

Natural Language Processing in AI: Language Interpreter

Sapna Sharma*

Assistant Professor, Department of Bachelor of Computer Application, Modern College, Business Administrative Campus, Pune, Maharashtra, India

Abstract

With the emergence of Artificial Intelligence in this technocrat world has led to the transformation in the world of technology. With the emergence of Artificial Intelligence it has led to different more technologies such as voice assistants which work for humans in order to make it easy in the workplace but in order to make it more of use and understandable. It has been proved to best suited for study of mathematical and computational modeling of various aspects of language. It widely includes a spoken language system which mainly integrates with speech and natural language. Natural Language Processing mainly focuses on the areas like Research and Language Text type of applications. It is being mentioned that Natural language is viewed in two terminologies as Rule based and Statistical based. Natural Language Processing has the base on heuristic search stemming with a tremendous root cause. Uses majority of English language rules and processing regulation for speechtual.

Keywords: Artificial Intelligence, Deep Learning, machine learning, Techniques of Natural Language Processing, Natural Language Processing, Lexical Input, Natural Language Generation, Natural Language Understanding.

***Author for Correspondence** E-mail: Sapnasharma8652@gmail.com

INTRODUCTION

Natural language processing is a form of artificial intelligence that helps machine to “read” text by simulating the human ability to understand language. NPL technology incorporates in a vivid methods which include linguistics, semantics, statics and machine learning to do extraction and it also includes entities relationship and understanding of context, which enables the understanding of what is being said in a well defined manner [1]. It not only concentrate on the single or combination of words in fact it helps the computer to understand the sentence which is being spoken by the person. It has many methods in order decode the ambiguities which are there in the language which include the following such as automatic summarization, part of speech, tagging, entity extraction and relation extraction as well as natural language understanding and recognition. NLP involves machines or robots in order to understand and process the language that human speaks and infer the knowledge from the speech which has been inputed. It also involves various type of active participation from machine in form of dialog [2].

Connection of NLP with AI:

- Both are very much interrelated with each other in different aspects and requirements. In other words we can tell that Natural language is a branch of Artificial Intelligence whose deals with interaction between
- Computers and human beings using natural language.
- The most important connectivity part of Natural language processing is to read the data or text and then decipher it, understand and make sense of human language in an absolute which make it more valuable and useful [3].
- From the below given figure we can analyze how natural language processing is working and how it is taking the data as an input from the speech recognition [4].
- Natural language processing works along with its two main components which get activated when the data is being taken by the natural language processing stage and is being processed by them (i.e. components) [5].

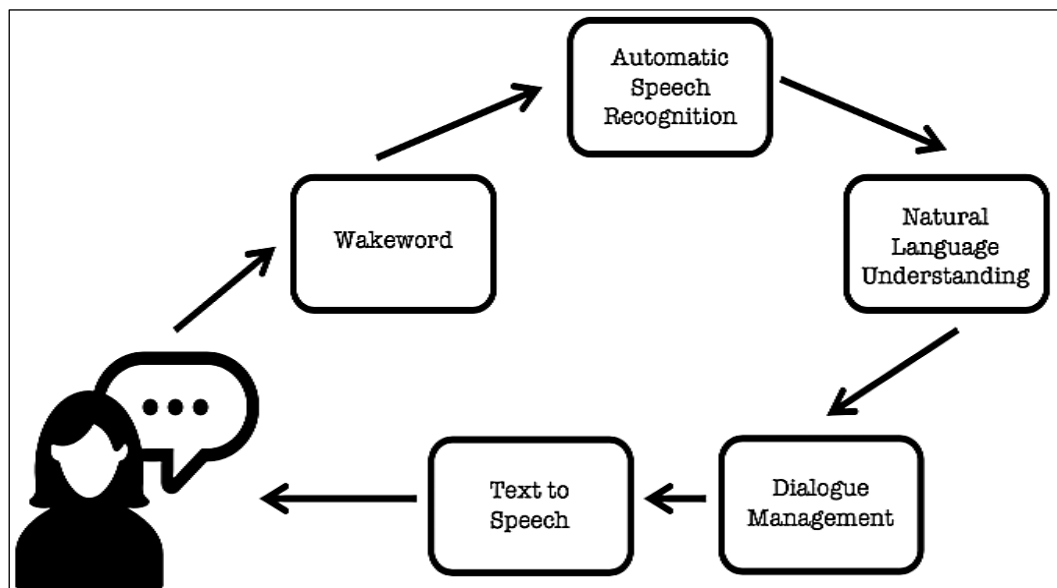


Fig. 1: Placement of Natural Language Processing AI System.

