COMPUTATIONAL LINGUISTICS(NLP)

Nisha Malar V[1], Swetha S[2], Yamini C L[3]

[1]*UG Student, Department of Computer Science and Engineering, Sri Ramakrishna Institute of Technology, Tamil Nadu, India*

[2]*UG Student, Department of Computer Science and Engineering, Sri Ramakrishna Institute of Technology, Tamil Nadu, India*

[3]*UG Student, Department of Computer Science and Engineering, Sri Ramakrishna Institute of Technology, Tamil Nadu, India*

ABSTRACT

Computational Linguistics or Natural language processing(NLP), is a type of artificial intelligence that deals with analysing, understanding, and generating natural human languages so that computers can process written and spoken human language without using computer-driven language. Natural language processing, sometimes also called "computational linguistics," uses both semantics and syntax to help computers understand how humans talk or write and how to derive meaning from what they say. This field combines the power of artificial intelligence and computer programming into an understanding so powerful that programs can even translate one language into another reasonable accurately. This field also includes voice recognition, the ability of a computer to understand what you say well enough to respond appropriately.

Keyword: - *Computational linguistics, Natural language processing, Natural language understanding, machine learning syntactic, semantic, ambiguous etc.,*

1. INTRODUCTION

Natural Language processing (NLP) could be a field of AI (AI) that creates human language intelligible to machines. NLP combines the ability of linguistics and computing to check the foundations and structure of language, and make intelligent systems (run on machine learning and NLP algorithms) capable of understanding, studying, and extracting meaning from text and speech.

Natural Language Understanding (NLU) is employed to grasp the structure and meaning of human language by analysing totally different aspects like syntax, semantics, pragmatics, and morphology. Then, computing transforms this linguistic information into rule-based, machine learning algorithms which will solve specific issues and perform desired tasks.

Take Gmail, for instance. Emails are mechanically categorised as Promotions, Social, Primary, or Spam, because of AN NLP task referred to as keyword extraction. By "reading" words in subject lines and associating them with pre-set tags, machines mechanically learn that class to assign emails.

2. NATURAL LANGUAGE PROCESSING/COMPUTATIONAL LINGUISTICS

Computational linguistics is an area of research and application that explores how computers can be used to understand and manipulates natural language text or speech to do useful things. The applications of Natural Language Processing include fields of study, such as machine translation, natural language text processing and summarization, user interfaces, multilingual and cross language information retrieval (CLIR), speech recognition, artificial intelligence (AI) and expert systems.

Fig 1: Natural Language Processing application fields

Often, NLP is running in the background of the tools and applications we use every day, helping businesses improve our experiences. Below, are some of the most common and most powerful uses of natural language processing in everyday life:

Everyday Examples of NLP

1. Email filters
2. Virtual assistants, voice assistants, or smart speakers
3. Online search engines
4. Predictive text and autocorrect
5. Monitor brand sentiment of social media
6. Sorting customer feedback
7. Automatic summarization
8. Chatbots
9. Automating processes in customer support
10. Machine translation
11. Natural language generation

2.1 HOW DOES NLP WORKS?

NLP entails applying algorithms to spot and extract the natural language rules specified the unstructured language data is regenerate into a type that computers will understand.

When the text has been provided, the pc can utilize algorithms to extract meaning associated with each sentence and collect the essential information from them. Sometimes, the pc might fail to grasp the meaning of a sentence well, resulting in obscure results.

For example, a humorous incident occurred in the 1950s throughout the translation of some words between the English and also the Russian languages.

Here is that the biblical sentence that need translation:

 "The spirit is willing, but the flesh is weak."

Here is that the result once the sentence was translated to Russian and back to English:

"The liquor is sweet, but the meat is rotten."

2.2 NATURAL LANGUAGE Vs COMPUTER LANGUAGE

Parameter	Natural language	Computer language
Ambiguous	They are ambiguous in nature.	They are designed to unambiguous.
Redundancy	Employs lots of redundancy.	Less redundant.
Literalness	Made of idiom and metaphor.	Mean exactly what they want to say.

Table 1: Comparison between natural language and computer language

3.   NATURAL LANGUAGE PROCESSING TASKS

There are no hard lines between these task types; but, several square measure fairly well-defined at this time. A given Macro NLP task could include a range of sub-tasks. Below are the main categories of NLP tasks.

