Module:-1

① Different phases of analysis in NLP

→ **a) lexical analysis or Morphological analysis:-** The 1st phase of NLP is lexical analysis or tokenization. This phase involves breaking down the text into smaller units called tokens, or morphemes such as words or punctuation marks. The purpose of this phase is to identify the basic building blocks of the text, which can be further analyzed in subsequent phases.

Ex:- Consider the sentence, "John went to the store to buy some milk". The tokens in this sentence would be "John", "went", "to", "the", "store", "to", "buy", "some" and "milk".

b) Syntactic Analysis:- The 2nd phase of NLP is syntactic or parsing. This phase involves analyzing the grammatical structure of the text by identifying the relationships b/w words and the roles they play in the sentence. This phase helps to create a parse tree that represents the structure of the sentence.

Ex:- "John went to the store to buy some milk", the parser would identify "John" as the subject, "went" as the verb and "to the store" as the prepositional phrase modifying the verb.

c) Semantic Analysis:- The 3rd phase of NLP is semantic analysis. This phase involves analyzing the meaning of the text by identifying the relationships b/w words and the concepts they represent. This phase helps to create a representation of the meaning of the sentence.

Module:-1

① Different phases of analysis in NLP

→ **a) lexical analysis or Morphological analysis:-** The 1st phase of NLP is lexical analysis or tokenization. This phase involves breaking down the text into smaller units called tokens, or morphemes such as words or punctuation marks. The purpose of this phase is to identify the basic building blocks of the text, which can be further analyzed in subsequent phases.

Ex:- Consider the sentence, "John went to the store to buy some milk". The tokens in this sentence would be "John", "went", "to", "the", "store", "to", "buy", "some" and "milk".

b) Syntactic Analysis:- The 2nd phase of NLP is syntactic or parsing. This phase involves analyzing the grammatical structure of the text by identifying the relationships b/w words and the roles they play in the sentence. This phase helps to create a parse tree that represents the structure of the sentence.

Ex:- "John went to the store to buy some milk", the parser would identify "John" as the subject, "went" as the verb and "to the store" as the prepositional phrase modifying the verb.

c) Semantic Analysis:- The 3rd phase of NLP is semantic analysis. This phase involves analyzing the meaning of the text by identifying the relationships b/w words and the concepts they represent. This phase helps to create a representation of the meaning of the sentence.

Ex:- In the sentence "John went to the store to buy some milk", the semantic analysis would identify that John is a person a store is a place where things are sold, and milk is a type of food.

d) Discourse analysis:- The 4th phase of NLP. This phase involves analyzing the relationship b/w sentences in a text and the overall meaning of the text. This phase helps to create a representation of the meaning of the text as a whole.

Ex:- In a conversation b/w 2 people, discourse analysis would involve identifying the topic of the conversation, the purpose of the conversation, and the intentions of the speakers.

e) Pragmatic Analysis:- The 5th phase of NLP involves analyzing the context in which the text is used and the goals of the user. This phase helps to create a representation of the meaning of the text in the context of the users goals.

Ex:- In a search engine, pragmatic analysis would involve analyzing the users query and identifying the information the user is looking for.

② Regular Expression for:-

9) To accept strings "book" or "books":

→ /books?/b

Here /b → word boundary to ensure we match only whole words "book" or "books"

• ? → matches both book and books.
↳ makes the preceding char optional.(s)

ii) To accept colour and color :-

→ /colou?r/

? → makes the u optional, so it can match both "color" and "colour".

iii) To accept any +ve integer with an optional decimal point.

→ #~ ^\d + (\.\d+)?$

^ → asserts the start of the string

\d+ → matches a digits [0-9]

(\.\d+)? → optional group with a decimal point followed by one or more digits.

$ → asserts the end of the string

iv) To check a string is an email address or not

→ A-fA-z

^[a-z A-Z 0-9 _\.-]⁺ + @ [a-z A-z 0-9_\.-]⁺ + [A-za-z 0-9_] [A-za-z 0-9_] $

^ → asserts the start of the string

[a-z A-z 0-9 _\.-]+ → match a +ve no of acceptable char's at the start of the string

@ → match the @ sign

[a-z A-z 0-9 _\.-]+ → match any domain name including a dot

[A-za-z 0-9_] [A-za-z 0-9_] $ → match 2 acceptable char's but not a dot. This ensures that the

email address ends with .xx, .xxx, .xxxx etc.

