

# Introduction to Natural Language Processing

M Yogi Reddy  
Assistant Professor  
CSE Department  
School Of Technology  
GITAM(Deemed to be ) University  
Hyderabad

# Agenda

- Natural Language Processing
- Natural Language Understanding
- Sentiment Analysis
- Segmentation and recognition

## Why NLP?

- Natural language processing helps computers communicate with humans in their own language and scales other language-related tasks.
- For example, NLP makes it possible for computer programs that understood text or speech.

# Let's us understand Language

- Method of communication



YES



NO



THANK YOU



SORRY



HELLO



I LOVE YOU



GOODBYE



PLEASE

- Natural Language

- Human Language
- Ex: Telugu, Hindi, English, French....

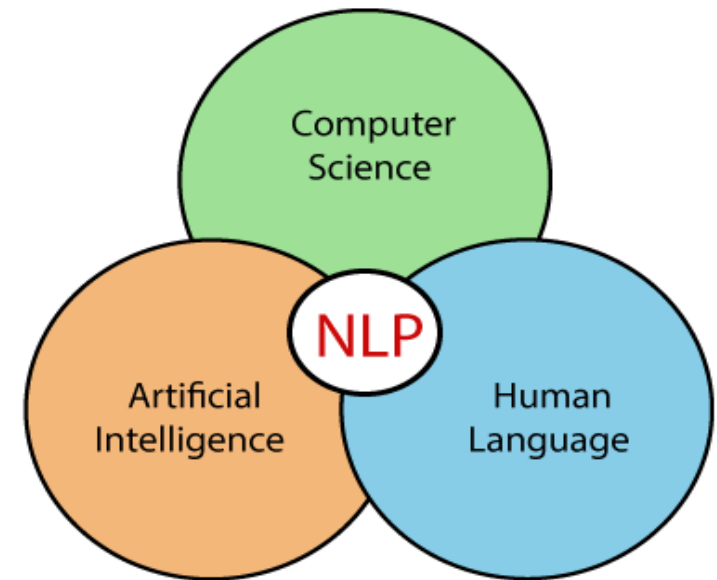
- Computer Language

- C, Fortran, Python....

■ What is Natural Language Processing ?

# Natural Language Processing

- NLP is a field of AI that gives the machines the ability to read, understanding and derive the meaning from human language
- Also known as Computational Linguistics (CL), Human Language Technology (HLT), Natural Language Engineering (NLE)
- **NLP = AI + Computational Linguistics**

