Submitted To: Mam Aasma Khan
Submitted By: Syed Muhammad Ali
ID & Semester: 14258 – BS(SE) 5th
Course: Natural Language Processing

**Question No. 1:**

a. Explain Part of Speech Tagging (POS) and explain POS tag ambiguity with two examples.

**ANSWER**

## ➢ POS Tagging:

It is a process of converting a sentence to forms – list of words, list of tuples (where each tuple is having a form *(word, tag)*). The tag in case of is a part-of-speech tag, and signifies whether the word is a noun, adjective, verb, and so on.

**Default tagging**: is a basic step for the part-of-speech tagging. It is performed using the

Default Tagger class. The Default Tagger class takes 'tag' as a single argument. NN is the tag for a singular noun. Default Tagger is most useful when it gets to work with most common part-of-speech tag. That's why a noun tag is recommended.

## ➢ POS tag ambiguity:

Common parts of speech in English are noun, verb, adjective, adverb, etc. The POS tagging problem is to determine the POS tag for a particular instance of a word. The main problem with POS tagging is ambiguity.

- For Example

### POS Tag Ambiguity

In English : I bank$_1$ on the bank$_2$ on the river bank$_3$ for my transactions.

Bank$_1$ is verb, the other two banks are noun

b. State difference between open vs. closed classes.

# ANSWER

## ➢ OPEN CLASS (CONTENT/LEXICAL)

- Lexical words deal with content and vocabulary.
- They have concrete meaning that goes beyond their function in a sentence.

## ➢ CLOSED CLASS (Grammatical/Function)

- Grammatical words deal with the formation of sentences.
- They have ambiguous meaning and serve to express grammatical relationships with other words within a sentence.

c. Apply Viterbi Algorithm on the below given bigram and lexical probabilities

| Initial Probabilities | |
|---|---|
| Noun | 1\3 |
| Verb | 0 |
| Other | 1\3 |

| Bigram Probabilities | | | |
|---|---|---|---|
| | Noun | Verb | Other |
| Noun | 1\4 | 1\4 | 0 |
| Verb | 1\4 | 0 | 1\4 |
| Other | 1\3 | 0 | 1\3 |

| Lexical Probabilities | | | | | |
|---|---|---|---|---|---|
| | O1=time | O2=flies | O3=like | O4=an | O5=arrow |
| Noun | 1\5 | 1\5 | 0 | 0 | 1\5 |
| Verb | 1\5 | 2\5 | 1\5 | 0 | 0 |
| Other | 0 | 0 | 1\5 | 2\5 | 0 |

# ANSWER

| **hi** | **P1(h1)** | **P2(h2)** | **P3(h3)** | **P4(h4)** | **P5(h5)** |
|---|---|---|---|---|---|
| **noun** | 1/3x1/5=1/15 | 1/5x1/4x1/15=1/300 | 0x1/150=0 | 0 | 1/5x1/4x1/16 |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | 875 |
| **verb** | 0 | 2/5x1/4x1/15=1/150 | 1/5x1/4x1/150=1/3000 | 0 | 0 |
| **other** | 0 | 0 | 1/5x1/3x1/150=1/2250 | 1/3x2/5x1/2250=1/16875 | 0 |

**Question No. 2:**
Apply Bayesian theorem over the below given string:
^John got many NLP books. ^He found them all very interesting.
Where for lexical probabilities assume John=0.5, got=0.3, many=0.2, NLP=0.1 and books=0.

# ANSWER

## POS Tags:

**^N V A N N. ^ N V N A R A**.

## Recording Numbers:

| | **^** | **N** | **V** | **A** | **R** | **.** |
|---|---|---|---|---|---|---|
| **^** | 0 | 2 | 0 | 0 | 0 | 0 |
| **N** | 0 | 1 | 2 | 1 | 0 | 1 |
| **V** | 0 | 1 | 0 | 1 | 0 | 0 |
| **A** | 0 | 1 | 0 | 0 | 1 | 1 |
| **R** | 0 | 0 | 0 | 1 | 0 | 0 |
| **.** | 1 | 0 | 0 | 0 | 0 | 0 |

## Bigram Probability:

| | **^** | **N** | **V** | **A** | **R** | **.** |
|---|---|---|---|---|---|---|
| **^** | 0 | 1 | 0 | 0 | 0 | 0 |
| **N** | 0 | 1/5 | 2/5 | 1/5 | 0 | 1/5 |
| **V** | 0 | 1/2 | 0 | 1/2 | 0 | 0 |
| **A** | 0 | 1/3 | 0 | 0 | 1/3 | 1/3 |
| **R** | 0 | 0 | 0 | 1 | 0 | 0 |
| **.** | 1 | 0 | 0 | 0 | 0 | 0 |

Where lexical probability is given i.e.
John=0.5, got=0.3, many=0.2, NLP=0.1 and books=0.