③ The morphological type (Noun phrase, Verb phrase, Adjective phrase) of the following sentence segments.

i) Important to Bill

Type:- Adjective phrase

→ "Important" is an adjective modifying the noun "Bill".

ii) ~~too~~ Looked up the tree

Type:- Verb phrase

→ "Looked up" is the verb phrase, where "looked" is the main verb and "up" is an adverb which is modifying the verb. "The tree" is the noun phrase that serves as the direct object to the verb "Looked up".

④ Corpus of Three sentences

. There is a big garden

. Children play in a garden

. They play inside beautiful garden

} Training set

→ Calculate P for the sentence "They play in a big Garden" / assuming a bi-gram lang model.

probability

for bi-gram = $P(s) = \prod_{i=1}^{n} P(w_i | w_{i-1})$

wkt

$$P(w_i | w_{i-1}) = \frac{count(w_{i-1}, w_i)}{count(w_{i-1})}$$

Given Training set are :-

           <s> There is a big garden </s>
           <s> children play in a garden </s>
           <s> They play inside beautiful garden </s>

Bi-gram model   pseudo word introduced to mark
                        the beginning of the sentence in

$P(\text{There} / <s>) = \dfrac{~~~~~~<s> \text{There}}{<s>} = \dfrac{1}{3} = 0.33$ bi-gram estimation

$P(\text{is} / \text{there}) = \dfrac{\text{there is}}{\text{there}} = \dfrac{1}{1} = 1$

$P(\text{a} / \text{is}) = \dfrac{\text{is a}}{\text{is}} = \dfrac{1}{1} = 1$

$P(\text{big} / \text{a}) = \dfrac{\text{a big}}{~~~\text{a}} = \dfrac{1}{2} = 0.5$

$P(\text{garden} / \text{big}) = \dfrac{\text{big garden}}{\text{garden}} = \dfrac{1}{3} = 0.33$

$P(\text{play} / \text{children}) = \dfrac{\text{play chil}}{1} \dfrac{1}{1} = 1$

$P(\text{in} / \text{play}) = \dfrac{1}{2} = 0.5$

$P(\text{a} / \text{in}) = \dfrac{1}{1} = 1$

$P(\text{garden} / \text{a})^{\cdot} = \dfrac{1}{2} = 0.5$

$P(\text{play} / \text{they}) = \dfrac{1}{1} = 1$

$P(\text{inside} / \text{play}) = \dfrac{1}{2} = 0.5$

P (beautiful | inside) = $\frac{1}{1}$ = 1

P (garden | beautiful) = $\frac{1}{1}$ = 1

Test sentence (s):- They play in a big garden

P (They | <s>) * P (play | they) x P (in | play) x

P (a | in) x P (big | a) x P (garden | big)

= 0.33 X 1 X 0.5 X 1 X 0.5 X 0.33

= $\underline{\underline{0.027225}}$

⑤ Construct the Surface Structure (CSS) and Deep structure (DS)