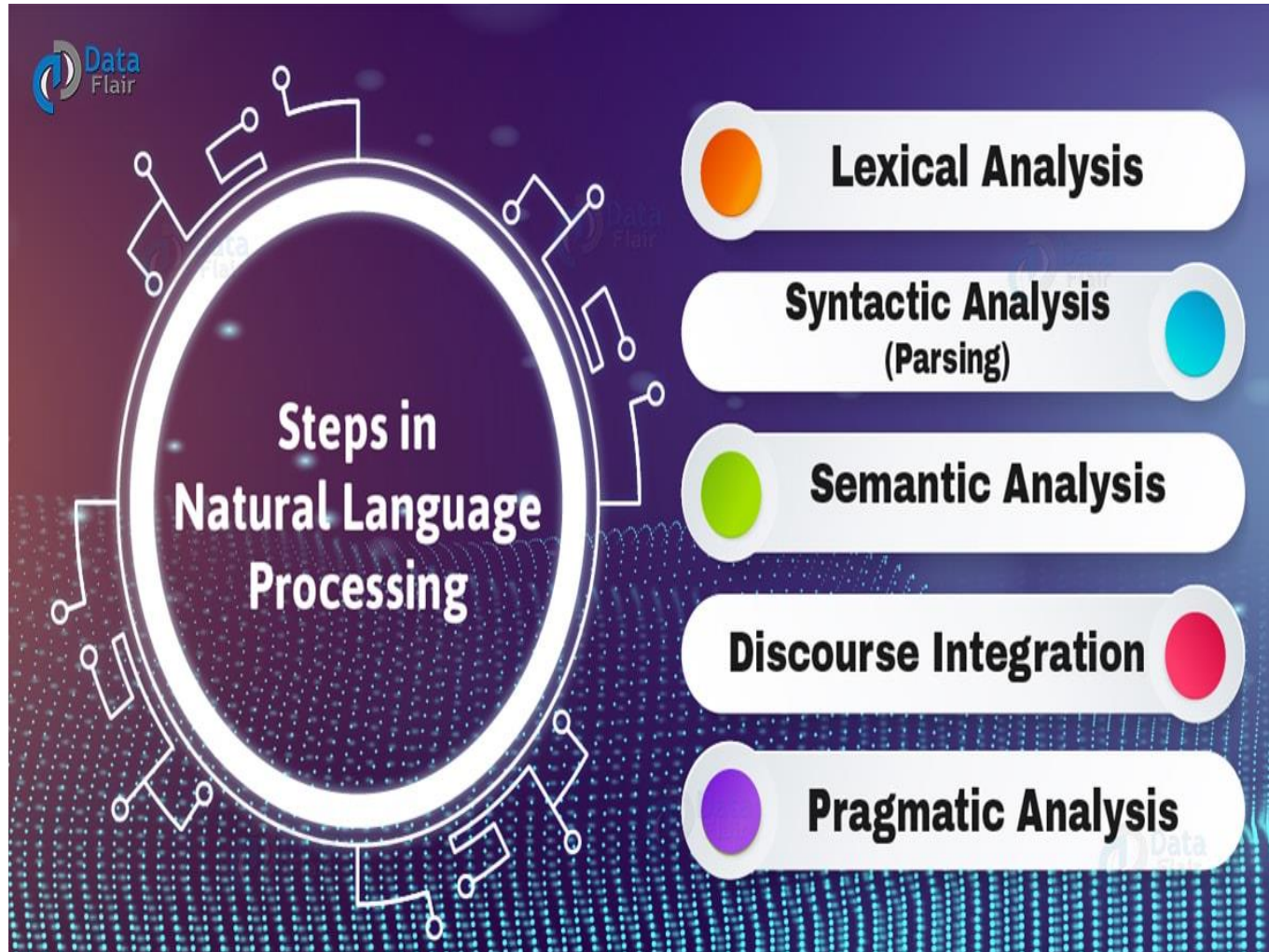


# Linguistics and language

- Linguistics is the science of language
- Its study includes:
  - Sounds which refers to phonology
  - Word formation refers to morphology
  - Sentence structure refers to syntax
  - Meaning refers to semantics
  - Understanding refers to pragmatics



# Steps in NLP



= Tokenization

= Stemming

= Lemmatization

= POS Tags and NER

= Chunking

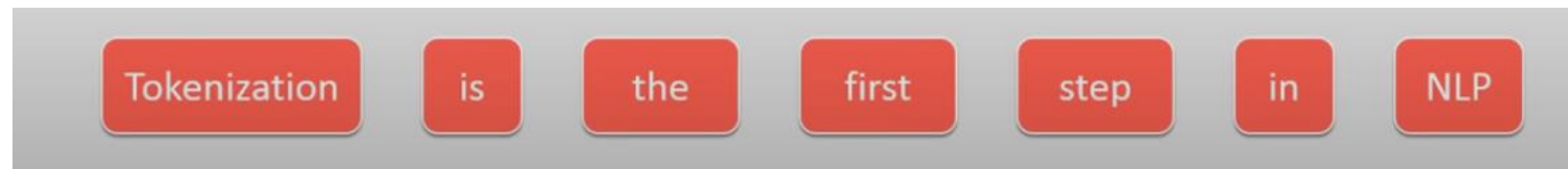


# Morphological /Lexical Analysis

- The first phase of NLP is the Lexical Analysis. 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.
- **Tokenization** is a way of separating a piece of text into smaller units called tokens.
- Here, tokens can be either words, characters, or subwords.



For Example:



# Syntactic Analysis

- Syntactic Analysis is used to check grammar, word arrangements, and shows the relationship among the words.
- Eg. “the girl the go to the school”. This would definitely be rejected by the English syntactic analyzer
- **Stemming** is basically removing the suffix from a word and reduce it to its root word.
- For example: “**Flying**” is a word and its suffix is “**ing**”, if we remove “**ing**” from “**Flying**” then we will get base word or root word which is “**Fly**”.



