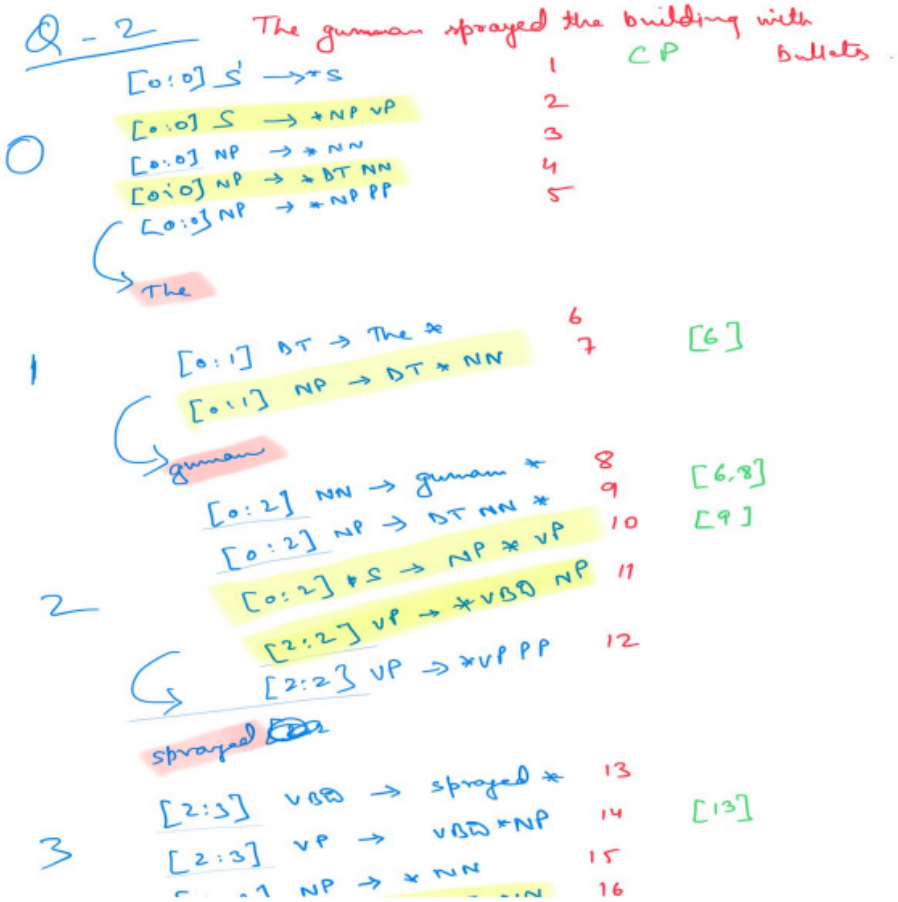
I	- मैंने	saw	- को देखा
saw	- को	saw a girl	- एक लड़की को देखा
saw	- देखा	a girl	- एक लड़की
a	- एक	a girl with	- से एक लड़की
girl	- लड़की	a telescope	- दूरबीन
with	- से	with a telescope	- दूरबीन से
a	- दूरबीन	a girl with a telescope	- दूरबीन से एक लड़की
telescope	- दूरबीन		