- TEXT CLASSIFICATION TASKS
  Representation: bag of words (does not preserve word order)
  Goal: predict tags, categories, sentiment
  Application: filtering spam emails, classifying documents based on dominant content

- WORD SEQUENCE TASKS
  Representation: sequences (preserves word order)
  Goal: language modelling – predict next/previous words, text generation
  Application: translation, chatbots, sequence tagging (predict POS tags for every word in sequence), named entity recognition

- A TEXT MEANING TASKS
  Representation: word vectors, the mapping of words to vectors (n-dimensional numeric vectors) aka embeddings
  Goal: how can we represent meaning?
  Application: finding similar words (similar vectors), sentence embeddings (as hostile word embeddings), topic modelling, search, question respondent

- SEQUENCE TO SEQUENCE TASKS

Many tasks in NLP can be framed as such examples are machine translation, summarisation, simplification, Q&A systems.
Such systems are characterised by encoders and decoders, which work in complement to find a hidden representation of text, and to use that hidden representation.

- DIALOG SYSTEMS
  Two main categories of dialog systems, categorized by their scope of use

Goal oriented dialog system focus on being useful in a particular, restricted domain; more precision, less generalizable.
Conversational dialog systems are concerned with being helpful or entertaining in a much more general context; less precision, more generalization.

4. COMPONENTS OF NLP
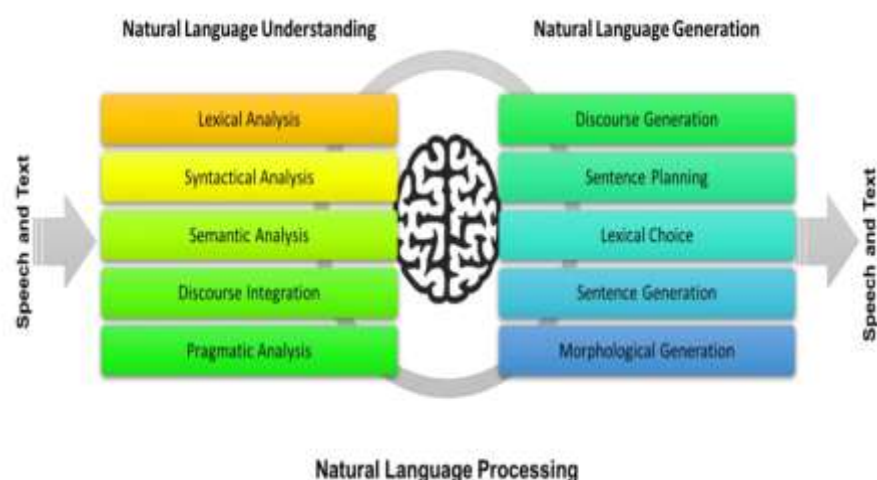
Natural Language Processing
Fig 2: Representation of NPL via components

LEXICAL ANALYSIS: It involves identifying and analysing the structure of words. Lexicon of a language means that the gathering of words and phrases in a language. Lexical analysis is dividing the entire chunk of text into paragraphs, sentences, and words.
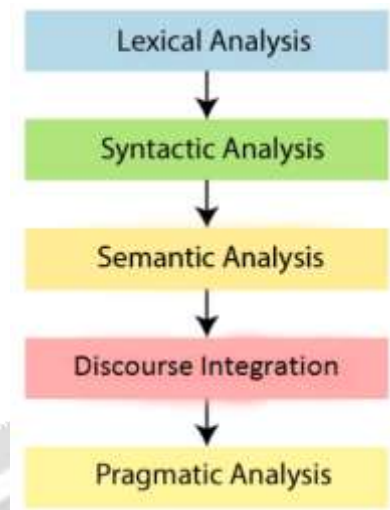

Fig 3: Steps in Natural Language Processing

SYNTACTIC ANALYSIS: It involves analysis of words within the sentence
for grammar and transcription words in a manner that shows the link among the words. The sentence like "The college goes to boy" is rejected by English syntactical analyser. Here are some syntax techniques that can be used:

- Morphological segmentation: It involves dividing words into individual units known as morphemes.
- Word segmentation: It involves dividing an oversized piece of continuous text into distinct units.
- Part-of-speech tagging: It involves identifying the part of speech for each word.
- Parsing: It involves endeavour grammatical analysis for the provided sentence.
- Sentence breaking: It involves inserting sentence boundaries on an outsized piece of text.