Affectation

Affects

Affections

Affected

Affection

Affecting

# Semantic Analysis

- Semantics concerns the (literal) meaning of words, phrases, and sentences
- This abstracts the dictionary meaning or the exact meaning from context
- The structures which are created by the syntactic analyzer are assigned meaning
- E.g.. “colorless blue idea” .This would be rejected by the analyzer as colorless blue do not make any sense together
- **Lemmatization**, on the other hand, takes into consideration the morphological analysis of the words.

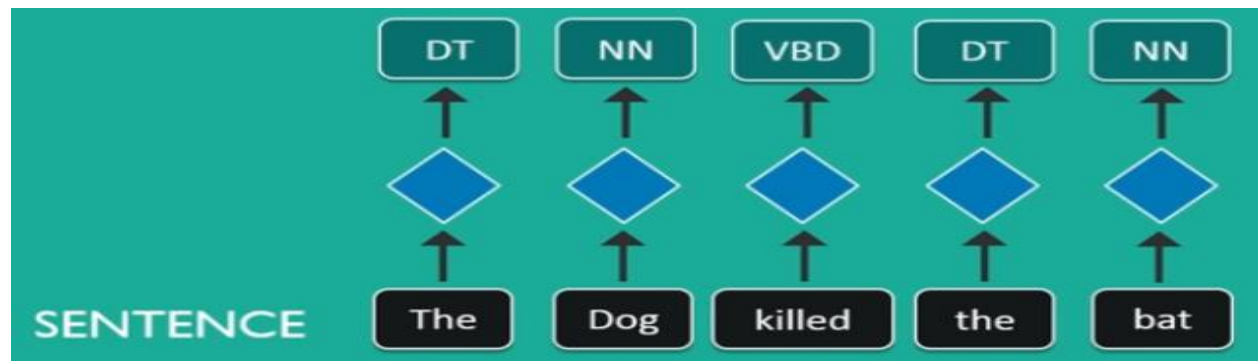
| Form     | Morphological information   | Lemma |
|----------|---|-------|
| studies  | Third person, singular number, present tense of the verb <b>study</b> | study |
| studying | Gerund of the verb <b>study</b>                                       | study |



# Discourse Integration



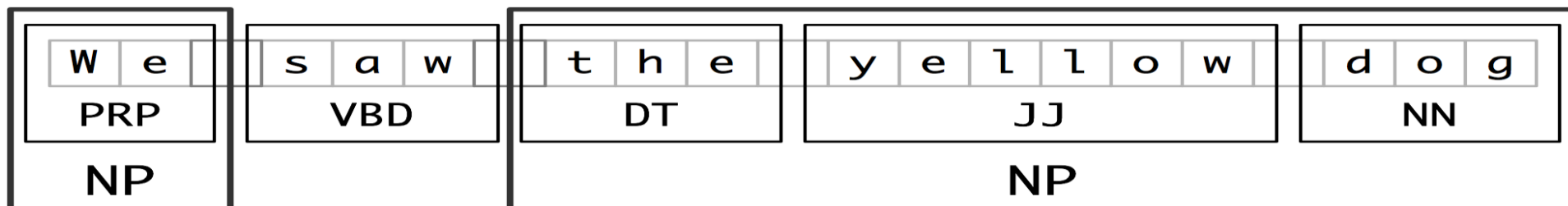
- Sense of the context
- The meaning of any single sentence depends upon the sentences that precedes it and also invokes the meaning of the sentences that follow it
- E.g. the word “it” in the sentence “she wanted it” depends upon the prior discourse context
- The tag in case of is a **Part-of-Speech(POS)** tag, and signifies whether the word is a noun, adjective, verb, and so on.



