Natural Language Processing
Quiz 1

MM: 20 (10 each)
Time: 1 hr

1. Consider the following sentences
   i. The monkeys climb the tree having the largest flowers.
   ii. The monkeys climb the tree having eaten the bananas.

   For each of the above two sentences:
   a. Draw the Phrase Structure Tree and the Dependency Tree for each of the above two sentences.
   b. Add context free grammar (CFG) rules (to the simple grammar explained in the class) to cover each of the above two type of sentences.
   c. Augment the CFG covered in the class with features specifications to handle subject verb agreement treating the above two sentences as examples.

2. Please compute the likelihood of sentence 'A', given the following data, in two ways
   i. Using P(Wi|Wi-1) (conditional probability)
   ii. Using P(Wi-1,Wi) (joint probability, hint: What is the denominator?)

   Data:
   - The boy ate a chocolate
   - The girl bought a chocolate
   - The girl then ate a chocolate
   - The boy bought a horse

   Sentence:
   - The boy bought a chocolate

M19CSE473- Natural Language Processing

Mid Semester Examination

Note: All questions are compulsory. Marks are mentioned next to the questions. Pace yourself. Best of Luck.
Time: 90 Minutes

MM: 100 marks

Section A (45 Marks):

A1) **(A1.1)** Consider the sentence — 25 marks

Ram saw a monkey eating a banana.

(a) Draw all possible phrase structure trees (PSTs) & dependency trees (DTs) for the sentence.

(b) Write down the CFG rules for producing the above PSTs.

(c) Write down feature specification associated with CFG rules for obtaining.

(i) agent or k1 of 'saw'

(ii) Agent or k1 of 'eat'

(A1.2) For the following sentence:

Ram saw a monkey eating a banana in the evening with a spoon.

(a) Draw all possible meaningful PSTs and DTs.

(b) What phenomena control unnecessary ambiguities in this case? List the phenomena and explain.

A2) Consider the following sentences; — 20 Marks

(i) Ram promised Mohan to go to school early

(ii) Ram persuaded Mohan to go to school early

(a) For each of the above sentences:

-Draw its phrase structure tree (PST)

-Draw its dependency tree (DT)

(b) What is the agent of the verb 'go' in each of the above sentences?

Is it marked explicitly in the dependency tree or would you have to infer it? Why?

(c) Write feature specification to get the agent or k1 of 'go'

Section B (30 Marks):

B1) Language Modeling may be used to determine the source (Author, Domain etc.) of a given text. What measures would you use to accomplish that? Why? — 2+8

B2) Why is Smoothing required? Explain a Smoothing Algorithm (except add-one) of your choice in detail. — 2+8

B3) Language Modeling may be used for next word prediction. What would be the algorithm to produce 10 distinct sentences from a given LM? What would be the impact of the order of the n-gram on the quality of the text produced? Why? — 2+2+6

Section C (25 Marks): 3+4+3

C1) What is the benefit of using HMMs for Language Modeling? Which Algorithm will be used for this? Write the pseudocode for this algorithm. 10

C2) In Baum Welch algorithm explain each expression and the final parameter re-estimations with detailed explanation. 1+3+1

C3) Given $\lambda=(A, B, \pi)$, would there be any differences in $P(O|\lambda)$ given by Forward algorithm vs Backward algorithm? Can you write a Backward Viterbi? Would the Q^* (best state sequence) change as compared to regular Viterbi?

M19CSE472 Natural Language Processing

End Semester Exam

November 22, 2019

Time: 3 hrs

MM: 180

Note: Marks are mentioned next to the questions.

1. Consider the following sentence:

   ```
   The girl gave the book which she had issued from the library
   to her friend who needs it.
   ```

   (a) Draw the phrase structure tree (PST) and dependency tree (DT) for the above sentence.

   (b) Give CFG rules (beyond the simple CFG grammar given in the class) which allow you to handle the relative clause, and for example, the above sentence.

