

## ARTIFICIAL INTELLIGENCE FOR ENGINEERING

### UNIT-3 : Natural Language Processing (NLP)

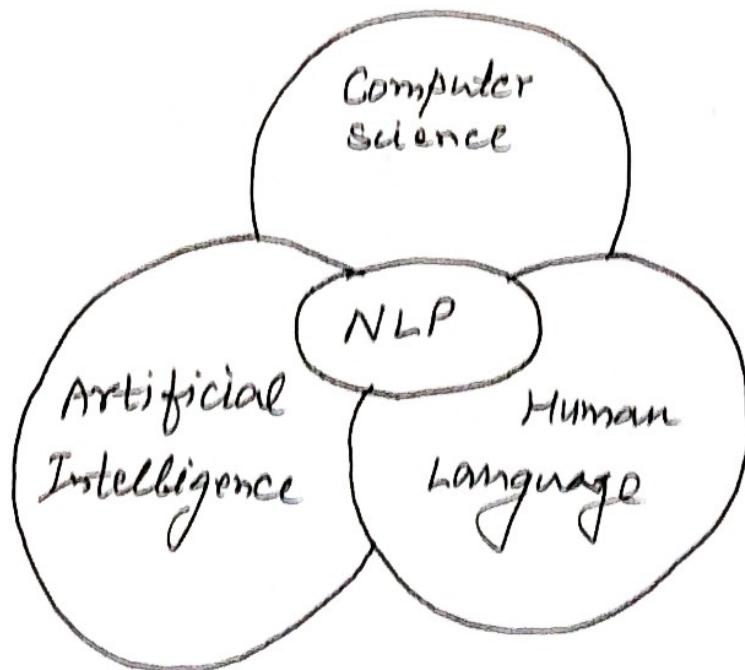
#### LECTURE - 11

Content : Natural Language Processing

Lecture Outcomes : Students will be able to learn the concept of NLP.

### Natural Language Processing (NLP)

Natural language processing is a part of computer science and artificial intelligence which deals with human language.



or

Natural Language Processing, which is a part of computer science, Human language and Artificial Intelligence. It is the technology that is used to machines to understand, analyse, manipulate and interpret human's languages.

## Applications of NLP

### 1. Sentiment Analysis

Sentiment analysis is also known as opinion mining. It is used on the web to analyze the attitude, behaviour and emotional state of the sender.



### 2. Speech Recognition

Speech recognition is used for converting spoken words into text. (Siri, google assistant, cortona etc.)

### 3. Chatbot

Chatbots use NLP to convert human language into desirable actions.

#### 4. Machine Translation

Machine translation is used to translate text or speech from one natural language to another natural language. Example: Google Translator.

#### 5. Spelling correction

Microsoft corporation provides word processor software like MS-word, PowerPoint for the spelling correction.

#### 6. Information Extraction

Information extraction is one of the most important applications of NLP. It is used for extracting structured information from unstructured or semi-structured machine readable documents.

#### 7. Spam Detection

Spam detection is used to detect unwanted e-mails getting to a user's inbox.



#### 8. Advertisement Matching

Advertisement matching uses NLP to recommend ads based on your history.

## NLP Terminologies

### Step-1 : Sentence Segmentation

Sentence segmentation is the first step for building the NLP pipeline. It breaks the paragraph into separate sentences.

### Step-2 : Tokenization

Tokenizer is used to break the sentence into separate words or tokens.

- Break a complex sentence into words
- Understand the importance of each of the words with respect to the sentence.
- produce a structural description on an input-sentence.

Example: The dog killed the bat.

The    dog    killed    the    bat

### Tokenization

### Step-3 : Stemming

stemming is used to normalize words into its base form or root form.

Example: celebrates, celebrated, celebrating



The big problem with stemming is that sometimes it produces the root word which may not have any meaning.

For Example: intelligence, intelligent, intelligently

intelligen → do not have any meaning.

Step-4: Lemmatization:

Lemmatization is quite similar to the stemming. It is used to group different inflected forms of the word, called Lemma. The main difference between stemming and lemmatization is that it produces the rootword, which has meaning.

For example: A Lemmatizer should map gone, going and went into go.