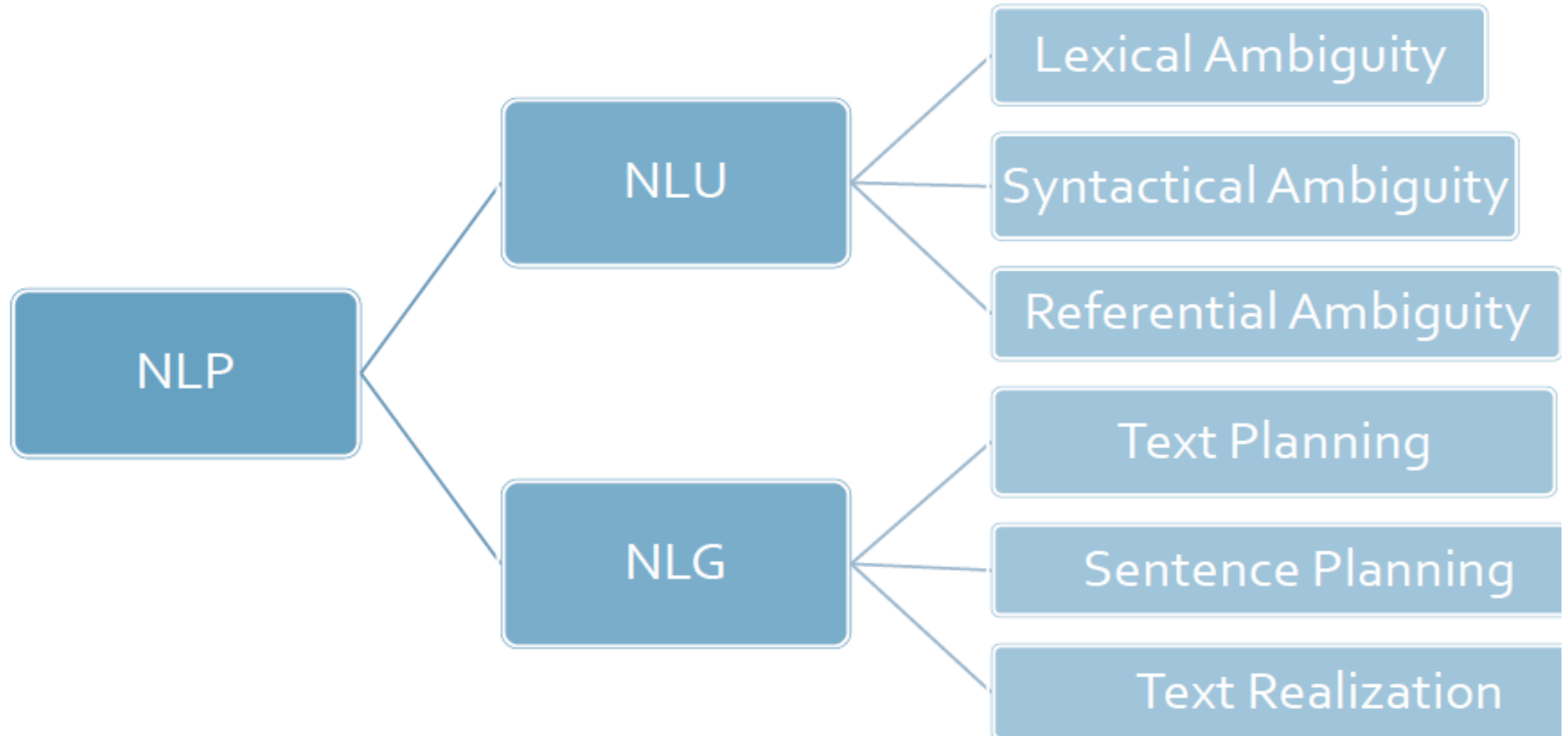
# Pragmatic Analysis



- Pragmatics concerns the overall communicative and social context and its effect on interpretation
- It means abstracting or deriving the purposeful use of the language in situations
- Importantly those aspects of language which require world knowledge
- E.g. “close the window?” should have been interpreted as a request rather than an order
- **Chunking**, one of the important processes in natural language processing, is used to identify parts of speech (POS) and short phrases.
- Chunking can break sentences into phrases that are more useful than individual words and yield meaningful results.



