



# POORNIMA

## COLLEGE OF ENGINEERING

### DETAILED LECTURE NOTES

Campus: ..... Course: .....

Class/Section: .....  
Name of Subject: .....

Date: .....  
Code: .....

#### Unit - 5 Natural language Processing

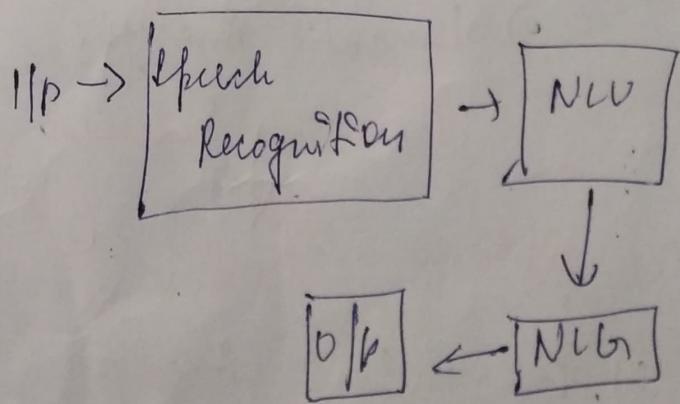
NLP stands for Natural language processing which is a part of CS, Human language and AI.

It is the technology used by machines to understand, analyze, manipulate and interpret Human language.

→ It helps developers to organize knowledge for performing tasks such as Translation, Automatic summarization, Named Entity Recognition (NER), Speech Recognition, Relationship extraction and topic segmentation.

#### Applications of NLP

- Speech Recognition
- Sentiment Analysis
- Machine translations
- Chat bots, etc



## Components of NLP

1) Natural Language Understanding : It helps the machine to understand and analyze human language by extracting the metadata from content such as concepts, entities, keywords, emotions, relations and semantic roles.

It involves the following tasks:

1) used to map the given input into useful representation

2) used to Analyze different Aspects of the language

## Challenges:

1) Lexical Ambiguity (Every word have diff meaning)

2) Syntactical Ambiguity (structure while passing should be same.)

3) Semantic Analysis

(meaning should be Appropriate)

The girl got hit by car while moving.

a) Pragmatic Analysis (context of the phrase give the multiple interpretation).

multiple meaning  
(sentiments, intonation, past experience, etc.)



# POORNIMA

## COLLEGE OF ENGINEERING

### DETAILED LECTURE NOTES

Campus: ..... Course: .....  
Name of Faculty: .....

Class/Section: .....  
Name of Subject: .....

Date: .....  
Code: .....

#### 2) Natural language Generation (NLG)

It is the process of producing meaningful phrases and sentences in the form of Natural language from some internal representation.

It involves:

- Text Planning : It includes the retrieving the relevant content from Knowledge Base.
- Sentence Planning : It includes choosing required words, forming meaningful phrases, setting tone of the sentences.
- Text Realization : Mapping Sentence Plan into sentence structure

## Applications of NLP:

- 1) Question Answering
- 2) Spam Detection
- 3) Sentiment Analysis / Opinion mining.
- 4) Machine Translation
- 5) Spelling correction
- 6) Speech Recognition
- 7) Chatbot.
- 8) Information Extraction

A  
JG

## Steps in NLP / Build an NLP Pipeline

Tokenization

↳ Sentence Segmentation  
Word Tokenization.

Stemming

1) Sentence Segmentation : It breaks the paragraph into separate sentences.

lemmatization

2) Word Tokenization : Used to break the sentence into separate words or tokens.

POS Tags

Xyz is an Almoanal human

Name Entity  
Recognition

2) Stemming : used to normalize words i.e. base form or root form.

