(2) statistical pos Tagging.

uses machine learning & probability to predict the best pos tag for a word.

* Instead of using rules, it learns from large datasets.

1. Hidden Markov Model (HMM) Tagging

* uses probability to predict the most likely pos tag.

2. Maximum Entropy model.

* uses more context features (e.g, surrounding words, word endings) to make better predicti ons.

3. Neural network - Based Tagging.

* uses Deep Learning models (like LSTMs) to improve accuracy.

Hybrid pos Tagging.

* combines Rule-Based and statistical approaches.

Issues in pos Tagging.

* Ambiguity -> some words have multiple meanings.
* unknown words -> New (or) rare words might not be recognized.
* complex sentences -> Long sentences with tricky structures confuse model.
* Different writing style -> Formal and informal language vary.

1. Markov chain

Markov chain is a Mathematical model that helps predict future events based on the current state. Ignoring past events.

Ex: If you're playing a board game, your next move only depends on your current position. not on how you got there.

2) Hidden Markov chains (HMM).

A Hidden Markov model (HMM) is a special type of Markov chain where some states are hidden (not directly visible).

Ex: speech recognition.

3) Likelihood computation.

* This refers to calculating how probable a sequence of events is given an HMM.

Ex: If you see a weather forecast you can calculate how likely it is that it will rain tommorrow based on past weather patterns.

④ HMM Training.

* This is the process of teaching an HMM using training data so that it can make accurate predictions.

Ex: Spam filter is trained with past emails to learn which words (or) patterns indicates spam.

UNIT-3

Syntactic Analysis

* Syntactic Analysis (or parsing) checks whether a sentence follows grammatical rules.
* It helps computers understand the structure of a sentence.

Ex: "The cat sleeps" ✓

"Cat the sleeps" ✗

Context-Free Grammars (CFG)

* Context-Free Grammars is a set of rules used to describe the structure of sentences in a language.
* It defines how words and phrases combine to from valid sentences.

Importance of CFG

Defines Sentence structure :-

Helps computers understand valid Sentence patterns.

Used in NLP :- Helps in machine translation, speech recognition, and AI chatbots

Essential in programming :- used to define syntax of languages like Python, Java etc.

Efficient parsing :- Allows computers to check correct grammar in text editors and compilers.

Production Rules in CFG

CFG consists of production rules that define how sentences are formed.

A production rule has the form.

$$A \rightarrow B\ C$$

A (Non_Terminal) = A symbol that can be expanded further.

B, C (Terminals (or) Non Terminals) = parts that form a sentence.

Ex: Sentence formation using CFG

Rules:

Sentence → Noun phrase + verb phrase

Noun phrase → "The cat"

verb phrase → "eats the food"

Final Sentence: "The cat eats the food."

components of CFG

1. Terminals (Σ)

Words in the language (cannot be replaced)

Ex: "cat", "dog", "runs"

2. Non-Terminals (N).

symbols that can be replaced using rules.

Ex: Sentence, Noun phrase.

3. start symbol (S).

The starting point of sentence generation

Ex: sentence.

(4) Production Rules (P)

Rules defining how words combine.

Ex: Sentence → Noun phrase + verb phrase.

Grammar Rules for English.

* Grammer rules are formal system that define how sentences are structured in English.
* They are essential for understanding the syntax and semantics of the Language.

clauses and sentences

* A clause is a group of words that has a subject and a verb.
* clauses are the building blocks of sentences.

Types

① Declarative clauses

These are simple statements that gives information.

structure →: S → NP VP

sentence → Nounphrase + verbphrase

b) Embedded clauses

These clauses exist inside another clause.

Ex: I said that there were two flights to Denver

I said — main clause

that there were two flights to Denver — Embedded clause.

c) complex sentences

A complex sentence has one independent clause and one or more dependent clauses.

Ex: Although it was raining, we decided to go for a walk.

Although it was raining — Dependent

we decided to go for a walk. — Independent

Tree banks :

A Tree bank is a database of sentences that have been analyzed and annotated with grammatical (syntactic) and meaning (semantic) information.

Types of Treebanks

A) Semantic Treebanks (meaning-Based)

* Focus on the meaning and Relationships between words in a sentence.

Ex : Robot commands Treebank

B) syntactic Treebank :

* Focus on sentence structure (how words fit together. grammatically).

C) Significance of the penn Tree bank.

penn Treebank is one of the most important English treebanks.

contains 4.5 millions words and 36 Part of speech tags

Ex : noun, verb, adjectives.

Chomsky Normal Form (CNF).

Chomsky Normal Form (CNF) is a simplified version of a context free grammer (CFG).

Ex: In CNF, Every production rule must follow these forms.

a. $A \rightarrow BC$ (where A, B, and C are non terminals, and B & C are not the start symbol)

b. $A \rightarrow a$ (where A is a non_terminal and "a" is a terminal)

c. $S \rightarrow \varepsilon$ (only if the language allows empty strings)

steps to convert CFG to CNF

step 1 : Remove useless symbols.

step 2 : Remove Nullable (ε-producting

step 3 : Remove unit productions (A→B)

step 4 → convert Long Right-Hand Sides

step 5 → convert Terminals in mixed Rules.

→ probabilistic and Lexicalized Parsing.