# Components of NLP





# Natural language Understanding

- Mapping the given input in natural language into useful representations.
- Analyzing different aspects of the language.
- Challenges in NLU :
  1. Lexical Ambiguity
  2. Syntactical Ambiguity
  3. Referential Ambiguity

# Lexical Ambiguity

- Lexical Ambiguity exists in the presence of two or more possible meanings of the sentence within a single word.

- **Example:**

Sai is looking for a **match**.

In the above example, the word match refers to that either Sai is looking for a **partner** or Sai is looking for a **match**(Cricket or other match).

- POS –can resolve Lexical Ambiguity

# Syntactical Ambiguity

- Syntactic Ambiguity exists in the presence of two or more possible meanings within the sentence.
- It is also termed as grammatical ambiguity
- **Example:**

The chicken is **ready to** eat

In the above example, **Is the chicken ready to eat his food? Or chicken is ready for someone.**

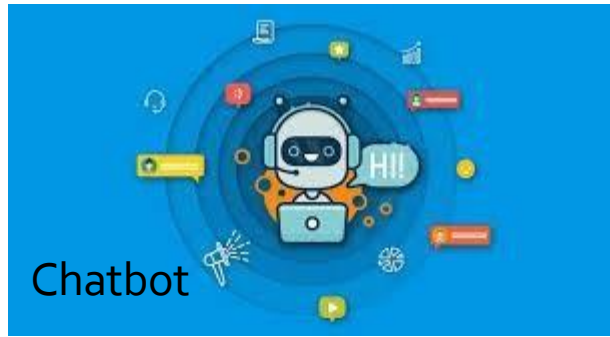
## Referential Ambiguity

- Referential Ambiguity exists when you are referring to something using the pronoun.
- **Example:**

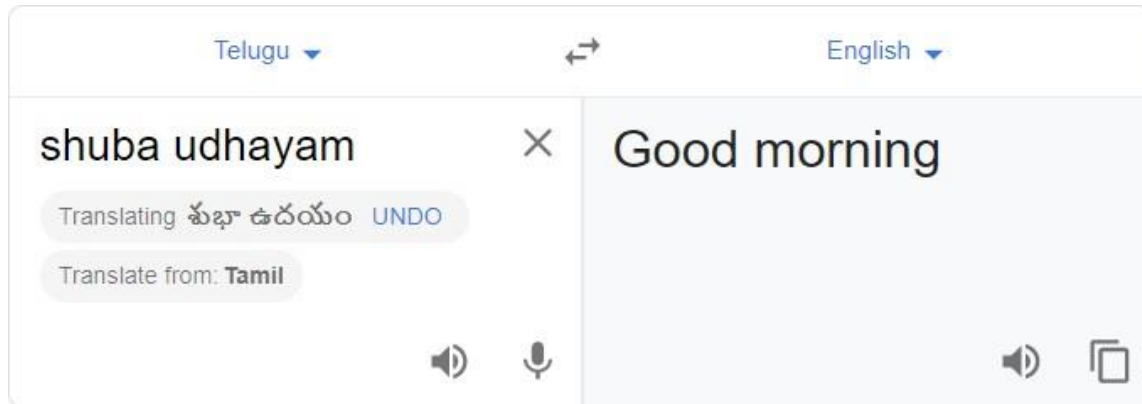
Kiran went to Sunita. She said, "I am hungry."

In the above sentence, you do not know that who is hungry, either Kiran or Sunita.