chunking



# POORNIMA

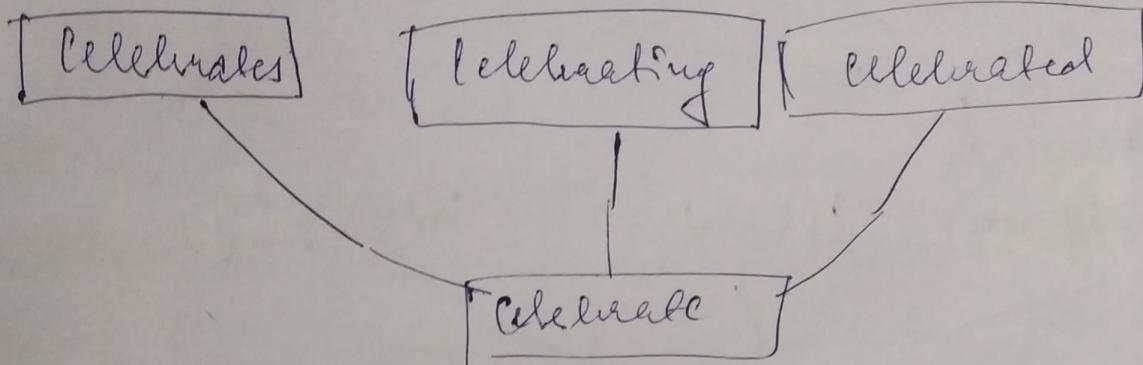
## COLLEGE OF ENGINEERING

### DETAILED LECTURE NOTES

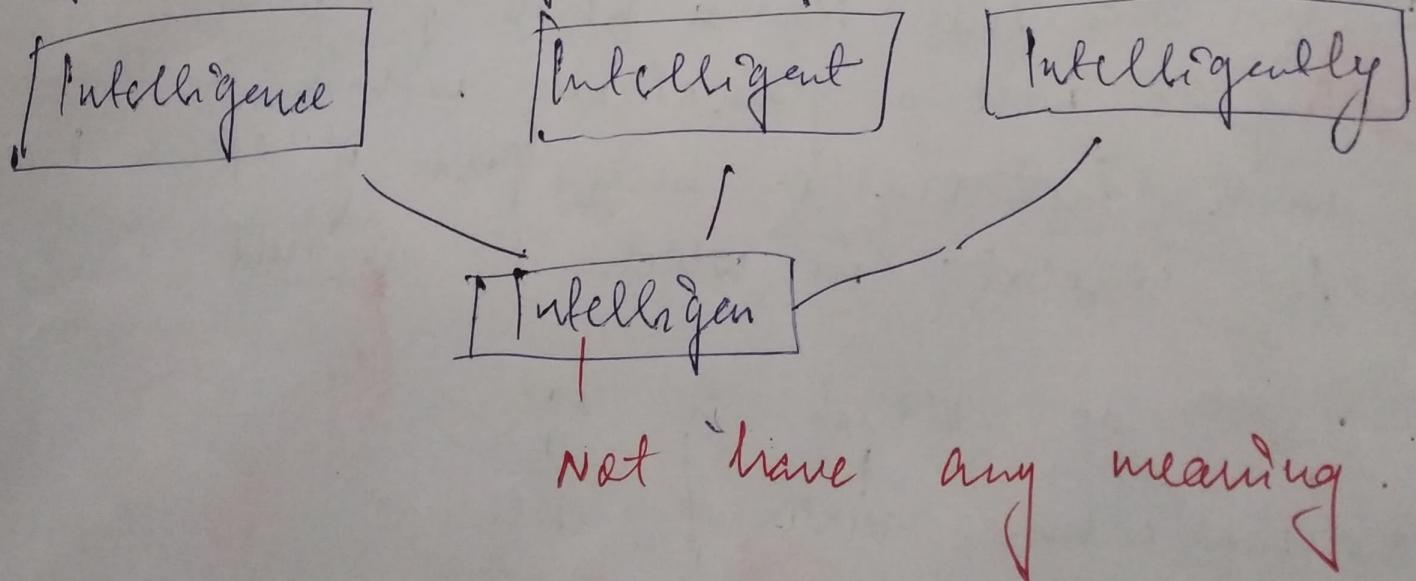
Campus: ..... Course: .....  
Name of Faculty: .....

Class/Section: .....  
Name of Subject: .....

Date: .....  
Code: .....