i) The police will catch the snatchers

Surface structure → Active sentence

S

NP   VP

The   police   Verb   NP

Aux   V   Det   Noun

will   catch   the   snatchers
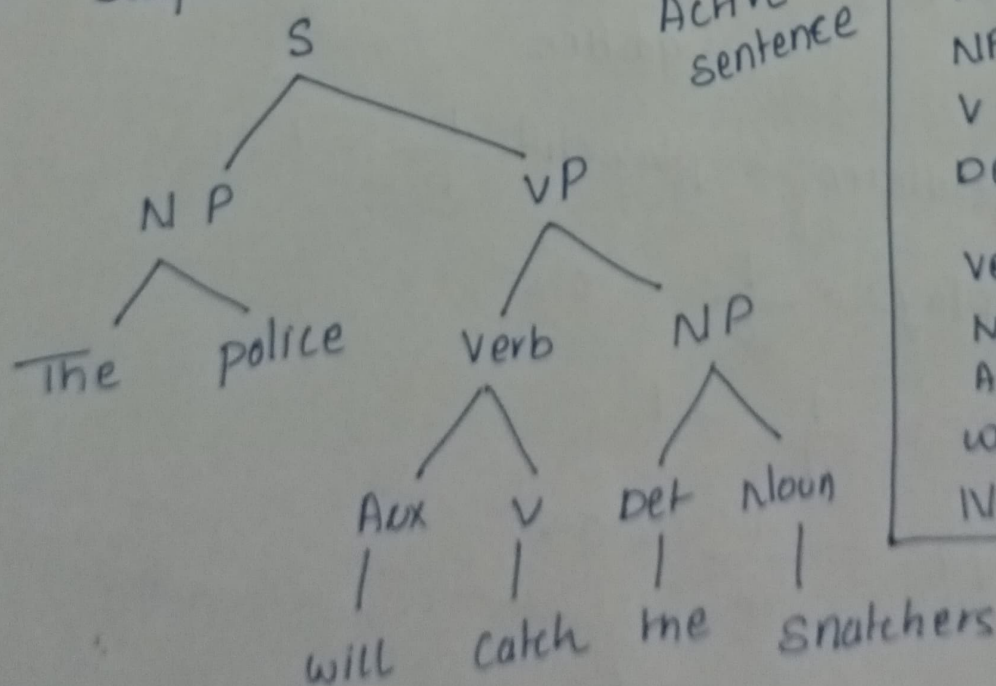
remember

S → NP + VP
VP → V + NP
NP → Det + Noun
V → Aux + Verb
Det → the, a, an
      those ----
Verb → catch ---
Noun → police...
Aux → am, is,
      were will---..
N → N + PP