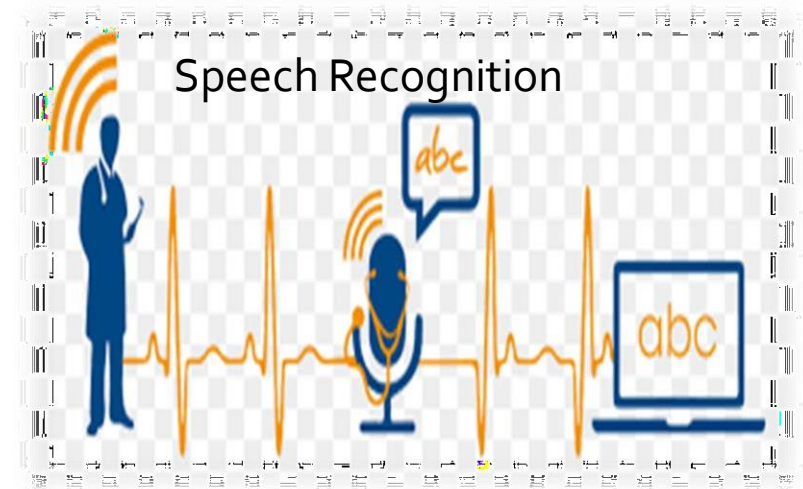
# Applications of NLP



Sentiment Analysis



Language Translation



## Available Tools For NLP

- NLTK – Natural Language toolkit
- Wolfram
- Apache OpenNLP
- Stanford Core NLP
- GATE – General Architecture for Text Engineering



# Sentiment Analysis

- Sentiment analysis is a machine learning technique that detects polarity (e.g. a *positive* or *negative* opinion) within text, whether a whole document, paragraph, sentence, or clause.



My experience  
so far has been  
fantastic!

POSITIVE



The product is  
ok I guess

NEUTRAL



Your support team  
is useless

NEGATIVE

## Sentiment Analysis - Example

### ■ Review 1 :

The Restaurant is great, Staff are really friendly and food is delicious.

Positive or Negative ?



### ■ Review 2 :

I would not recommend this restaurant to anyone, food is terrible and is really expensive.

Positive or Negative ?



# Types of Sentiment Analysis

- **Fine-grained Sentiment Analysis**

It involves determining the polarity of the opinion. It can be a simple binary positive/negative sentiment differentiation. This type can also depending on the use case (for example, as in five-star Amazon reviews).

- **Emotion detection**

It used to identify signs of specific emotional states presented in the text. Usually, there is a combination of lexicons.

- **Aspect-based sentiment analysis**

It goes deeper. Its purpose is to identify an opinion regarding a specific element of the product. For example, the brightness of the flashlight in the smartphone.

# Sentiment Analysis Algorithms

- Sentiment analysis uses various Natural Language Processing (NLP) methods and algorithms.
  1. Rule Based algorithm
  2. Automatic algorithm and etc.

# Rule Based algorithm

- Rule-based sentiment analysis is based on an algorithm with a clearly defined description of an opinion to identify. Includes identify subjectivity, polarity, or the subject of opinion.
- The rule-based approach involves basic NLP routine. It involves the following operations with the text corpus:
  - Stemming
  - Tokenization
  - Part of speech tagging
  - Parsing

