NLP Sample Problems

(1) Consider the corpus

1. Ram ate mango
2. Rai ate orange
3. Ram cooks apple
4. Rai ate apple

Assume a bi-gram language model-
What is the probability of the sentence

"Ram ate apple"

(2) Consider the PCFG given below:

$S \rightarrow NP\ VP$ (1)
$NP \rightarrow Det\ N$ (0.7)
$NP \rightarrow N$ (0.3)
$VP \rightarrow V$ (0.8)
$VP \rightarrow V\ NP$ (0.2).

$N \rightarrow$ boy [0.4]
$N \rightarrow$ man [0.6]
$Det \rightarrow$ The [1]
$V \rightarrow$ saw [0.6]
$V \rightarrow$ ate [0.4]

What is the probability of the sentence

"The boy saw the man"?

Note: If multiple parse trees are possible for sentence S, then $P(S) = P(t_1) + P(t_2) + \cdots + P(t_n)$.

(3) Design an FST for identifying regular and irregular past tense.

(4) Consider the CFG given below

$S \rightarrow NP\ VP$

$NP \rightarrow N$

$NP \rightarrow Det\ NP$

$VP \rightarrow V$

$VP \rightarrow V\ NP$

$NP \rightarrow Adj\ N$

$N \rightarrow old \mid man \mid saw \mid mill.$

$Det \rightarrow the$

$Adj \rightarrow old \mid small$

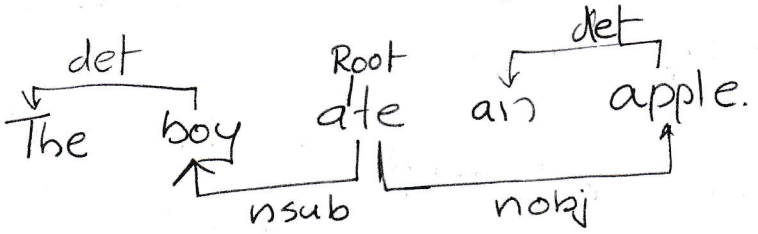
$V \rightarrow man \mid saw$

Apply CKY parsing ~~or~~ for the Sentence

"The old man saw the mill"

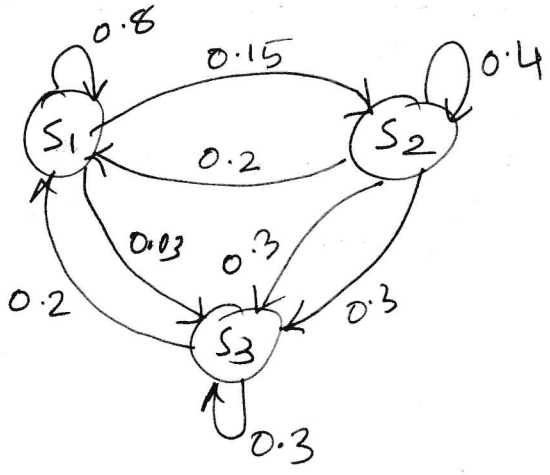
Note: grammar needs to be in CNF.

(5) Consider the dependency graph:



Use S-R parser to verify the dependency structure. write down the configuration & state at each stage.

(6) Consider the state transition diagram with 3 states.



Assume the previous states

$t_0 = S3$, $t_1 = S1$, what is the probability of $t_2 = S1$. ? [consider a first order Markov model.]

(7). Consider the emission probabilities & transition probabilities.

	w_1	w_2	w_3
T_1	0.1	0.3	0.6
T_2	0.2	0.2	0.6
T_3	0.4	0.1	0.5

	T_1	T_2	T_3
T_1	0.01	0.7	0.29
T_2	0.9	0	0.1
T_3	0.4	0.5	0.1

compute the following probability

$P(x_1 = w_1, x_2 = w_2, y_1 = T_1, y_2 = T_3)$?

Assume all POS tags are equally likely to be at the starting position.

(8) Assume word W_i appears in 800 wikipedia mentions, with following context links—

C1 → 300

C_2 → 268

C_3 → 132

C4 → 480

No links → 20

what is keyphrasenen (w_i).

Commonness (c_i, w_i) $c_i = (c_1, c_2, c_3, c_4)$

(9) Consider the ~~sentence~~ system generated summary and reference summaries → compute $ROUGE_1$, $ROUGE_2$. Sup

S: The boy sleeps on bed.

R1: boy sleeps on a red bed

R2: small boy sleeps on wooden bed.

a bed in the room.

(10) wordnet / WSD — } refer quiz-3 questions.
Semantic relations.

(11)

Total - 17 questions.

Questions		Marks	
10	x	3	= 30
7	x.	10	= 70
			100.

10 marks question contains sub questions.

CO1 — 18 marks

CO2 — 30 mark

CO3 — 29 mark.

CO4. — 23 marks.

CO1 → basic steps, levels of NLP, morphology, ..

CO2 → FST (morphological parsing), CFG, CYK, ~~PCFG~~.
~~Hmm~~, POStags, n-gram models, Distributional Semant.

CO3 → WSD, PCFG, Hmm,

CO4 → Entity linking, Text classified, Summarization