WORKING OF NLP

For working of Natural Language Processing following levels are followed and are as follows (Figure 2):

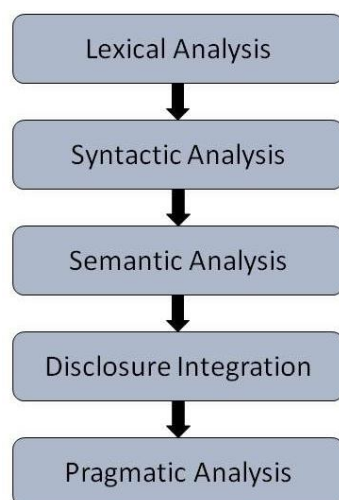


Fig. 2: Levels of NLP Working.

Explanation

Lexical Analysis

Identifies and analyze the structure of words (Collections of words and phrases in a language). All the chunks of text are being divided into small paragraphs, sentences and words.

Syntactic Analysis

Shows the relationship between words and grammar not only in this task but also rejects the

sentences which are not grammatically correct with the help of “Syntactical Analyzers”.

Semantic Analysis

Checkout for meaningfulness and also map’s the syntactic structure and objects in the task domain.

Discourse Integration

It works for both predecessor sentence and immediate successor sentence.

Pragmatic Analysis

Involves re-interpretation of current data and for what is meant for, which is being used for derivation of different aspects of language which is an absolute use for a real world knowledge system [6].

Components of NLP

Natural Language processing follows two different components and they are as follows (Figure 3):

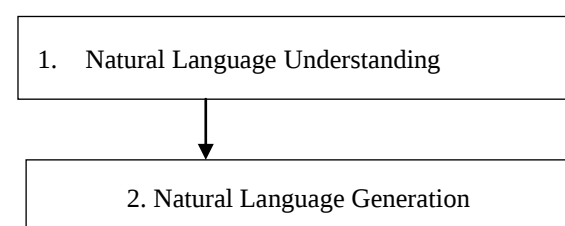


Fig. 3: Component level of Natural Language Processing.

Natural Language Understanding

Firstly it is a process in which the input out of natural language into meaningful and useful representation and secondly it analyses the different aspects of the language [7].

Natural language is very much rich in forms and various structures but at the same time it is very ambiguous too.

Natural Language Generation

It can be defined as generating meaningful phrases and sentence in way of natural language and involves the following points around it [8]:

Text Planning

It is known for its Retrieval of authentic data.

Sentence Planning

In this Choosing of words and then forming meaningful phrases.

Text Recognition

Map's sentence plan into sentence structure.

Used Techniques of NPL

Natural Language Processing follow up two different type of techniques:

- a. Syntactic Analysis:
- b. Semantic Analysis:

Syntactic Analysis

It is a combination of words in such a manner that it forms a well defined sentence. NPL pairs up with grammatical rules. Applies some grammatical rules on specified groups and out of which derive some meaning out of it [9].

Techniques

1. Lemmatization
2. Morphological Segmentation
3. Word Segmentation
4. Part of Speech tagging
5. Parsing
6. Sentences Breaking
7. Stemming

Lemmatization

It combines various forms of data into a single sentence or form in order to do easy analysis [10].

Morphological Segmentation

It divides different words into an individual unit which are also called as "Morphemes".

Words Segmentation

It is the process of breaking the large and continuous data into distinct data forms.

Part of Speech Tagging

Process of identifying part of speech for each and every word.

Parsing

Do the analysis on the grammatical part of the provided sentences.

Sentence Breaking

Placement of sentence boundaries on a very large chunk of data or text.

Stemming

It is the process of cutting the vivid changing forms of data from their root generation.

Sematic Analysis

It carries the appropriate meaning of the text in use and it is a bit difficult too to understand this aspect. Various algorithms are being used in order to do interpretation sentence structuration.

Techniques

1. Named Entity Recognition
2. Word sense disambiguation
3. Natural language generation

Named Entity Recognition

Determining of text which can be categorized and identified onto pre-set groups. It is a small piece of text or moreover a series of text which tends to the same thing. Subfield of Artificial Intelligence which is more concerned with computer processing and analysing natural language.

Word Sense Disambiguation

It is all about meaning to a word or text. There is one of the problems that is being faced by Natural Language Processing at most level Lexical ambiguity which actually violates the accuracy of words or the texts. This problem is being resolve by Word sense

disambiguation, for this WDS uses two different types of input [11]:

- a. **A Dictionary:** It is used to evaluate sense that is disambiguated.
- b. **Test Corpus:** It works on target base for more precise correct sense and is divided into two inputs: @ Lexical sample and @ All words.