The big problem with learning is that sometimes it produces the root word which may not have any meaning as:





# POORNIMA

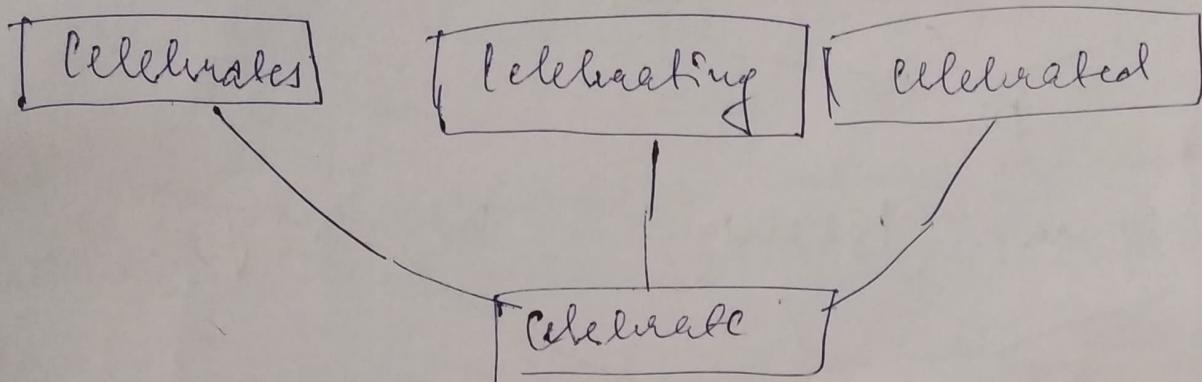
## COLLEGE OF ENGINEERING

### DETAILED LECTURE NOTES

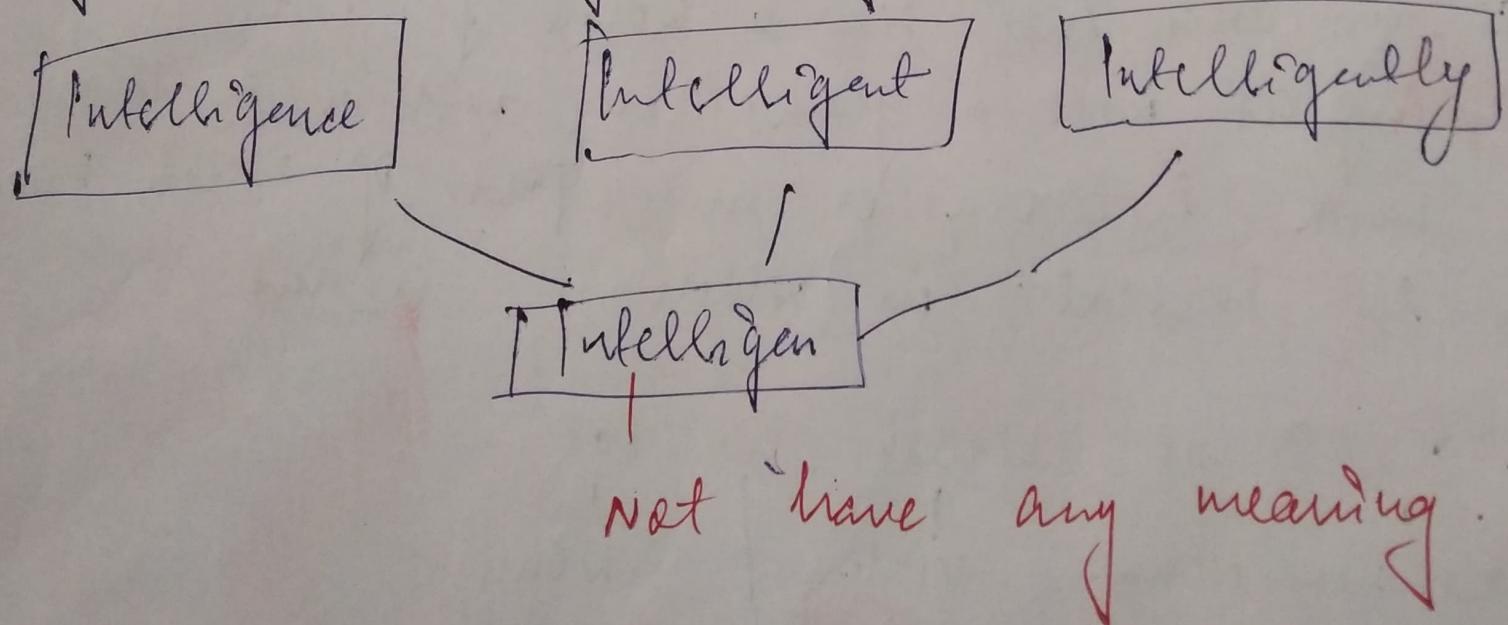
Campus: ..... Course: .....  
Name of Faculty: .....

Class/Section: .....  
Name of Subject: .....