Step-5: Stop words

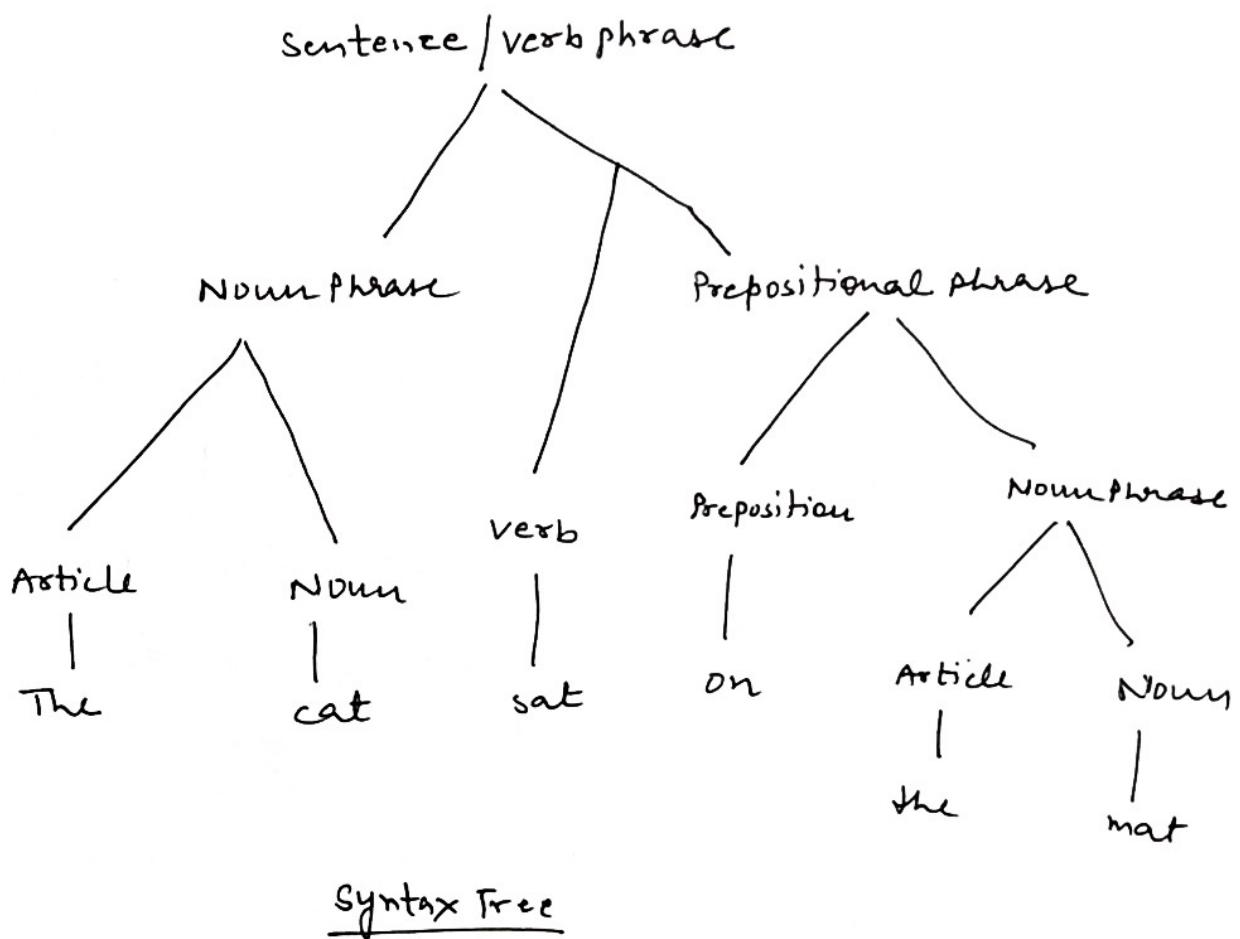
There are a lot of words that appear very frequently like however, just, if, before, Although, Really etc..

NLP will flag these words as stop words.

stop words might be filtered out before doing any statistical analysis.

## Step-6 : Syntax Tree ( Parsing )

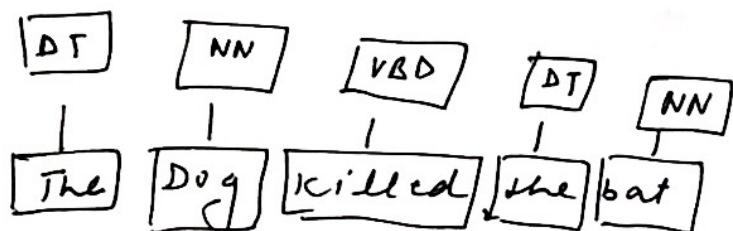
Syntax tree is a tree representation of syntactic structure of sentences or strings.