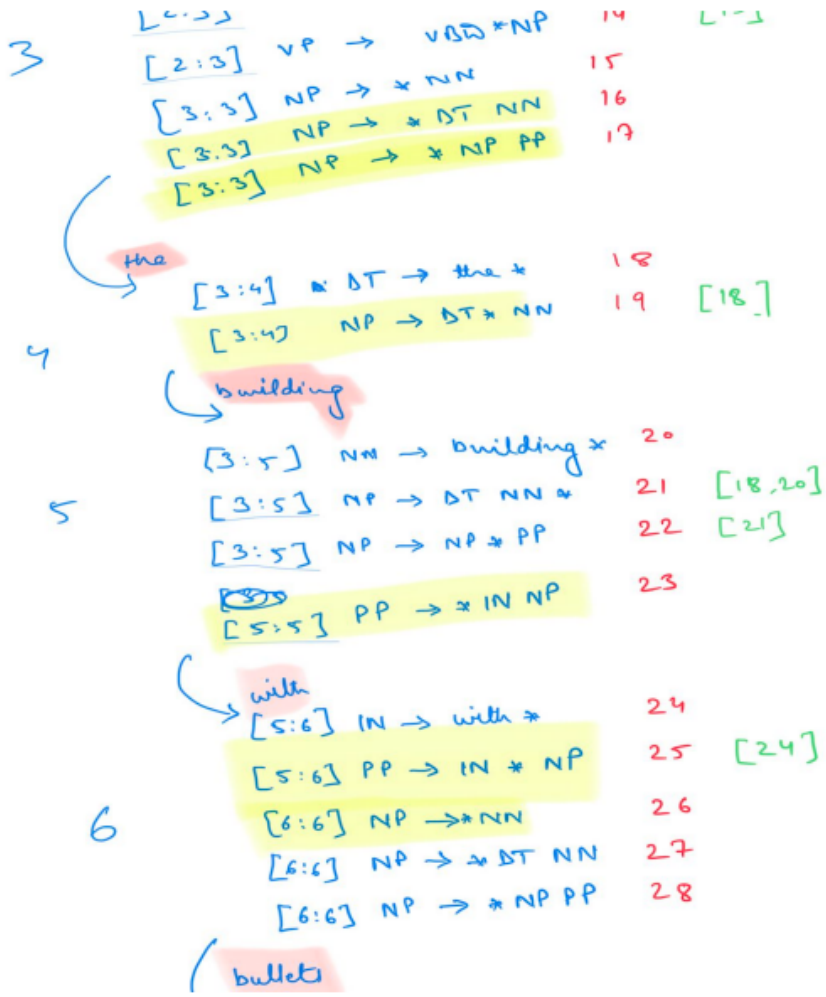
---

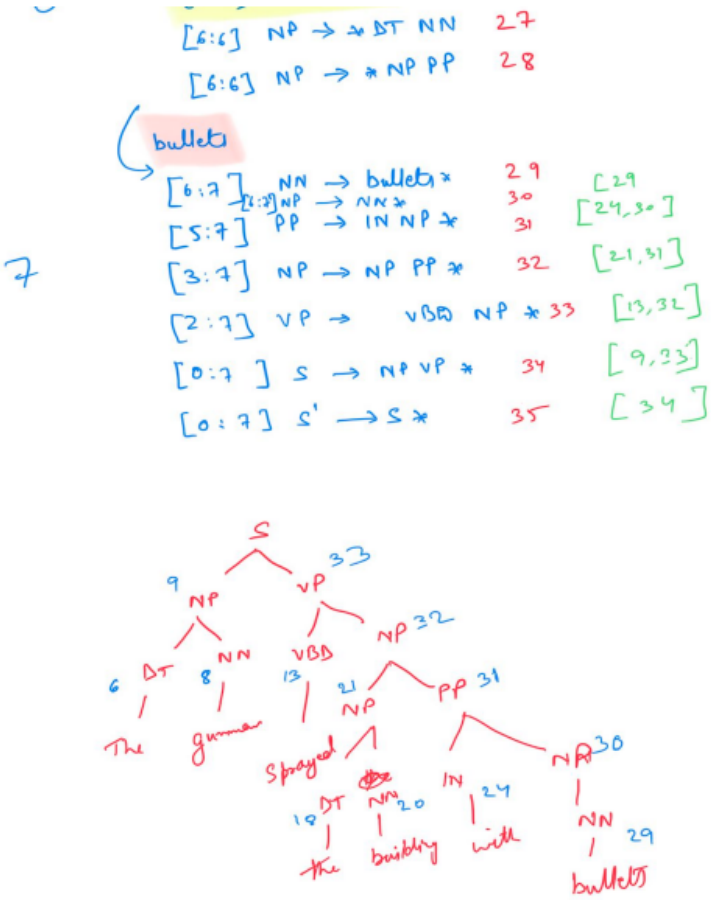
2. Given the following grammar G and a sentence S, show the complete parsing process (including parse tree construction) using Earley parser.

