

# Natural Language Processing: Comprehensive Review

Vinothina V<sup>1\*</sup>

Dept. of Computer Science, Kristu Jayanti College Autonomous, Bengaluru, India

\*Corresponding Author: [vinothina.v@kristujayanti.com](mailto:vinothina.v@kristujayanti.com), Tel.: 8861511222

DOI: <https://doi.org/10.26438/ijcse/v7si9.8486> | Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

**Abstract**— The emerging and most popular processing technique of Natural Language Processing (NLP) deals with computational algorithms to analyze and process human language. It also represents human language. The application of NLP enabled system ranges from search engine to voice assistant, machine translation and dialogue generation. It is concerned with interaction between computers and natural language such as human language. High-revenue non NLP-domain such as finance, government surveillance or marketing is the areas benefited from NLP. The main objective of this paper to review state-of-the-art of NLP, its benefits and various processes in NLP enable the systems to deal with real-time data.

**Keywords**— Machine Learning, Internet of Things, Artificial Intelligence, Google, Voice Assistant

## I. INTRODUCTION

NLP is the combination of different disciplines of computer science, Artificial Intelligence and Linguistics. The major task of NLP includes automatic summarization, discourse analysis and information retrieval. NLP started with rule-based methodology which used grammar rules and dictionary entries to process human language. But developing NLP enabled system was difficult due to huge amount of entries compilation. Hence corpus-based and statistical systems are got into the picture to develop NLP enabled system. Corpus based is study of language based on real time usage. The different methods deployed in NLP such as application methods, acquisition methods, and evaluation methods are in statistical sense as it uses probability and other concepts from statistics. NLP enabled systems build using statistics also used linguistic rules. Hence articulated framework is used to build NLP System [2].

Due to rapid advancement in computer power and data storage, nowadays we can utilize huge amount of actual linguistic data along with language analyser to build effective NLP system. Using these linguistic databases, information can retrieved in different views which supports NLP [3]. NLP is strongly interconnected with research, science, linguistics, e-learning, evaluation systems and yields positive results [4]. Mainly NLP is used to create own massive datasets from existing one, transforming data into something useful for processing, incorporate human intelligence and build generalizable model [5].

SHRDHU, a Natural language System developed in 1960 uses restricted vocabularies. ELIZA is simulation of

psychotherapist provided human like interaction. Other systems, Margie, sam, pam, qualm, and alespin are conceptual ontologies which represented real –world information into computer understandable format developed in 1970's. Till 1980's, NLP used complex set of hand-written rules, then started using machine language. The objective of this paper is to review state-of-the-art of NLP, its real-world applications and algorithms used to develop NLP enabled system.

In the following section, related works are discussed, in section III, state-of-the-art, benefits, challenges and techniques of NLP are discussed and last section dealt with conclusion.

## II. RELATED WORK

The review of NLP using deep learning models, issues in linguistics processing and applications of computational linguistics are given in [6]. Daniel et.al also addressed future research areas. Webber [7], discussed basic concepts and terms of NLP both in generation and Interpretation. Theoretical, applied research results and NLP relations to knowledge management system were also discussed. In paper [9], review on various approaches of NLP, and applications are discussed. Systematic survey on NLP applications in Health care systems are addressed in [11].