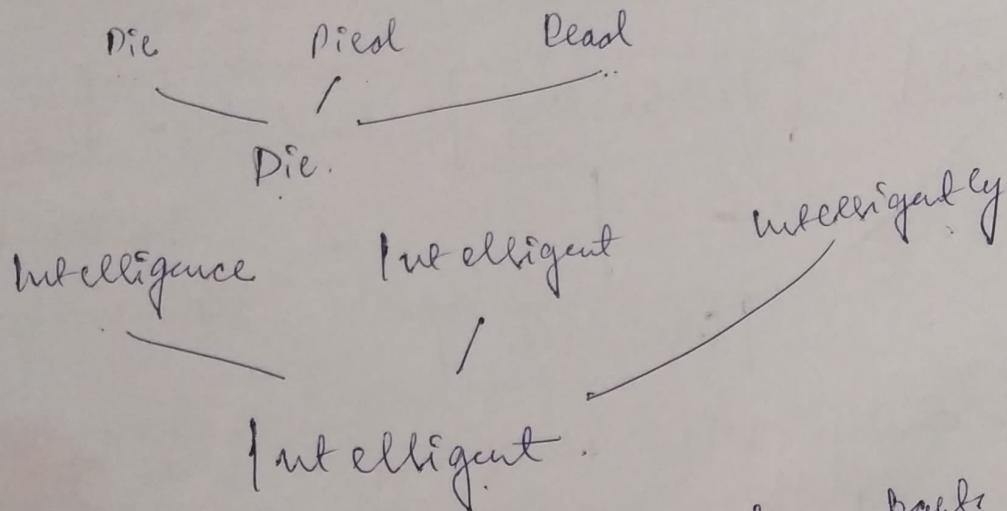
Date: .....  
Code: .....



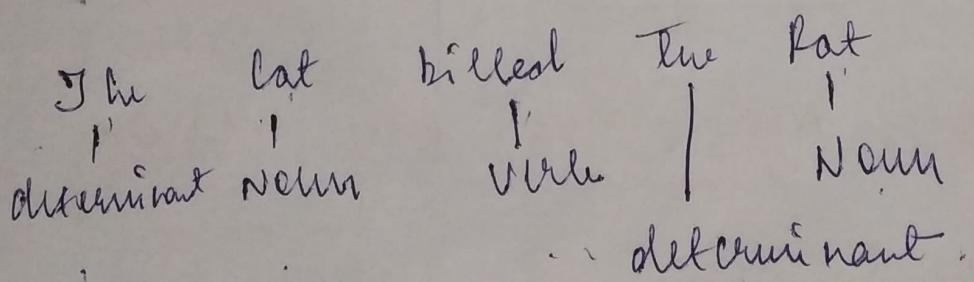
The big problem with learning is that sometimes it produces the root words which may not have any meaning as:



- 3) Lemmatisation = quite similar to stemming  
 • used to group different inflected forms of the root word, **LECTURE**  
 • called lemmas  
 • main difference → it produces the which has a meaning.



- 4) Post tags : It stands for the parts of speech, which includes Noun, Verb, adverb and Adjective. It indicates that how a word fits with its meaning as well as grammatically within the sentences. A word has one or more parts of speech based on the context in which it is used.



Whitcapp  
 noun / verb  
 me  
 when you are free



# POORNIMA

## COLLEGE OF ENGINEERING

### DETAILED LECTURE NOTES

Campus: ..... Course: .....

Name of Faculty: .....

Class/Section: .....

Name of Subject: .....