Passive :— The snatchers will be caught by the Police

```
                            S
                  ┌─────────┴─────────┐
                 NP                    VP
              ┌───┴───┐        ┌────────┴────────┐
             Det     Noun    Verb                NP
              │       │     ┌──┴──┐         ┌─────┴─────┐
             The   Snatchers Aux  Verb     by          NP
                          ┌──┴──┐   │              ┌────┴────┐
                        will   be  caught         det      Noun
                                                   │         │
                                                  The      Police
```
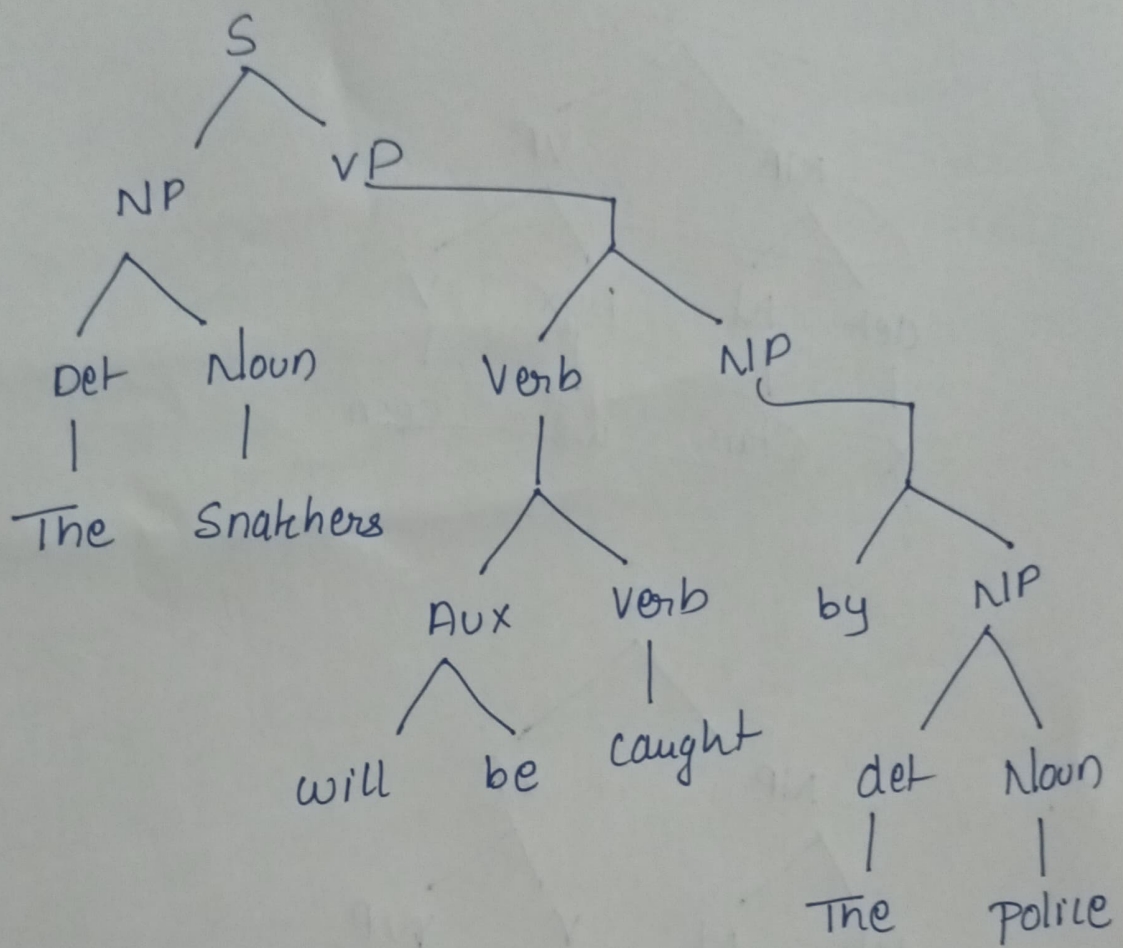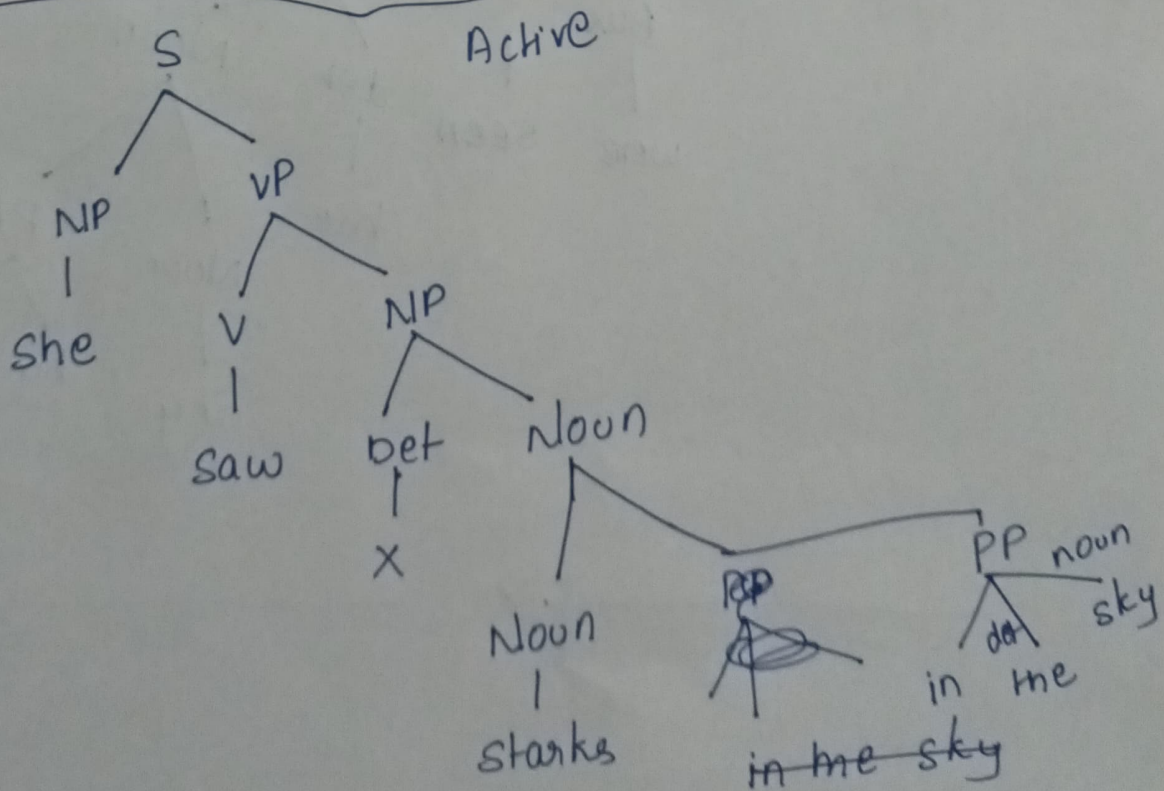
ii) She saw stars in the sky.   Active

```
                  S
            ┌─────┴─────┐
           NP           VP
            │        ┌───┴───┐
           She       V       NP
                     │    ┌───┴────┐
                    saw  Det      Noun
                     │    │    ┌────┴──────────┐
                     X   Noun  PP            PP  noun
                          │                ┌─┴─┐  sky
                        starks            det
                                           in  the
                                        in the sky
```

Passive :- Stars were seen in the sky by her

Tree diagram 1:
S
- NP
  - det — X
  - N — Stars
- VP
  - Aux — were
  - V — seen

Tree diagram 2:
S
- NP
  - det — X
  - N — stars
- VP
  - V
    - Aux — were
    - V — seen
  - VP
    - in
    - det — the
    - Noun
      - Noun — sky
      - PP — by her