## Step-7 (Parts of speech : pos tags)

POS stands for parts of speech, which includes Noun, verb, adverb and adjective. It indicates that how a word functions with its meaning as well as grammatically with in the sentence.

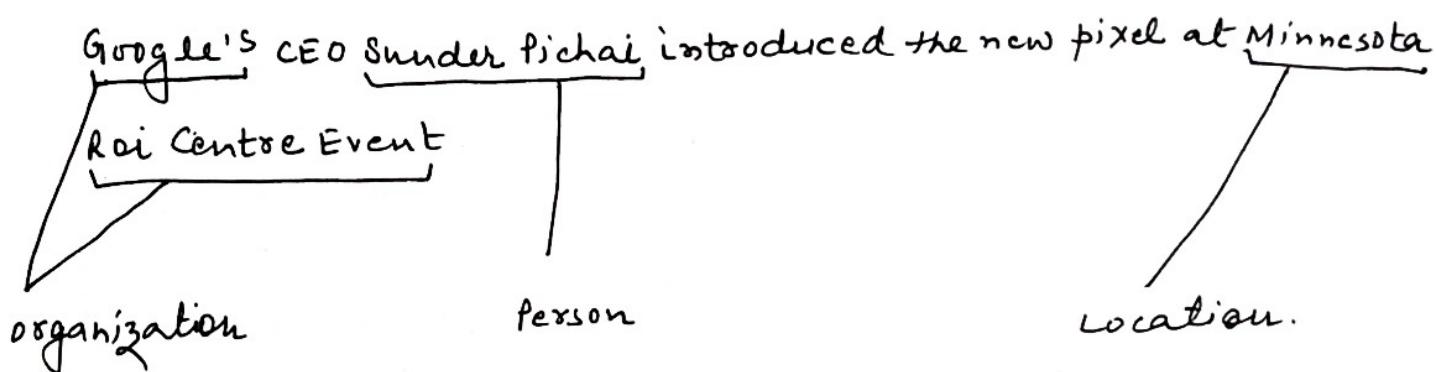
Example :



## Step - 8 : (Named Entity Recognition - NER)

NER is the process of detecting the named entity such as :

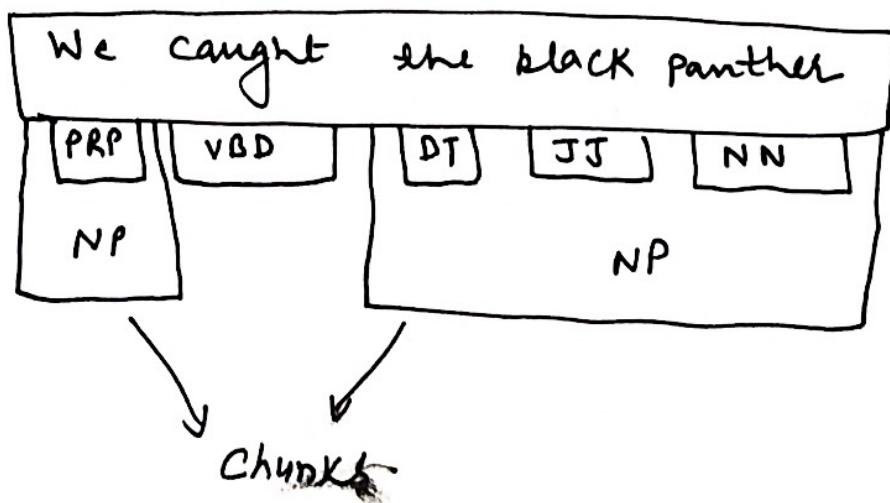
- Movie
  - Monetary Value
  - Organization
  - Location
  - Quantities
  - Person



## Step - q : Chunking :

Picking up individual pieces of information and grouping them into bigger pieces.