G:
- S → NP VP
- NP → NN | DT NN | NP PP
- VP → VBD NP
- VP → VP PP
- PP → IN NP
- NN → building | bullets | gunman
- VBD → sprayed
- DT → the
- IN → with

Sentence: The gunman sprayed the building with bullets [10 marks]




3
[2:3] VP → VBD * NP
[3:3] NP → * NN
[3:3] NP → * DT NN
[3:3] NP → * NP PP
the
[3:4] DT → the *
[3:4] NP → DT * NN
4
building
[3:5] NN → building *
[3:5] NP → DT NN *
[3:5] NP → NP * PP
5
[5:5] PP → * IN NP
with
[5:6] IN → with *
[5:6] PP → IN * NP
[6:6] NP → * NN
[6:6] NP → * DT NN
[6:6] NP → * NP PP
6
bullets
14
15
16
17
18
19
[18]
20
21
[18,20]
22
[21]
23
24
25
[24]
26
27
28

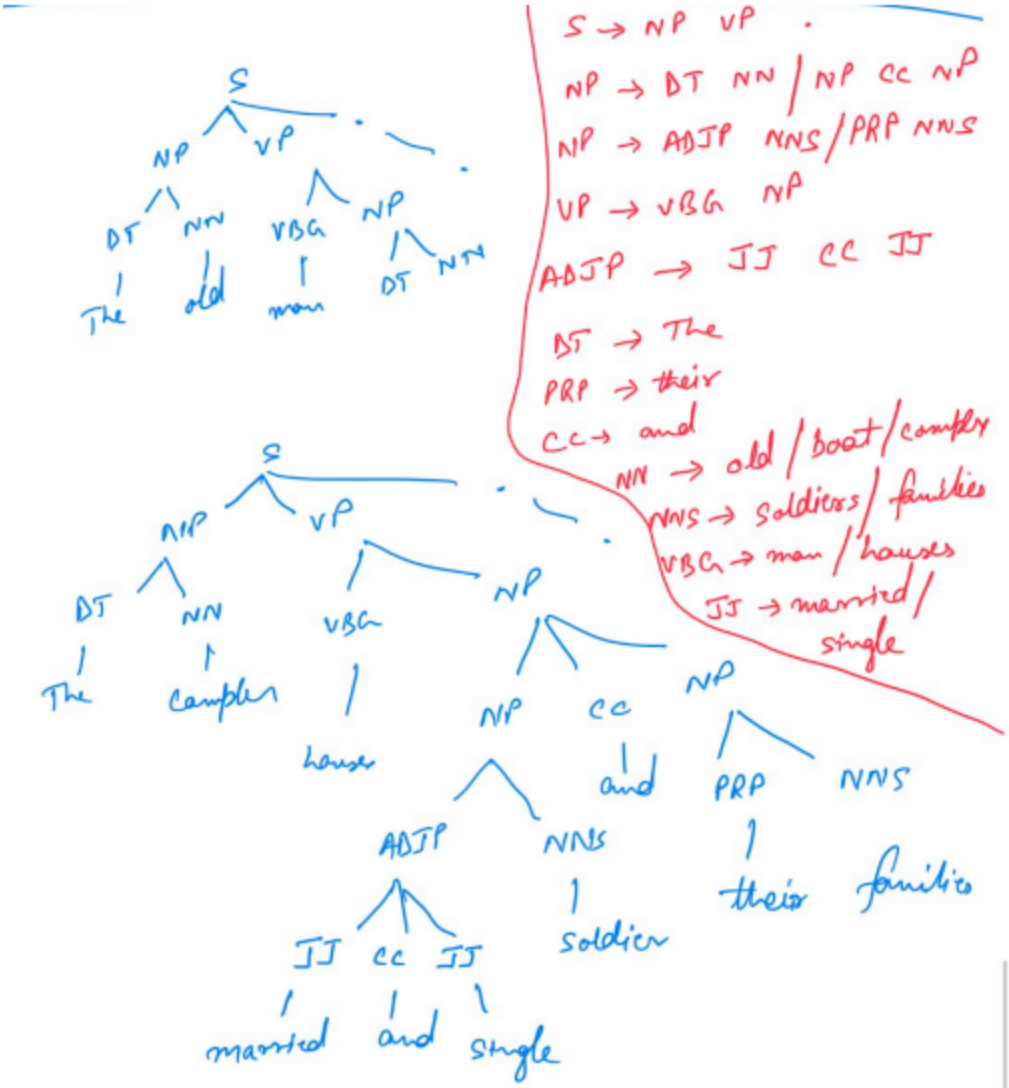


3. For the following sentences, give the PoS tag sequences and devise a (combined) grammar G to parse both. Also, show the parse trees. [10 marks]
    a. The old man the boat.
    b. The complex houses married and single soldiers and their families.

    [Note: PoS tag: (1+1), Grammar: (2+4), Parse tree: (1+1)]

Q-3 The old man the boat .
DT NN VBG DT NN .

The complex houses married and single soldiers and their families .
DT NN VBG JJ CC JJ NNS CC PRP NNS .

