



ROHINI

COLLEGE OF ENGINEERING & TECHNOLOGY

Approved by AICTE and Affiliated to Anna University (An ISO Certified Institution) | Accredited with A+ Grade by NAAC
Recognized under Section 2(f) of University Grants Commission, UGC ACT 1956
(AUTONOMOUS)

INTRODUCTION TO NATURAL LANGUAGE PROCESSING (NLP)

Natural Language Processing (NLP) helps computers understand, interpret and produce human language. It studies language as data and develops a model that can analyse linguistic structure, meaning and context in both written and spoken communication.

Simple Example of NLP: **“Ravi is happy with the new phone.”**

An NLP system can:

- Detect Ravi as a person
- Identify phone as an object
- Recognize sentiment as positive
- Understand topic as product review

How Natural Language Processing Works

1. Text or Speech Input

- **Receiving text data:** The system takes written language like sentences or documents which is called text acquisition.
- **Receiving voice input:** When the input is audio, it is first converted into text using Speech Recognition.

2. Pre-processing

The text is cleaned and prepared. It can include:

- **Removing punctuation or noise:** Cleaning unwanted characters or symbols from text is done using text normalization.
- **Splitting into words:** Breaking sentences into smaller units so they can be processed easily.
- **Converting to lowercase:** Changing all words into the same case for uniform processing is known as case folding.
- **Removing common words:** Eliminating frequent words like is, the, and to focus on meaningful terms.
- **Reducing words to base form:** Converting words like running to run to reduce computational power.

3. Language Analysis

The system studies structure and meaning:

- **Grammar detection:** Identifying nouns, verbs, and other parts of speech in a sentence is done.
- **Word relationships:** Finding how words connect to each other in a sentence.
- **Context understanding:** Determining the actual meaning of a word based on surrounding text.
- **Finding names and places:** Detecting entities like person names, locations, or dates.
- **Sentiment detection:** Identifying whether text expresses positive, negative or neutral emotion.

4. Text Representation and Embedding Techniques

Since machines process numbers, this stage converts text into numerical vectors.

- **Text representation:** In this step, text is converted into numbers using statistical features or vector representations so machines can process it.
- **Traditional representations:** Earlier methods represent text using word counts and importance scores.
- **Word embeddings:** Modern methods represent words as dense vectors capturing similarity and meaning.
- **Contextual embeddings:** Advanced models generate word meanings based on the surrounding sentence.

5. Model Training

Once text is numeric, models are trained to learn patterns and perform NLP tasks.

- **Model training:** After text is converted into vectors, algorithms learn patterns from data to perform tasks like classification or translation.
- **Traditional machine learning:** Earlier NLP systems relied on statistical algorithms that learn from manually prepared features.
- **Deep learning approaches:** Modern NLP uses neural networks that automatically learn language structure from large data.
- **Pre-trained models:** Large language models trained on massive datasets can be reused and fine-tuned for tasks.

6. Output Generation

The system produces results such as:

- Text reply
- Voice response

- Translation
- Summary
- Prediction

Common NLP Tasks

- **Text classification:** Assigning predefined labels to text like spam or topic categories.
- **Sentiment analysis:** Detecting whether text expresses positive, negative or neutral emotion.
- **Machine translation:** Automatically converting text from one language to another.
- **Named Entity Recognition:** Identifying names of people, places, dates, etc in text.
- **Text summarization:** Generating a shorter version of a document while keeping key meanings.
- **Question answering systems:** Systems that read text and return exact answers to queries.

Real-Life Applications

- Voice assistants like Alexa, Google Assistant, etc
- Chatbots in customer support
- Email spam filtering
- Auto-correct and predictive typing
- Language translation tools
- Social media sentiment tracking
- Document search and recommendation systems