   (c) What do you understand by empty noun phrase in a sentence? Justify why they are needed? Explain with respect to the above example sentence.

   (d) Give feature specification for inferring the gender and number of the empty noun phrase in a relative clause S'.

   (e) Show the values of gender and number of the empty noun phrases for the sentence above, and how the values are derived using your feature specification.

   [30 Marks]

2. Consider the following sentence in Hindi (with English gloss) followed by one of its translations.

   ```
   jisa laDakii ne bhaagate huey sher ke bachche ko dekha
   which girl  -ne running       lion 's cub    -ko saw

   vaha jangal meiM rahatii hai
   she  forest -in  live    -s

   (The girl who saw the running lion cub lives in the forest.)
   ```

   There are other translations also. For those who might not know Hindi, 'bhaagate huey' (running, in participial form) in the above sentence may modify any of 'sher' (lion), 'bachche' (cub), or 'dekhaa' (saw).

   (a) Write down all the possible dependency trees for the above sentence.

   (b) Make karaka charts for the verbs 'bhaag' (run), 'dekha' (saw), and 'rahataa' (live). Their karakas would be like those of the corresponding words in English only.

   (c) Give karaka chart transformation rule for participial form of 'bhaagate huey' (running, in participial form).

(d) Show constraint graph for the above sentence.

(e) Show integer programming inequalities for the above constraint graph.

(f) Show all the possible solution graphs.

[40 marks]

3. For the following sentence show parsing process as asked below:

```
Shyam drank milk.
  n     v    n
```

The following is the state of the stack after n (for the word 'Ram') is shifted on the stack:

		Stack Entry		
SN	Op	OPEN	REM-SNT	REM-RULES
1	s (shift)	n	2	R1

(a) What is the next operation?

(b) Show the stack after v (corresponding to 'drank') is shifted on the stack.

(c) What happens in the next operation (after v has been shifted). Show the stack.

[**For information:** The CFG rules for the language are as follows:

```
CFG Rules::
   R1 :  S   -->  NP VP
   R2a:  NP  -->  n
   R2b:  NP  -->  det n
   R3 :  PP  -->  prep NP
   R4a:  VP  -->  v
   R4b:  VP  -->  v NP
   R4c:  VP  -->  v NP NP
   R4d:  VP  -->  v PP-STAR
```

The entry on the stack above has three parts:

- OPEN - Open list (string of symbols being processed, terminals and non-terminals)
- REM-SNT - Remaining sentence (words yet to be seen)
- REM-RULES - Remaining rules (rules yet to be applied)]

[15 marks]

4. Consider the following sentence:

```
The robot put a book on the table in the almirah.
```

(a) Show the phrase structure tree and dependency tree. (Show only one meaningful tree for each of the above two types of tree.)

(b) Give the predicate calculus formula for each of the NPs. Show the derivation, clearly writing down the rules.

(c) Show the logic formula for the declarative sentence above.

(d) For the imperative sentence:

`Put a book on the table in the almirah.`

Show the procedure invocation with its arguments for the verb 'put' (PUT-PROC).

[20 marks]

5. Describe and compare Kneser-Ney and Witten-Bell Smoothing methods. Also, succinctly list down the key ideas for both.

[15 marks]

6. Derive HMM's objective function (what are you maximizing?). Why is Viterbi or Forward Algorithms required for this maximization? How can HMM be used for identifying Names in a text. Describe how the data would be annotated (what the training data would look like.)

[15 marks]

7. How would you use HMM for clustering words based on grammatical categories if no annotated training data is available but large amount of raw text is available.

[20 marks]

8. Compare GloVe, Word2Vec (Skip-Gram) and SVD based word representations highlighting the situations in which one might be preferred over another.

[15 marks]

9. Briefly describe Walker's algorithm or Decision List algorithm for Word Sense Disambiguation.

[10 marks]