Example :



# ARTIFICIAL INTELLIGENCE FOR ENGINEERING

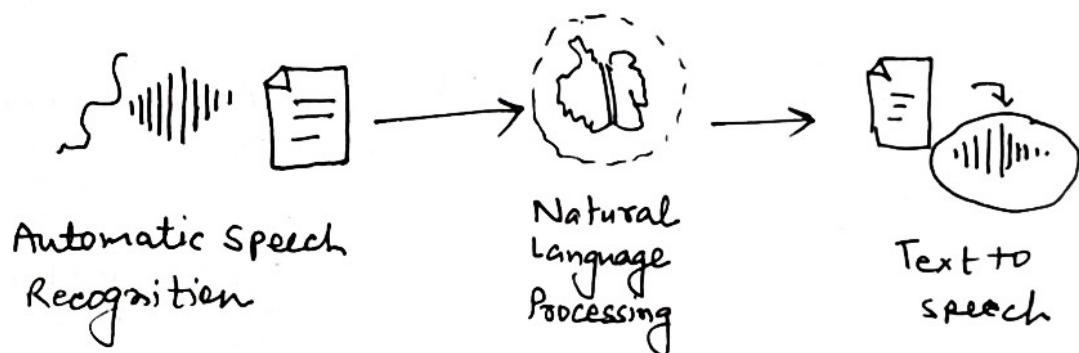
## UNIT-3: Natural Language Processing

### LECTURE-12

Content: Speech Recognition

Lecture Outcomes: Students will be able to learn the concept of speech recognition.

### Speech Recognition



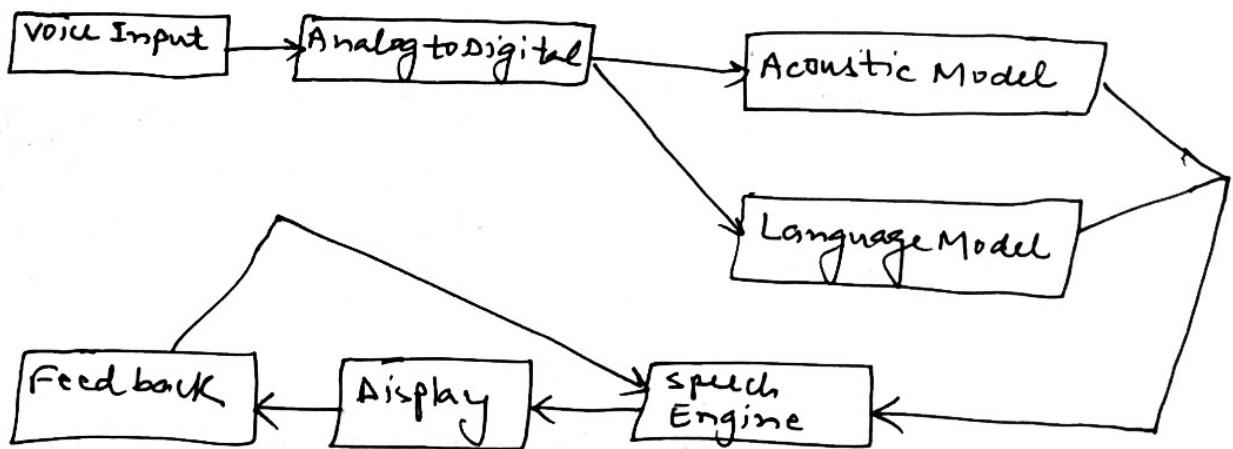
Speech recognition is the process ~~of~~ that enables a computer to recognize and respond to spoken words and then converting them in a format that the machine understands. The machine may then convert it into another form of data depending on the end goal.

For Example: Google dictate and other transcription programs use speech recognition to convert your spoken words into text while digital assistants like Siri and Alexa respond in text format or voice.

speech processing system has mainly three tasks:

- speech recognition that allows the machine to catch the words, phrases and sentences we speak.
- Natural language processing to allow the machine to understand what we speak.
- speech synthesis to allow the machine to speak.

### Speech Recognition Process



- Speech recognition works using algorithms through acoustic and language modeling. Acoustic modeling represents the relationship between linguistic units of speech and audio signals ; language modeling matches sounds with word sequences to help distinguish between words that sound similar.
- Often Hidden Markov models are used as well to recognize temporal patterns in speech to improve accuracy within the system.

- Other methods used in speech recognition may include natural language processing (NLP) or N-grams. NLP makes the speech recognition process easier and take less time.

## Application

- The most frequent applications of speech recognition within the enterprise include the use of speech recognition in mobile devices. For example, individuals can use this functionality in smartphones for call routing, speech to text processing, voice dialling, and voice search. Speech recognition on iPhones, for example, is tied to other functions like the keyboard and Siri.
- Speech recognition can also be found in word processing application like Microsoft Word, where users can dictate what they want to show up as text.