Natural Language Generation

It is a kind of software process which actually transforms the structural data into natural language. It is being used by many of the organizations for different purposes. For example in the Psychological department they use it for their specific task and it is called language production. It is widely used for translators of artificial intelligence's computer languages; examples are "**Decompilers**" and "**transpilers**". It is specialized in putting in concepts in words. NLG doesn't always work in the same form(one way). It totally works on ambiguity and error containable data. Has following stages too [12]:

1. Content Determination
2. Document Structuring
3. Aggregation
4. Lexical Choice
5. Referring expression generation
6. Realization

Content Determination

Decision is being taken about what information need to mention in the following text.

Document Structuring

What level of information needed to be conveyed with respect to the organization.

Aggregation

In-order to reduce the individual `execution of data and long procedure to process it aggregation method is being performed. It means combining or merging of similar type of data with same functionality into one particular common group which not only reduce unnecessary processing but also enhance the readability and natural form of data.

Lexical Choice

This stage will ensure about the getting proper concept with respect for data or text.

Referring Expression Generation

At this stage the reference is being provided with respects to the original data which identify in and out of the original data and its entities.

Realization

All these stages works on proper English language presentation. At this stage actuality of text is being performed and also ensure that it follows up proper rules and protocols of the English Language processing [13].

EVALUATION OF NATURAL LANGUAGE GENERATION

A lot of things and research are needed to be performed by NLG in order to achieve precision in work and have three different techniques for it (Figure 4):

- a. Task Based
- b. Human rating
- c. Metrics

Task Based

It provides the evaluated and generated text to the human being and also provide knowledge about how to completely perform task in out of best manner.

Human Rating

At this specific stage the data which is being provided to human being is asked for provide the rating about the data, in order to under how perfect it is as per the requirement.

Metrics

At this level the data or the text is being compared with one another for the same input.

VARIOUS FORM OF USAGE OF NLG

Core Tasks Performed

The below diagram Figure 5 includes the various core tasks that are essential.

Most General Level Application

The following diagram Figure 6 shows the applications which strictly follow the NLP.

Industrial Specific Level Work

The following diagram Figure 7 shows different type industries use this technology.

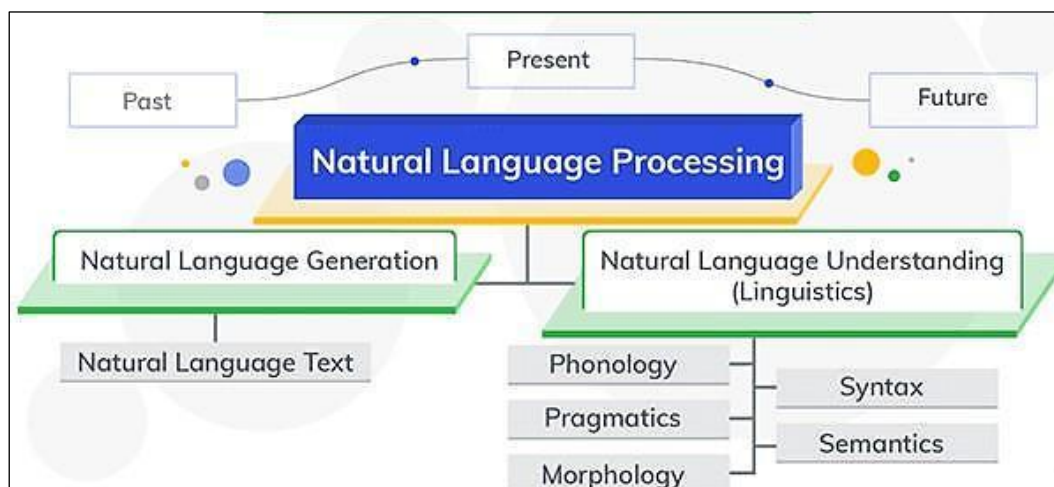


Fig. 4: Diagrammatic Information of Classified NLP.



Fig. 5: Diagram wrt Core Task Performed.



Fig. 6: Diagram wrt Various forms of Applications.



Fig. 7: Diagram wrt Various Industries use NLP.

USAGE OF NATURAL LANGUAGE PROCESSING

Various sectors are using Natural Language Processing as per the essential requirement and accomplishment of the application, here are some of the following:

Machine Translation

Various types of algorithms manage out automation translation without any human interaction so far needed. But it is one of the most challenging tasks with respect to various forms of languages. A lot things are needed into consideration.

Speech Recognition

It is one of the hottest topics which is very much trendy and essential in usage for various sectors. It has made the work of human beings more easy and flexible too.

Sentiment Analysis

It is a very different form of *method in Natural Language Processing*. It totally works on the pointed out opinion which is being provided by human beings. For example it helps to know the customer satisfaction about the provided goods and services.

Question Answering

At this level a system is built where the system will automatically provide the question and answer which is being posed by the human being with respect to natural language.