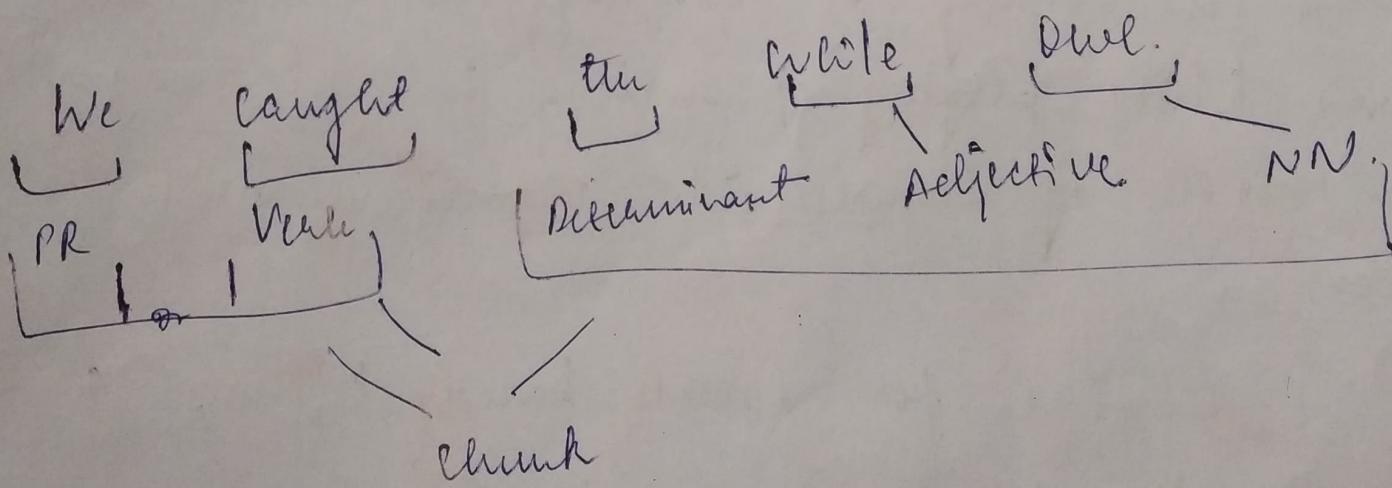
Date: .....

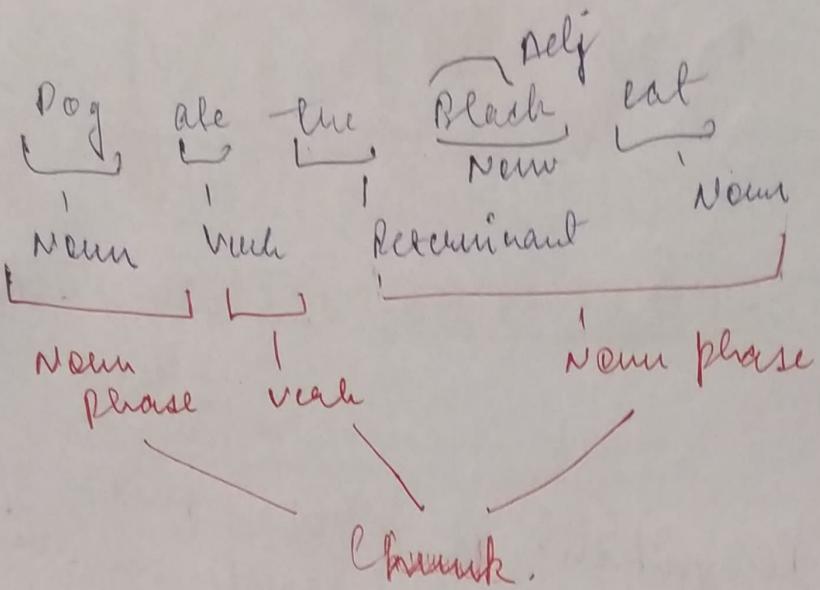
Code: .....

5) Name Entity Recognition: It is the process of detecting the Named Entity such as person name, movie name, organization name or location.

Eg: Google, Inc <sup>Organization</sup> Sunder Pichai <sup>Person</sup> introduced the new Pixel 3 at New York <sup>Location</sup> Central mall <sup>Organization</sup>

6) Chunking: Picking individual pieces of information and grouping them into bigger pieces.





Knowledge Required in NLP.

System must have knowledge about

- 1) What the words mean.
- 2) How words combine to form sentences
- 3) :

Different forms of knowledge are as follows:

### 1) Phonemic and Phonological knowledge

- It is the study of language at the level of sounds while ~~play~~ phonology is the study of combination of sounds into organized units of speech.
- They are useful for speech based systems as they deal with how words are related to the sounds that realize them.

COLLEGE OF ENGINEERING  
DETAILED LECTURE NOTES

Campus: ..... Course: .....  
Name of Faculty: .....

Class/Section: .....  
Name of Subject: .....

Date: .....  
Code: .....

2) Morphological Knowledge:

- concerns with word formation.
- Study of patterns of formations of words by the combination of sounds into minimal distinctive units of meaning called morphemes.
- Morphological knowledge concerns how words are constructed from morphemes.