## How Does It Work ??

PolarizedWords

- Define two lists of polarized words

Count

- Assign +1 for positive word
- Assign -1 for negative word

Positive  
Vs  
Negative

No of +ve words > no of -ve words  
Then Text is (positive) polarity



# Example

Camera is awesome

Screen is good

Processing is fast

Battery backup is so poor

Design is not bad

Finally I say this product is good  
and I satisfied

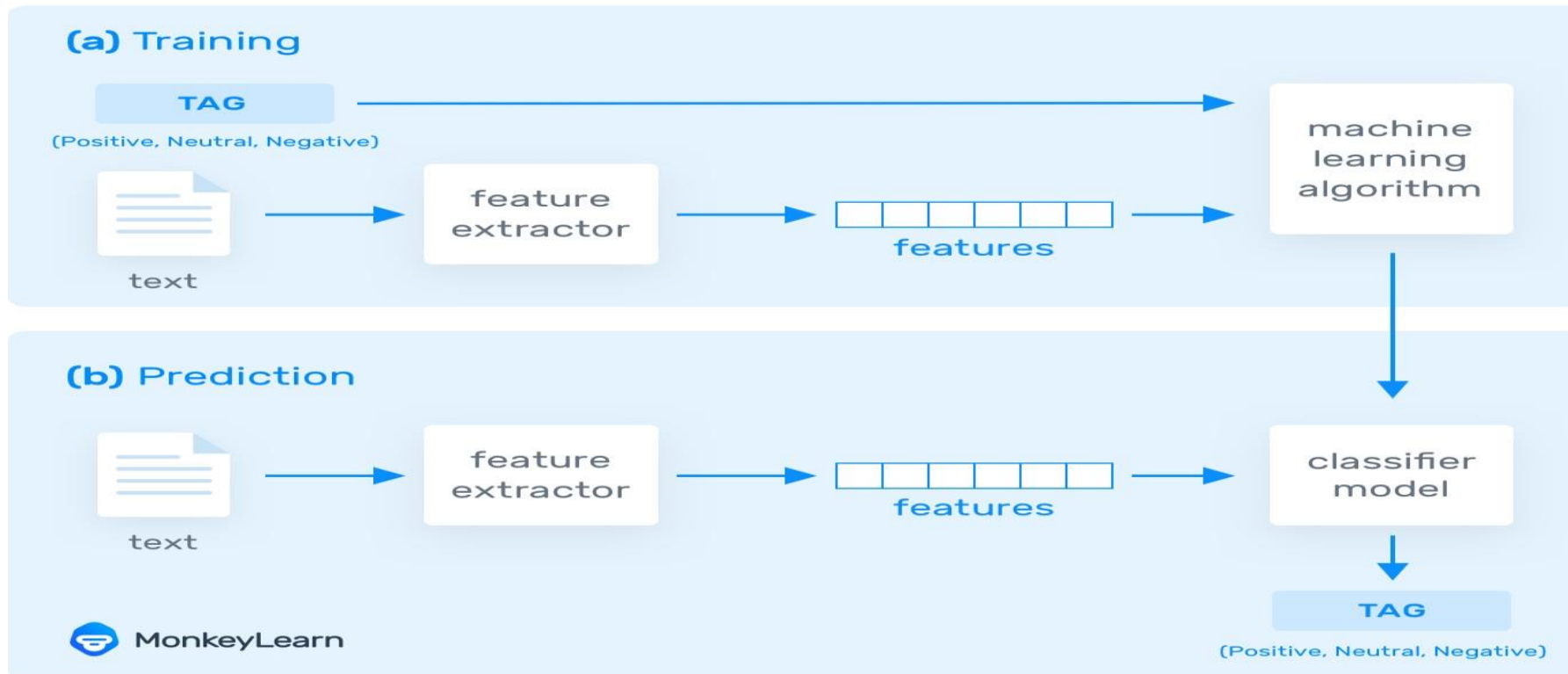
| Positive Words                       | Negative Words |
|--------------------------------------|----------------|
| Awesome<br>Fast<br>Good<br>satisfied | Poor<br>Bad    |

Now Apply Rule Based Approach to above review and  
classify whether given review is positive or negative

# Automatic Algorithm

- A sentiment analysis task is usually modeled as a classification problem, whereby a classifier is fed a text and returns a category, e.g. positive, negative, or neutral.

## How Does Sentiment Analysis Work?



# Advantages of Sentimental Analysis

- Social media monitoring
- Brand monitoring
- Voice of customer (VoC)
- Customer service
- Market research

# Applications of Sentiment Analysis

- Consumer information
  - Product reviews
- Marketing
  - Consumer attitudes
  - Trends
- Politics
  - Politicians want to know voters' views
  - Voters want to know politicians' stances and who else supports them
- Social
  - Find like-minded individuals or communities

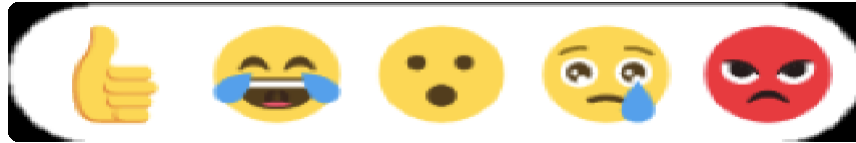


## Segmentation and recognition

- **Text segmentation** is the process of dividing written text into meaningful units, such as words, sentences, or topics.
- Word Segmentation
- Sentence Segmentation
- Topic Segmentation

# Recognition

- Emotion Recognition is the identification of emotions usually through facial expression and verbal communication such as happy, angry, sad, etc.
- **Happy, Sad, Angry, Fearful, Excited, Bored or anything**







**“Thank you”**