## ARTIFICIAL INTELLIGENCE FOR ENGINEERING

### UNIT-3 : Natural Language Processing

#### LECTURE -13

content : component of NLP, Natural language understanding (NLU)

lecture Outcomes : students will be able to know about the components of NLP.

### Components of Natural Language Processing

There are two components of NLP as given

- i) Natural Language Understanding (NLU)
- ii) Natural Language Generation (NLG)

Here in this lecture we will discuss about NLU only.

#### 1. Natural Language Understanding (NLU)

- Mapping the given input in natural language into useful representations .
- Analyzing different aspects of the language .

Natural language understanding (NLU) is a subtopic of natural language processing which focuses on machines understanding human language. Interesting applications include text categorization, machine translation and question answering.

NLU makes it possible for machines to understand the overall context and meaning of natural language

NLU is used in natural language processing (NLP) tasks like topic classification, language detection and sentiment analysis.

- Sentiment analysis automatically interprets emotions within a text and categorizes them as positive, negative, or neutral.
- Language detection automatically understands the language of written text.
- Topic classification is able to understand natural language to automatically sort texts into predefined groups or topics.

NLP: Ambiguity or NLU: Ambiguity

- Lexical Ambiguity
- Syntactic Ambiguity
- Referential Ambiguity

## 1. Lexical Ambiguity

- She is looking for a match.
- The fisherman went to the bank.

## 2. Syntactic Ambiguity :

- The chicken is ready to eat.
- Visiting relatives can be boring.
- I saw the man with the binoculars.

## 3. Referential Ambiguity

- The boy told his father the theft. He was very upset.

## ARTIFICIAL INTELLIGENCE FOR ENGINEERING

### UNIT-3: Natural Language Processing

#### LECTURE-14

Content : Natural language Generation

Lecture Outcomes : students will be able to learn the concept of NLG .

The second components of NLP is natural language generation (NLG)

#### 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 retrieving the relevant content from knowledgebase .
- Sentence Planning - It includes choosing required words , forming meaningful phrases , setting tone of the sentence .
- Text Realization - It is mapping sentence plan into sentence structure .

## Applications

- Retail & Wholesale
- Banking & Finance
- Manufacturing
- Media

## Difference between NLU and NLG

NLU	NLG
<ul style="list-style-type: none"><li>• NLU is the process of reading and interpreting language</li><li>• It produces non-linguistic outputs from natural language inputs.</li></ul>	<ul style="list-style-type: none"><li>• NLG is the process of writing or generating language</li><li>• It produces constructing natural language outputs from non-linguistic inputs</li></ul>

## Phases of NLP

There are the following five phases of NLP

### 1. Lexical Analysis and Morphological

This phase scans the source code as a stream of characters and converts it into meaningful lexemes. It divides the whole text into paragraphs, sentences, and words.

### 2. Syntactic Analysis (Parsing)

Syntactic analysis is used to check grammar, word arrangements and shows the relationship among the words.

Example: Agra goes to the program.

### 3. Semantic Analysis

Semantic analysis is concerned with the meaning representation. It mainly focuses on the literal meaning of words, phrases and sentences.

### 4. Discourse Integration

Discourse integration depends upon the sentences that precede it and also invokes the meaning of the sentences that follow it.

## 5. Paragmatic Analysis

Paragmatic is the fifth and last phase of NLP.  
It helps you to discover the intended effect  
by applying a set of rules that characterize  
cooperative dialogues.

for Example:

"Open the door" is interpreted as a request  
instead of an order.