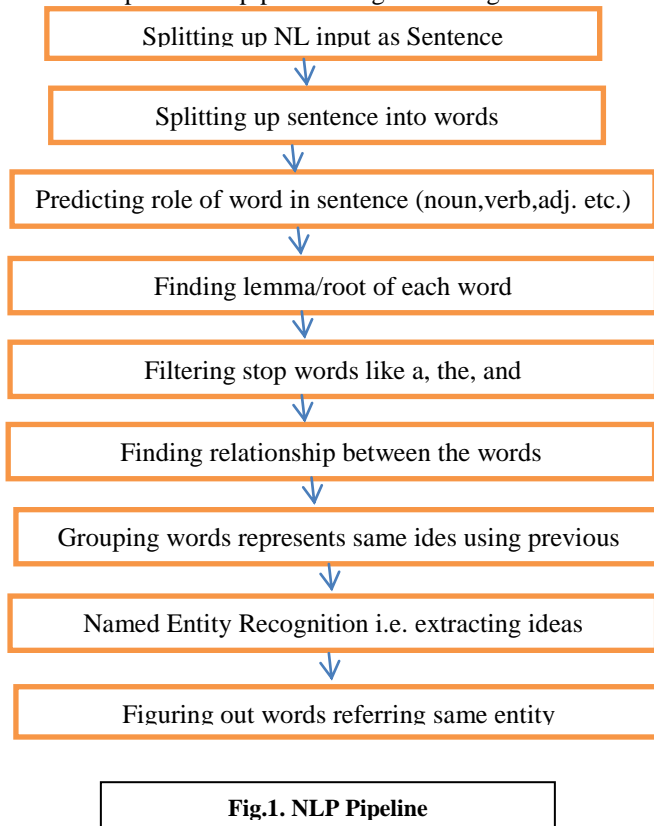
## III. PROPOSED WORK –PROCESSES IN NLP

NLP involves making computers to perform useful tasks with the natural languages humans use. Speech and text are input and output of an NLP system. Natural language

understanding (NLU) and natural language generation (NLG) are two components of an NLP system. The functions of NLU are: 1. transforming the given input language into useful representation 2. Analysing different aspects of language. From the internal representation, producing phrases and sentences in the form of natural language are the significant task of NLG. This significant task involves the following task as follows:

- Text planning – retrieving the relevant content from knowledge base.
- Sentence planning – choosing required words, forming meaningful phrases, setting tone of the sentence.
- Text Realization –mapping sentence plan into sentence structure

Various steps in NLP pipeline are given in Fig.1.



Each step in the NLP pipeline is challenging task. Finding verb, noun, adjectives, etc. is done with predefined linguistic rules. Lemmatization i.e. finding the root word of difference forms is also determined using rule based methods. Normalizing stop words before any processing depends upon the application as stop words differ from application to application. Parse tree shows relationship between the words. Implementation of grouping words is depends upon end goal. NER (Named Entity Recognition) tag the nouns with real-word entity which they represent. For example, place name,

company name, people's name, geographic locations etc., are the real world entities.

Now, the natural language text is transformed in to structured way. The most challenging step in NLP pipeline is figuring out the words which refers same entity across the sentences i.e. pronouns like he, she, it etc. This process is called coreference resolution. With the help of parse tree and NER, more useful information can be extracted. But the implementation of coreference resolution is difficult. Advancements in deep learning algorithms provide accurate result and applied in coreference resolution to overcome the problem [10].

### 3.1 Challenges in NLP Pipeline

NLP is very ambiguous. It must be disambiguated. There are different ambiguities as listed in Table 1.

**Table.1** Types of Ambiguity in NLP

Ambiguity	Reason
Speech Recognition (Lexical ambiguity)	Pronouncing the different words same
Sentence can be parsed in different way. (syntax level ambiguity)	Preposition attachment
Word Disambiguation	Sense Words have different possible meaning.
Referential	Referring to something

Phonological Analysis, Morphological analysis, Lexical Analysis, syntactic analysis, semantic analysis and pragmatic analysis are the different analysis processes done as NLP. Phonological analysis is concerned with organization of speech sounds. Morphological analysis is defined as the scientific study which involves identifying the structure of words in a language and finding the relationship between the words.

Lexical analysis tokenizes the input text in to words. The role of syntax analysis phase is construction of sentences and finding the relationship amongst the words in sentence. Moreover, syntax analysis easier the process of Pragmatic analysis and semantic analysis process as they extract meaning from input. Semantic analysis focuses on the interactions among word-level meanings in the sentence. Pragmatic analysis deals with how sentences in different contexts are combined to form discourse such as paragraphs, documents and dialogues [11]. Syntax, Semantic, discourse analysis and pragmatic are analysis factor of NL. Generation, on the other hand, is commonly factored into three components such as Content Determination , Text Planning and Realization as a NL Text [7].

### 3.2 Applications of NLP

NLP is mainly used in healthcare section for extracting information, retrieving information, question and answering, communication, document categorization and machine translation. In education, NLP is used for providing information and learning through natural acquisition. It is an effective approach for teachers, students, authors and educators for providing assistance for writing, analysis, and assessment procedures. Various educational contexts such as research, science, and linguistics, e-learning are widely integrated and resulting positive outcomes in other educational settings such as schools, higher education system, and universities.

In business, NLP addresses customer pain through sentiment analysis, gather market intelligence from unstructured data and reduce customer frustration through hybrid bots. In business, NLP is integrated with various phases of finance ecosystem from approving loans, credit scores, to managing assets, to assessing risks with machine learning approaches[13]. In Banking, NLP might allow a company to garner insights that can be used to assess a creditor's risk or gauge brand-related sentiment across the web [14].