My successful

3) Syntactic Knowledge:

- Syntax is the level at which we study how words combine to form phrases, phrases combine to form clauses and clauses join to make sentences.
- It deals with how words can be put together to form correct sentences.

# POC LEG PL

## 4) Semantic Knowledge:

- 1) It concerns meanings of the words and sense.
- 2) Refines the meaning of a sentence is very difficult due to ambiguities involved.

## 5) Pragmatic Knowledge

- Pragmatics is the extension of the meaning or semantics.
- It deals with the contextual aspects of meaning in particular situations.
- Concerns how sentences are used in different situations.

## 6) Discourse Knowledge

- Discourse concerns connected sentences. It is a study of chunks of language which are bigger than a single sentence.
- Discourse language concerns inter-sentential links i.e. how the immediately preceding sentences affect the interpretation of the next sentence.
- It is important for interpreting sentences by temporal aspects of the information conveyed.



# POORNIMA

COLLEGE OF ENGINEERING

## DETAILED LECTURE NOTES

Campus: ..... Course: .....  
Name of Faculty: .....

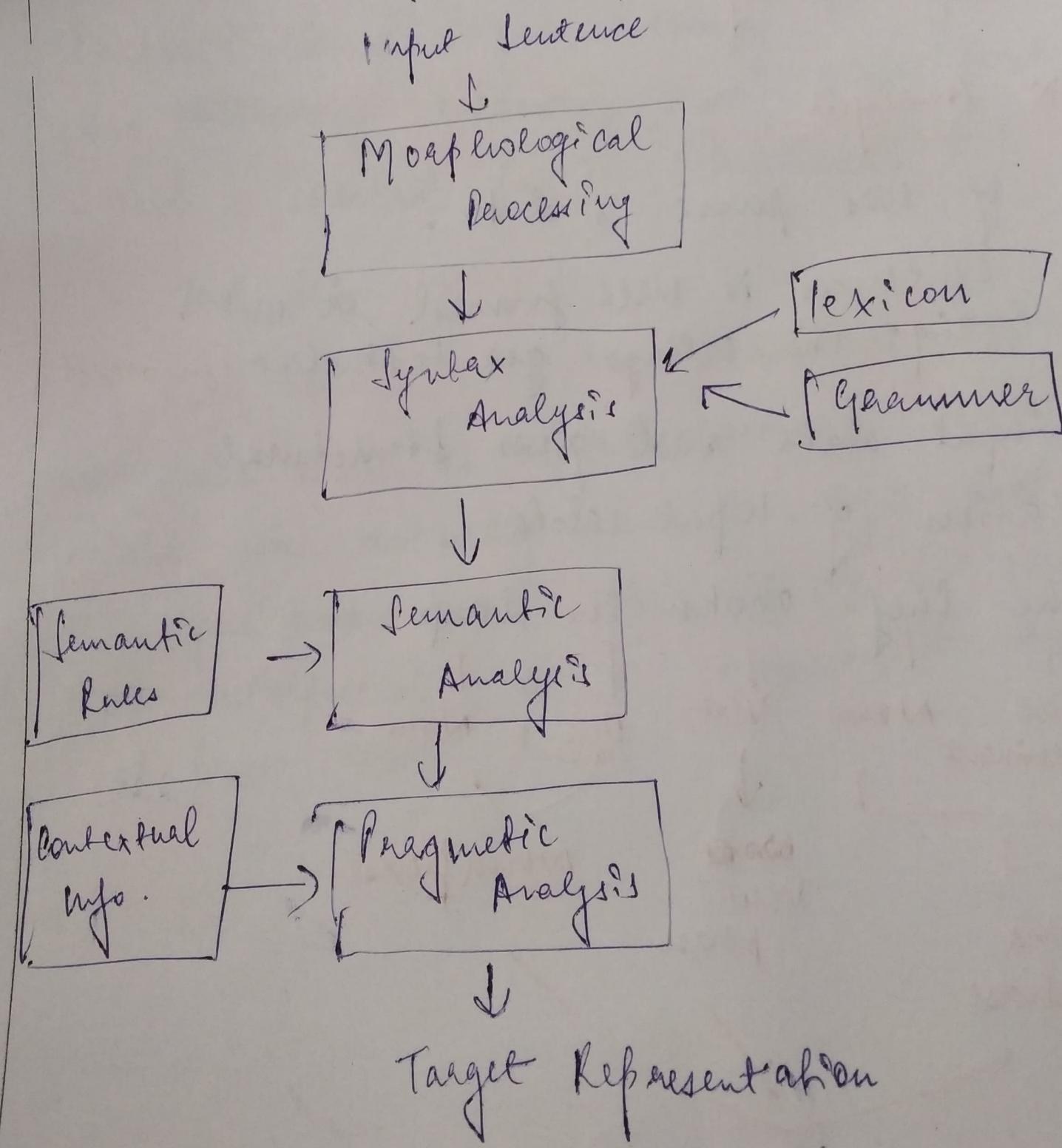
Class/Section: .....  
Name of Subject: .....