# ARTIFICIAL INTELLIGENCE FOR ENGINEERING

## UNIT-3: Natural Language Processing

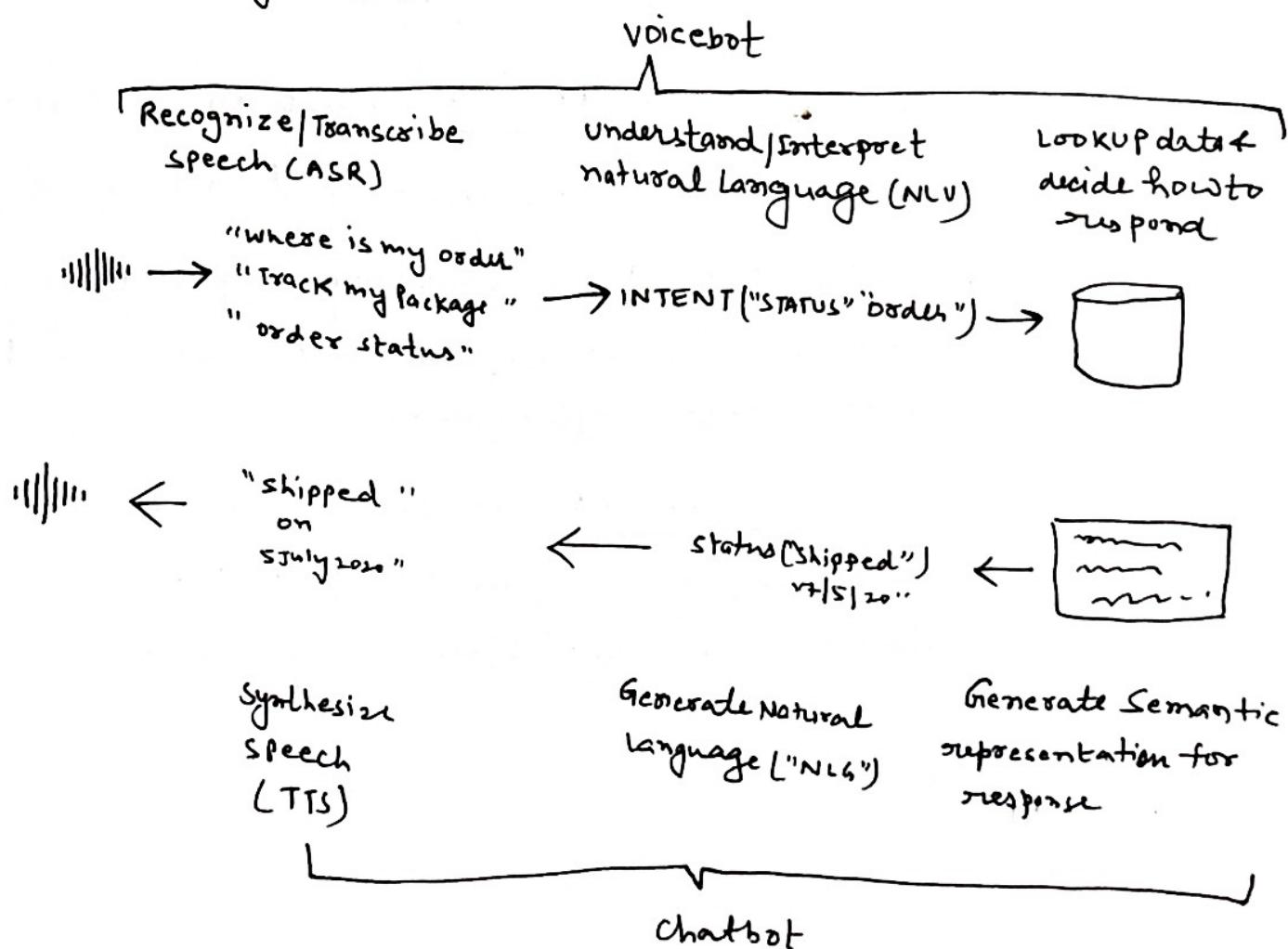
### LECTURE-15

Content: Chatbots

Lecture Outcomes: Students will learn the basic structure of chatbots.

#### Introduction to chatbots

Chatbots are not a recent development. They are simulations which can understand human language, process it and interact back with humans while performing specific task.



## Types of chatbots

There are many types of chatbots can be majorly classified as follows:

- Text-based chatbot: In a text-based chatbot, a bot answers the user's questions via text interface.
- Voice-based chatbot: In a voice or speech based chatbot, a bot answers the user's questions via a human voice interface.

There are mainly two approaches used to design the chatbots, described as follows:

- In a Rule-based Approach, a bot answers questions based on some rules on which it is trained on. The rules defined can be very simple to very complex. The bot can handle simple queries but fail to manage complex ones.
- Self-learning Approach are the ones that use some machine learning based approaches and are definitely more efficient than rule based bots.