→ probabilistic parsing.

* probabilistic parsing assigns probabilities to different possible sentence structure.

* It helps select the most likely correct structure for ambiguous sentences.

Lexicalized Parsing

Lexicalized Parsing includes word-specific (lexical) information in Parsing.

* It improves accuracy by considering specific words in the sentence.

UNit - 4

Reasoning in NLP:

* Semantics (Meaning of words and sentences).
* pragmatics (Meaning Based on context). (sarcasm & humor)

1) **Semantics** : semantics is the study of literal meaning of words, phrases, and sentences.

* It focuses on definitions and grammer, without considering context.

Ex :

sentence : "The sky is blue."

Semantics explains that "sky" means the space above, and "blue" is a color.

* It doesn't care who is speaking (or) where they are.

Key features :

* **word meaning** : "Dog" means a four-legged animal.

Sentence meanings :-

"I ate an apple." means someone consumed a fruit.

NO context needed :-

The meaning stays the same everywhere.

2) Pragmatics? (Meaning Based on context).

* pragmatics is the study of how context affects meaning.

* The same sentence can mean different things depending on who says it, where, and how.

Example :- "can you pass the salt".

Semantics :- Asking if someone is physically able to pass the salt.

pragmatic :- A polite way of asking someone to give you the salt.

Key Features of pragmatics :

Context matters → "It's cold here" can mean "close the window".

Speaker's Intention :→ saying "Nice job!" can be sincere or sarcastic

Culture & Social Rules :- In some cultures, indirect speech is more polite.

Logical Representation,

* Logical Representation refers to the process of converting natural language text into a structured, logical form, that machines can understand and reason with.

Propositional logic

* The simplest logic
* A proposition is a statement that is either true (or) false.

Connective symbols	word	Technical term	Example
$\wedge$	AND	Conjuction	$A \wedge B$
$\vee$	OR	Disjunction	$A \vee B$.
$\rightarrow$	implies if--Then.	Implication	$A \rightarrow B$.
$\Leftrightarrow$	If and only if	Biconditional	$A \Leftrightarrow B$.
$\neg$ (or) $\sim$	Not	Negation	$\neg A$ (or) $\neg B$

Ex :- Natural Language :- It is raining (or) it is sunny.

$\Rightarrow P \vee Q$.

Where P = "It is raining", Q = "ti is sunny.

Two types of proposition

1) Atomic proposition

2) compound proposition.

=> Atomic proposition (simple proposition)

* It is a single, basic statement that does not contain any logical connectives (like AND, OR, NOT)

* It cannot be broken down further.

Ex: "The sky is blue"

* Each of these statements is independent and cannot be split into smaller logical statements.

=> compound proposition (complex proposition).

* It is formed by combining two or more atomic proposition using logical operators (AND, OR, NOT, If-Then etc).

* It can be broken down into simpler proposition.

Ex: "It is raining and the ground is wet.

$P \wedge Q$.

* If you study then you will pass.

$P \rightarrow Q$

Logical connectives.

Logical connectives. (also called logical operators) are symbols (or) words used to connect statements (proposition) to form more complex

table from propositional logic

Logical connectives.

Negation :-

A sentence such as $\neg P$ is called negation of P.

A literal can be either positive literal or negative literal.

Ex :- The sun is not cold.

P : The sun is cold.

$\neg P$

P	$\neg P$
T	F
F	T

conjuction (T and T = T)

A sentence which has $\wedge$ connective such as, $P \wedge Q$ is called a conjuction

Ex : prolog is simple and Easy

P : prolog is simple

Q : prolog is Easy

P	Q	P ∧ Q
T	T	T
T	F	F
F	T	F
F	F	F

Disjuction

A sentence which has ∨ connective, such as P ∨ Q is called disjunction, where p is proposition.

Ex: After 10th Student can take either A-Group (or) B roup !!

P = After 10th student can take A-Group

Q = After 10th student can take B-Group

P ∨ Q.

P	Q	P ∨ Q
T	T	T
T	F	T
F	T	T
F	F	False

Implication

A sentence such as $P \rightarrow Q$ is called an implication.

* Implication are also known as if then rules.

Ex: If it is raining then the street is wet

P = It is raining

Q = street is wet

$P \rightarrow Q$.

P	Q	P → Q
T	T	True
T	F	False
F	T	T
F	F	T

Bi-conditional

A sentence such as $P \Leftrightarrow Q$ is a Bi-conditional sentence.

Ex :- If I am breathing, then I am alive.

P :- I am breathing, Q :- I am alive.

$P \Leftrightarrow Q$

P and q are true (or) both p and q are false.

P	Q	P <=> Q
T	T	T
T	F	F
F	T	F
F	F	T

Ex :- if it rains, then I will stay at home.

if p then q

$p \rightarrow q$

Ex :- if $a=b$ and $b=c$ the $a=c$

$(p \wedge q) \rightarrow r$

$3^2 = 9$

P	q	r	$p \wedge q$	$(p \wedge q) \rightarrow r$
T	T	T	T	T
T	T	F	T	F
T	F	T	F	T
T	F	F	F	T
F	T	T	F	T
F	T	F	F	T
F	F	T	F	T
F	F	F	F	T

example in pdf unit 4.