Date: .....  
Code: .....

### 7) World Knowledge:

- It is Noticing but everyday knowledge that all speakers share about the world.
- Includes General knowledge about the structure of the world by what each language user must know about the other user's beliefs and goals.
- Essential to make language understanding much better.

## Phrasal levels of NLP



## Morphological Processing

- Purpose: Break tokens of language input into sets of tokens corresponding to Paragraphs, sentences and words

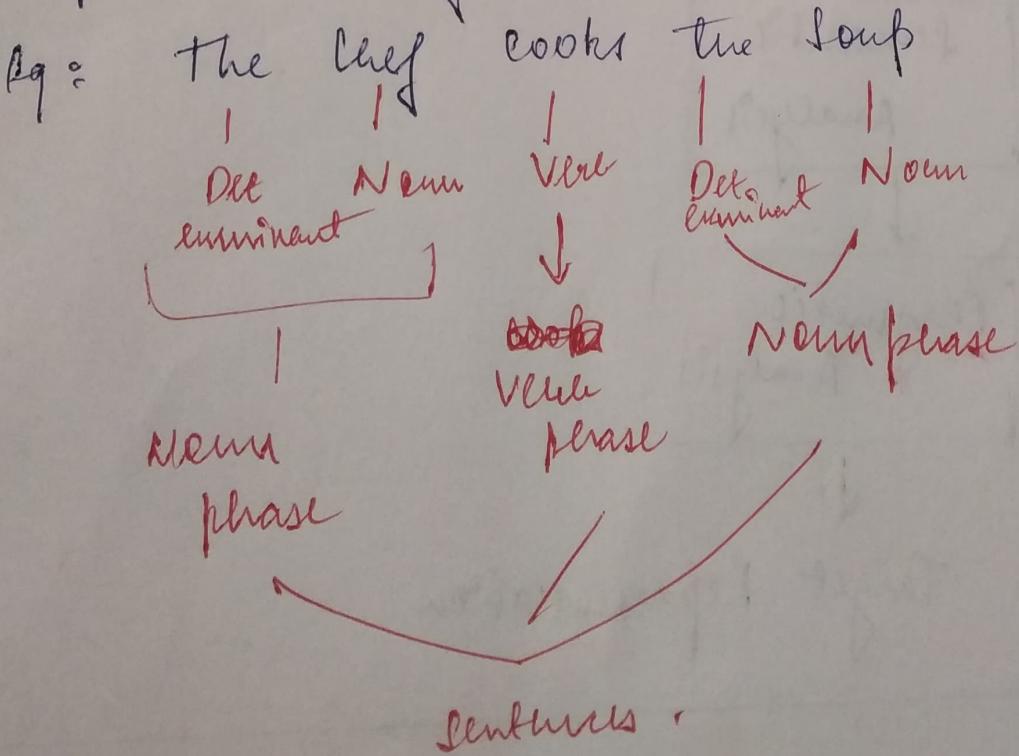
Example      un easy  
                /    \  
        un      easy

## Syntax Analysis

Purpose of this phase is to:

- 1) check sentences is well formed or not.  
Eg: The collage goes to Student (incorrect)

- 2) Takes input data and gives structural representation of input data.



# DETAILED LECTURE NOTES

Campus: ..... Course: .....

Class/Section: .....  
Name of Subject: .....

Date: .....  
Code: .....

## Semantic Analysis

Purpose: Draw exact meaning or you can say dictionary meaning from given text.

- Text is checked for its inference.

## Not ice (Wrong)

Pragmatic Analysis: The feature representing what was said is interpreted to determine what was actually meant. e.g.: the sentence "Do you know what time it is?" should be interpreted as a request to be told at that time.