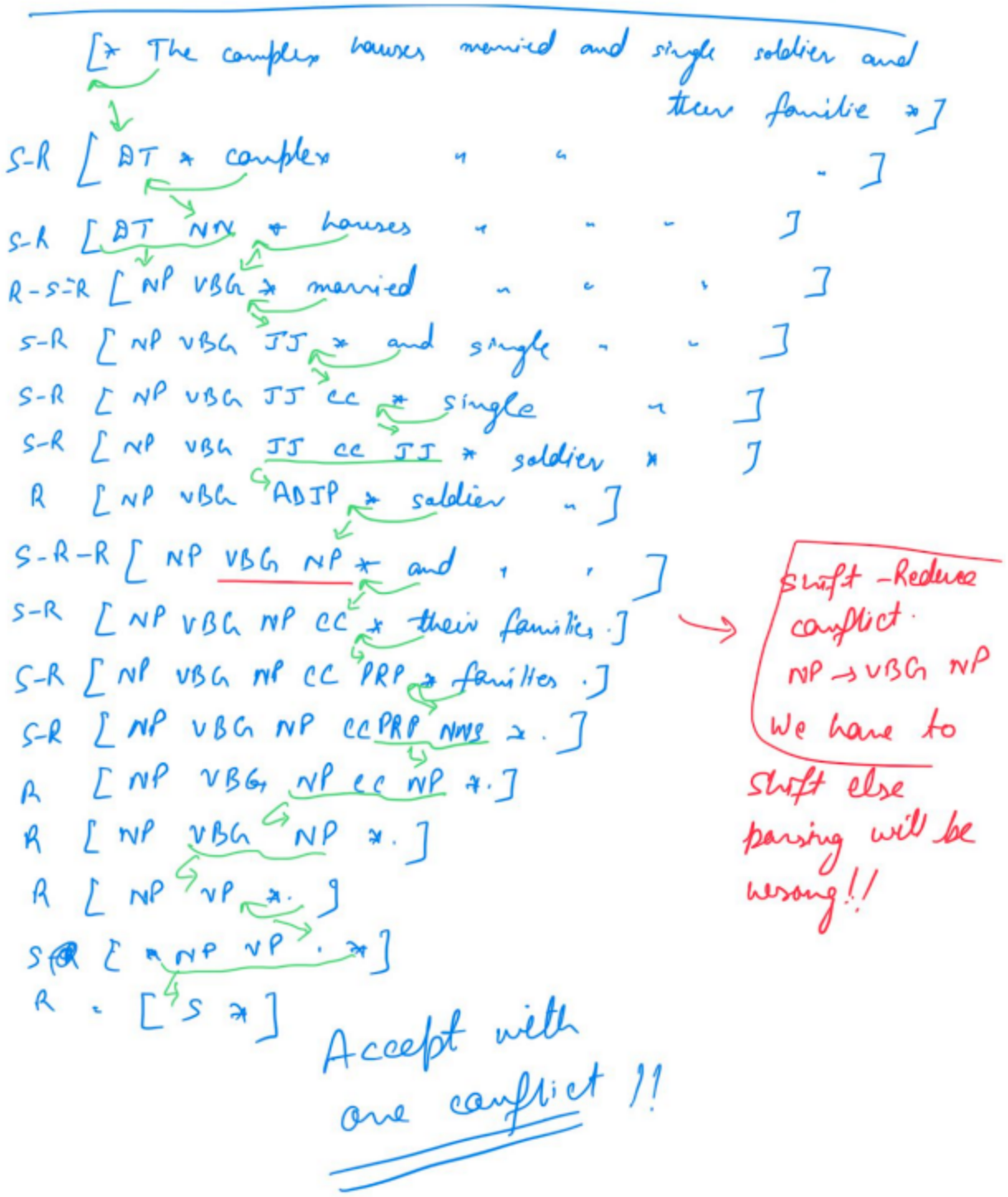


4. Using the grammar generated in the previous question, parse the sentences in 3a and 3b following bottom-up (shift-reduce) parser. Report the number of conflicts observed during the parsing process for each sentence. [4+4+2 marks]

Q4

[* The old man the boat.]
Shift [The * old man the boat.]
Reduce [DT * old man the boat.]
Shift [DT old * man the boat.]
Reduce [DT NN * man the boat.]
Reduce [NP * man the boat.]
Shift [NP man * the boat.]
Reduce [NP VBG * the boat.]
Shift [NP VBG the * boat.]
Reduce [NP VBG DT * boat.]
Shift [NP VBG DT boat * .]
Red [NP VBG DT NN * .]
Red [NP VBG NP * .]
Red [NP VP * .]
Shift [NP VP . *]
Reduce [S *]

Accept, No conflict



5. Extract the minimal set of rules for a grammar G for the following two parse trees.
    a. (S (NP (PRP My) (NN cat)) (ADVP (RB also)) (VP (VBZ likes) (S (VP (VBG eating) (NP (NNS mangoes))))) (. .))
    b. (S (NP (PRP He)) (VP (VBZ is) (VP (VBG enjoying) (NP (PRP his) (NN coffee)))) (. .))

[2 marks]

S → NP VP | VP | NP ADVP VP

ADVP → RB
NP → PRP NN | PRP | NNS
VP → VBZ VP | VBZ S | VBG NP
PRP → He | My | His
NN → Cat | Coffee
NNS → mangoes
RB → also
VBZ → likes | is
VBG → enjoying | eating
Note:
- 1 mark deduction if grammar is given separately for both trees
- 0.25 deduction if missed 1-2 productions.
- 1 mark deduction if missed either terminals/lexicon or non-terminals/production

6. Convert the following grammar into Chomsky Normal Form (CNF). [2 marks]

G: S → Da | b
D → bc | d

Q.8

$S \rightarrow Da \mid b$
$D \rightarrow bc \mid d$

$S \rightarrow DD' \mid b$
$D' \rightarrow a$
$D \rightarrow D^2 D^3 \mid d$
$D^2 \rightarrow b$
$D^3 \rightarrow c$