Automatic Summarization

Information overload is one of the problems or in another word can say that drawback which is being faced by the system and person. So to overcome this problem automatic summarization is an answer which prepares shorts and accurate forms of long data. This actually helps in different ways to the concerned people. Many systems have been into the picture for this work to be accomplished.

Market Intelligence

It plays a very important role in today's market which actually tells about the different trendy customers and their way of purchase. A lot of data is being collected from different sources in order to make a clear broad image of customers and their activities they perform.

Text Classification

It is a way of categorizing the data or text. It makes the work of understanding what type of data it is actually and what is the use of it. Analyzation and assignment of category is done over here.

CONCLUSION

As per the demand of the real world needs and requirements, it has become very much essential to make such system which can understand what exactly human try to do. These types of system helps human in order to accomplish their work in much more speedy way and on time. They have lead to a tremendous change in the technology world and we have come across many of such technique. It is not so easy to generate such a machine for the usage, it's not just a simple machine who has to perform certain tasks, in fact it is a smart machine which has changed this world at a very different level and has created a benchmark for the other kinds of machine. They have everything that a human has, infact more appropriate than a human too. But if we give a second thought about

the such creation, we can conclude the interdependency on such a machine for many of the tasks in order to achieve badges (rewards) fastest, to become number one and many more things. It has led to less importance to the human being and its intelligence with respect to everything. It has led to the depreciation of human beings regarding their work and job. Though it contains the intelligence power of human and thinking ability, it has led to too much dependency among human beings in-order for the process. Being such a smart process it is leading to a lot of ambiguity and error in data. It has a very big problem with isolation of languages, a lot of problems where a single word is having different meaning and many aspects where it is lacking behind. It needs to analyze more with aspects of more and accurate level of output out of it. Some time due to inefficiency or some wrong procedure the outcome is unpredictable and would lead to wrong information generation. With all these points we come to a point that these concepts though have eased our but still need a lot of accuracy and completeness is required with respect to current situation.

REFERENCES

1. David M. W. Powers and Christopher C. R. Turk (1989). *Machine Learning of Natural Language*, Springer-Verlag London, edition 1st.
2. Dan Jurafsky (1999). *Speech and Language Processing: An Introduction to Natural Language Processing*, Prentice Hall, Englewood Cliffs, New Jersey.
3. Tjong Kim Sang, E. F., & De Meulder, F. (2003, May). Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. [Online] Available from: Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition [Accessed June 2020].
4. Ferrucci D, Lally A, Verspoor K (eds) (2009) Unstructured information management architecture (UIMA) Version 1.0. OASIS Standard, 2 Mar 2009. [Online] Available from: Unstructured Information Management Architecture

- (UIMA) Version 1.0 [Accessed June 2020].
5. U.S. Tiwary (2008). Natural Language Processing and Information Retrieval. Oxford University Press, edition 1st.
 6. Ranjan, P., & Basu, H. V. S. S. A. (2003). Part of speech tagging and local word grouping techniques for natural language parsing in Hindi. [Online] Available from: Part of Speech Tagging and Local Word Grouping Techniques for Natural Language Parsing in Hindi 1 [Accessed June 2020].
 7. Diab, M., Hacioglu, K., & Jurafsky, D. (2004, May). Automatic tagging of Arabic text: From raw text to base phrase chunks. [Online] Available from: Automatic Tagging of Arabic Text: From Raw Text to Base Phrase Chunks [Accessed June 2020].
 8. Robert Baud, Laurence Alpay, Christian Lovis (1994). Let's Meet the Users with Natural Language Understanding. Book series: Studies in Health Technology and Informatics. Volume 12: Knowledge and Decision in Health Telematics, 103-108p.
 9. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE*. 1989; 77(2): 257–286p.
 10. Matthews J. Basic A* Path finding Made Simple, AI Game Programming Wisdom. Hingham, MA: Charles River Media, Inc.; 2002, pp. 105–113.
 11. Eslami M, Shareef H, Mohamed A. Application of artificial intelligent techniques in PSS design: a survey of the state-of-the-art methods. *Electr Rev*. 2011; 87(4): 188–197.
 12. Firschein O, Fischler MA, Coles LS, Tenenbaum JM. Forecasting and assessing the impact of artificial intelligence on society. In: *International Joint Conferences on Artificial Intelligence*. Stanford, California, 1973, pp. 105–120.
 13. Ramesh AN, Kambhampati C, Monson JR, Drew PJ. Artificial intelligence in medicine. *Ann Royal Coll Surg Engl*. 2004; 86(5): 334p.

Cite this Article

Sapna Sharma. Natural Language Processing in AI: Language Interpreter. *Journal of Artificial Intelligence Research & Advances*. 2020; 7(2): 21–27p.