SEMANTIC ANALYSIS: It draws the precise meaning or the dictionary meaning from the text. The text is checked for significance. It's done by mapping syntactic structures and objects in the task domain. The semantic analyser disregards sentence like "hot ice-cream".
Here are some techniques in semantic analysis:

- Named Entity Recognition (NER): It involves deciding the components of a text which will be known and categorised into predetermined groups. Samples of such groups include names of individuals and names of places.
- Word Sense Disambiguation: It involves giving meaning to a word based on the context.
- Natural Language Generation: It involves using databases to derive linguistics intentions and convert them into human language.

DISCOURSE INTEGRATION: The meaning of any sentence depends upon the meaning of the sentence simply before it. Additionally, it also brings about the meaning of immediately succeeding sentence.
PRAGMATIC ANALYSIS: throughout this, what was said is re-interpreted on what it really meant. It involves etymologizing those aspects of language that need real world information.

5. FREQUENTLY USED ALGORITHMS IN NLP

- TOKENIZATION

Is the process of segmenting running text into sentences and words? In essence, it's the task of cutting a text into pieces called tokens, and at the same time throwing away certain characters, such as punctuation.

- STOP WORDS REMOVAL

Includes getting rid of common language articles, pronouns and prepositions such as "and", "the" or "to" in English. In this process some very common words that appear to provide little or no value to the NLP objective are filtered and excluded from the text to be processed, hence removing widespread and frequent terms that are informative about the corresponding text.

Stop words can be safely ignored by carrying out a lookup in a pre-defined list of keywords, freeing up database space and improving processing time.

- STEMMING

Refers to the process of slicing the end or the beginning of words with the intention of removing affixes (lexical additions to the root of the word).

- LEMMATIZATION

The objective of reducing of word to its base form and grouping together different forms of the same words. For example, verbs in past tense are changed into present (e.g., "went" is changed to "go") and synonyms are unified (e.g., "best" is changed to "good"), hence standardizing words with similar meaning to their root.

- TOPIC MODELING

Is as a way for uncovering hidden structures in sets of texts or documents. In essence it's clusters texts to get latent topics based on their contents, processing individual words and assigning them values based on their distribution. This method relies on the assumptions every document consists of a combination of topics which each topic consists of a collection of words, which implies that if we are able to spot these hidden topics, we can unlock the meaning of our texts.

6. PROS AND CONS OF NPL

6.1 PROS

- Large-scale analysis: natural language processing helps machines automatically perceive and analyse immense amounts of unstructured text data, like social media comments, client support tickets, on-line reviews, news reports, and more.
- Automated processes in real-time: natural language processing tools will facilitate machines learn to sort and route information with very little to no human interaction – quickly, with efficiency, accurately, and round the clock.
- Tailored to your industry: natural language processing algorithms is tailored to your desires and criteria, like advanced, industry-specific language – even satire and ill-used words.

6.2 CONS

- Complex query Language- the system might not be able to give the right answer if the question is poorly worded or ambiguous.
- The system is made for one and specific task only; it's unable to adapt to new domains and issues as a result of restricted functions.
- NLP system doesn't have an interface that lacks features that permit users to further interact with the system.

7. CONCLUSION

NLP provides a large set of techniques and tools which might be applied in all the areas of life. By learning them and exploitation them in our everyday interactions, our life quality would extremely improve, we have a tendency to improve the lives of those who surround us. NLP techniques facilitate us improving our communications, our goal reaching and therefore the outcomes we tend to receive from each interaction. They also enable as overcome personal

obstacles and psychological issues. NLP facilitate US using tools and techniques we have already got in us without being conscious of it.

By learning NLP techniques properly, we might be able to come through our goals very satisfactory method, and overcome any obstacle that stands before of us. As the NLP suppose says "everybody has the resources they might ever need", we already have what it takes, and we solely ought to learn about it.

NLP and machine learning applications play a crucial role in supporting machine-human communications. With a lot of analysis during this sphere, there are a lot of developments to form machines smarter at learning and understanding the human language.

8. REFRENCES

[1]. From Languages to Information, Stanford.

[2]. Natural Language Processing, National Research University

   Higher School of Economics (Coursera).

[3]. Neural Network Methods for Natural Language Processing, Yoav Goldberg.