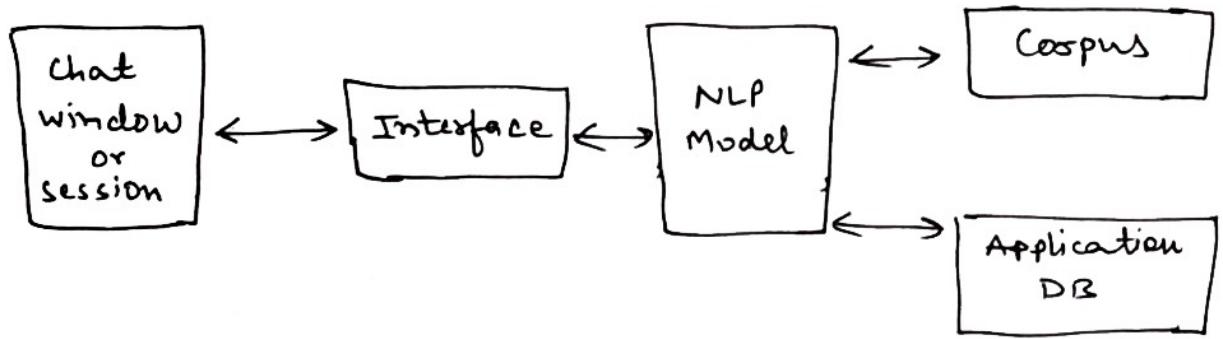
There are many types of chatbots available depending on the complexity, a few of them can be majorly classified as follows:

- Traditional chatbots: Traditional chatbots are driven by system and automation, mainly through scripts with minimal functionality and the ability to maintain only system context.
- Current chatbots: Current chatbots are driven by back and forth communication between the system and humans. They have the ability to maintain both system and task context.
- Future chatbots: Future chatbots can communicate at multiple levels with automation at the system level. They have the ability to maintain the system, task and people context.

### Architecture of Chatbots

Typical chatbot architecture should consist of the following:

- Chat window/ session/ frontend application interface
- The deep learning model for NLP
- Corpus or training data for training the NLP model.
- Application database for processing actions to be performed by the chatbot.



## Different chatbots

- Virtual reception assistant
- Virtual help desk assistant
- Virtual home assistant [Google Home]
- Virtual entertainment assistant [Amazon Alexa]
- Virtual phone assistant [Apple Siri]

## ARTIFICIAL INTELLIGENCE FOR ENGINEERING

### UNIT-3 : Natural Language Processing

#### LECTURE-16

Content : Machine Translation

Lecture Outcomes : Students will be able to know the concept of machine translation.

#### Machine Translation

Machine translation is the task of automatically converting one natural language into another, preserving the meaning of input text, and producing fluent text in the output language.

There are many challenging aspects of MT :

- 1). The large variety of languages, alphabets and grammars.
- 2). The task to translate a sequence.
- 3). There is no one correct answer.

Three major approaches of machine translation are :

- Rule-based Machine Translation (RBMT) 1970 - 1990
- Statistical Machine Translation (SMT) 1990 - 2010
- Neural Machine Translation (NMT) 2014 -

## 1. Rule-based Machine Translation

A rule based system requires expert's knowledge about the source and the target language to develop syntactic, semantic and morphological rules to achieve the translation.

### Advantages :

- No bilingual text required
- Domain independent
- Reusability (existing rules of languages - can be transferred when paired with new languages)

### Disadvantages :

- Requires good dictionaries
- Manually set rules

## 2. Statistical Machine Translation (SMT)

This approach uses statistical models based on the analysis of bilingual text corpora.

### SMT examples :

- Google Translate (between 2006 and 2016, when they announced to change to NMT)
- Microsoft Translator (in 2016 changed to NMT)

### Advantages :

- Less manual work from linguistic experts
- One SMT suitable for more language pairs
- Less out of dictionary translation with the right language model, the translation is more fluent.

### Disadvantages :

- Requires bilingual corpus .
- specific errors are hard to fix .
- Less suitable for language pairs with big differences in word order .

### 3. Neural Machine Translation (NMT)

The neural approach uses neural networks to achieve machine translation. Compared to the previous models, NMTs can be built with one network instead of a pipeline of separate tasks.

A problem with neural network occurs if the training data is unbalanced, the model can not learn from the rare samples as well as frequent ones.

## NMT Examples

- Google Translate (from 2016)
- Microsoft Translate (from 2016)
- Translation from Facebook

## Advantages

- End-to-end models (no pipeline of specific tasks)

## Disadvantages

- Requires bilingual corpora
- Rare word problem