Natural Language Processing

Assignment 5

Type of Question: MCQ

Number of Questions: 9 Total Marks: (8×1)+(1×2)=10

==========

Question 1.

Which of the following are true? **[1 mark]**

A) Given a CFG and its corresponding CNF, they both produce the same language.
B) For a given grammar, there can be more than one CNF.
C) It requires '2n+1' productions or steps in CNF to generate a string w of length 'n'.
D) None of the above

Answer: A, B
Solution: Let n be the length of a string. We start with the (non-terminal) symbol S which has length n=1. Using (n−1) rules of form NT→NT NT (where NT represents a non−terminal) we can construct a string containing 'n' non-terminal symbols. Then on each NT symbol of said string of length 'n' we apply a rule of form NT→T. i.e. we apply n rules. In total we will have applied
(n−1) + n = 2n−1 rules.

==========

Question 2:
Consider the CFG given below:
$S \rightarrow aSb|D$
$D \rightarrow Dc|\epsilon$
How many non-terminals should be added to convert the CFG into CNF? **[1 mark]**

A) 3
B) 2
C) 4
D) 5

Answer: D
Solution: The final CNF is:

$S' \rightarrow AE|AB|DC|c$
$S \rightarrow AE|AB|DC|c$
$E \rightarrow SB\ D \rightarrow DC|c$
$A \rightarrow a\ B \rightarrow b\ C \rightarrow c$

==========

For question 3 to 6 consider the following PCFG fragment:

$S \rightarrow$ NN VP	0.50	$S \rightarrow$ VP NN	0.50
$NP \rightarrow$ NN PB	0.40	$P\ B \rightarrow$ PP NN	0.30
$V\ P \rightarrow$ VB NN	0.30	$V\ P \rightarrow$ VB NP	0.20
$V\ P \rightarrow$ NN VB	0.25	$V\ P \rightarrow$ NN PB	0.15
$P\ P \rightarrow$ with	0.10	$P\ P \rightarrow$ without	0.10
$V\ B \rightarrow$ play	0.30	$V\ B \rightarrow$ enjoy	0.20
$V\ B \rightarrow$ watch	0.25	$NN \rightarrow$ children	0.15
$NN \rightarrow$ cricket	0.15	$NN \rightarrow$ friends	0.20
$NN \rightarrow$ football	0.10	$NN \rightarrow$ music	0.12

For a sentence S = w1w2w3w4, assume that the cells in the table are indexed as follows:

	1	2	3	4	
w_1	11	12	13	14	1
w_2		22	23	24	2
w_3			33	34	3
w_4				44	4

Question 3:

Using CKY algorithm, find the probability score for the most probable tree for the sentence S_1 = "**children play cricket with friends**". **[1 mark]**

A) 5.06×10^{-4}
B) 2.73×10^{-3}
C) 1.62×10^{-6}
D) None of the above

Answer: c
Solution: Calculate the probability using the Bottom-Up method as explained in the lecture. 2

==========\

Question 4:

Using CKY algorithm, find the number of parse trees for the sentence S_2 = **children enjoy music** and the probability score for the most probable tree. **[2 marks]**

A) 1, 4.95×10^{-3}
B) 2, 0.36×10^{-3}
C) 3, 0.99×10^{-3}
D) 2, 0.54×10^{-3}

Answer: D
Solution:
There are two parse trees.
$S \rightarrow NN_{11}\, VP_{23} = 0.5 \times 0.15 \times (0.3 \times 0.2 \times 0.12) = 0.54 \times 10^{-3}$ $S \rightarrow VP_{12}\, NN_{33} = 0.5 \times (0.25 \times 0.15 \times 0.2) \times 0.12 = 0.45 \times 10^{-3}$

We get the above probabilities with CKY algorithm.

==========

Question 5:
Consider the expression below:

P("children watch football enjoy music", $N_{34}|G$) = P_j P("children watch football enjoy music" $|N^j_{34}, G$)
What does the L.H.S. represent? **[1 mark]**

A) Probability of the sentence "children watch football enjoy music", given a grammar G.
B) Probability of the sentence "children watch football enjoy music", given a grammar G and some rule which derives the segment "football enjoy".
C) Probability of the sentence "children watch football enjoy music", given a grammar G and that there is some consistent spanning of the segment "football enjoy", i.e. from word 3 to 4.
D) None of the above

Answer: C
Solution: Refer to Inside-Outside Probabilities.
==========

Question 6:
Suppose after parsing the sentence S_2 = children enjoy music with CKY algorithm, the non-terminals that appear in position 12 and 23 are NT_1 and NT_2 respectively. Compute the outside probabilities for $\alpha_{NT1}(12)$ and $\alpha_{NT2}(23)$. (1 mark)

A) 0, 0.075
B) 0.25, 0
C) 0.30, 0.06
D) None of the above

Answer: D
Solution:
$\alpha_{VP}(12) = 0.5 \times 1 \times 0.12 = 0.060$
$\alpha_{VP}(23) = 0.5 \times 1 \times 0.15 = 0.075$

==========

Question 7:
Which of the following grammars are valid CNF? **[1 mark]**

1. A → B 2. A → BCD 3. A → BC
   B → CD B → b B → ϵ
   C → c C → c C → c
   D → d D → d

A) 1.
B) 2.
C) 3.
D) None of the above

Answer: d
Solution: Valid CNF form is as follows:
$A \rightarrow BC$
$A \rightarrow a$

==========

Question 8:

Consider the CFG given below:
S -> ASA | aB
A -> B | S
B -> b | ε
How many non-terminals need to be added to convert the above grammar into CNF? (1 mark)

A) 1
B) 4
C) 2
D) 3

Answer: D

Solution: The final CNF is:
S' -> AX | YB | a | AS | SA
S -> AX | YB | a | AS | SA
A -> b | AX | YB | a | AS | SA
B -> b
X -> SA
Y -> a

==========

Question 9:

Which of the following are true with respect to a Top-Down and Bottom-Up Parser? [1 mark]

A) A Top-Down Parser never explores options that will not lead to a full parse.
B) A Bottom-Up Parser never explores options that will not lead to a full parse.
C) A Top-Down Parser never explores options that do not connect to the actual sentence.
D) A Bottom-Up Parser never explores options that do not connect to the actual sentence.

Answer: A, D
Solution: Follow the lecture slides.