In recent days, Home automation one of significant benefits of Internet of Things. Home automation includes the feature of controlling the fans, lights and other electrical appliances in a house. There is a web application using which the fans, lights and other electrical appliances can be controlled over the Internet. The web application uses chatbot algorithm using that user send text to control the electrical appliances. The messages sent using the chatbot is processed using Natural Language processing techniques [15].

In e-governance, NLP plays a major role in areas like employment, News, Who's who of government, Legal Sub-portal, Grievance Handling, survey reports, government schemes, directory services for all government offices, links to all other government sites, parliamentary discussions and trade related queries[16].

Financial market can be categorized in to equity, credit, risk management and sustainability. In each category NLP is applied to analyze past earning calls and annual report to calculate the future growth of an organization. Using NLP Security factor exposures can be predicted by analyzing companies' latest reports. Environmental, Social and Governance scores of any companies can be analyzed using NLP by having right data such as press releases and reports.

#### IV. CONCLUSION

It is very difficult to design a natural language framework as natural human language is complex due to difficulty in capturing the linguistic meaning due to speech impediments, heavy accents and quiet voices. In all the situations hundred

percent accuracy may not be possible in NLP. Machine learning and deep learning approaches are applied to build computational algorithms to process and generate and NL. Optimal result is possible with reusable part of open source programming. But still representing entire knowledge, the different linguistic rules and cultures as knowledge base is challenging task. In this paper, processes involved, applications and challenges of NLP may help research practitioners to get basic knowledge of NLP and encourage them to proceed further in solving the challenges of NLP.

#### REFERENCES

- [1] E.U Reshma, P C Remya, A review of different approaches in natural language interfaces to databases, International Conference on Intelligent Sustainable Systems (ICISS), <https://doi.org/10.1109/ISS1.2017.8389287>,IEEE,2017.
- [2] Joakim Nivre, On Statistical Methods in Natural Language Processing.
- [3] Hitoshi Isahar, Resource-based Natural Language Processing, International Conference on Natural Language Processing and Knowledge Engineering, <https://doi.org/10.1109/NLPKE.2007.4368002>, IEEE,2007.
- [4] Ponnusamy, Ramalingan, A Systematic Survey of Natural Language Processing (NLP) Approaches in Different Systems, vol 4,2018.
- [5] Machine Learning in the real world: <https://techcrunch.com/2015/11/26/machine-intelligence-in-the-real-world/>
- [6] Daniel W, Otter, Julian R..Medina, Jugal K.Kalita, "A Survey of the usage of Deep Learning in Natural Language Processing", Computation and Language, Cornell University.
- [7] Webber B.L. (1986) Natural Language Processing: A Survey. In: Brodie M.L., Mylopoulos J. (eds) On Knowledge Base Management Systems. Topics in Information Systems. Springer, New York, NY.
- [8] Graeme Ritchie, "Survey of Natural Language Processing", ACM Newsletter Issue 80, pp, 66, 1982.
- [9] M.VasunthraDevi, Dr.R.Ponnusamy, "A Systematic survey on Natural Language Processing Approaches in Different systems", International Journal of Computer Science and Engineering Vol 4, 7,pp-192-198.
- [10] Natural Language Processing is Fun: <https://medium.com/@ageitgey/natural-language-processing-is-fun-9a0bff37854e>
- [11] OLaronke G.Iroju, Janet O.Olaleke, "A Systematic Review of Natural Language Processing in Healthcare,IJ Information Technology and Computer Science, 08,pp-44-50,2015.
- [12] Dr. Khaled M. Alhawiti," Natural Language Processing and its Use in Education", International Journal of Advanced Computer Science and Applications, Vol.5, No.12,2014.
- [13] Machine learning in Business: <https://emerj.com/ai-sector-overviews/machine-learning-in-finance/>
- [14] <https://emerj.com/ai-sector-overviews/natural-language-processing-banking-current-applications/>
- [15] Cyril Joe Baby, Faizan Ayyub Khan, J. N. Swathi, "Home Automation using IoT and Chatbot using Natural Language Processing", International Conference on Innocations in Power and Advanced computing Technologies ISBN, 978-1-5090-5682-8.
- [16] M A Moni, "Application of Natural Language Processing Techniques in e-Governance", International conference on Language and Development, 2009.