© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 02 | April - June 2022
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed

A Review Study on Computational Linguistics and Natural Language Processing

Dr. Aditi Dubey

Associate Professor, Deptt of Humanities
University Institute of Technology, RGPV, Bhopal
email :aditidtbey@rgtu.net

Abstract

The study of how computers and human languages interact is known as "natural language processing," a subfield of computer science and artificial intelligence. Mathematical and computational models are used to analyse many elements of language and construct a wide range of systems in natural language processing (NLP). Included in this category are systems that use speech and natural language together. In computer science, natural language processing plays an important role since many elements of the subject deal with the linguistic aspects of computation. To better interpret and handle natural language text or voice, computers are increasingly turning to natural language processing (NLP). Translating and summarizing material in natural language, user interfaces, cross-linguistic information retrieval (CLIR), speech recognition, artificial intelligence (AI) systems, and expert systems are all examples of natural language processing applications.

Keywords: Artificial Intelligence, Natural language Processing; cross-linguistic information retrieval, speech recognition

Introduction

Language and natural language processing (NLP) are the focus of this study, as well as linguistics' cooperation with machine translation (MT). Only one direction of such cooperation is addressed, namely the use of language theory in NLP and its possible applications, which can range from providing computer-aided research tools and aids for the study of language to implementing formal linguistic theories, verifying models, and using NLP to implement these theories and models.

Reasons for applying Linguistics to NLP

The field of linguistics is meant to know how to structure a piece of writing. An enlightened approach holds that the basic purpose of linguistics is to investigate the mental processes that underlie language. In order to replicate or imitate the native speaker's language ability, linguistics studies the indirect implications of his/her speech output, as well as the rules that control this output, because direct observation is difficult.

It sores down to matching sounds and meanings for the native speaker when it comes to language proficiency. Rather than rely on an unstructured media made up of many interconnected tiers of unrelated material, this is accomplished through a very organized medium. Text linguistics/discourse analysis and phonetics/phonology, morphophonology/morphology and morphology are among these levels.

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 02 | April - June 2022
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed

Using an MSMD (mechanism, symbol, manipulative device) strategy entails all of the elements listed below.

- without the use of any explicit criteria or methodological rigour, gathering all essential data
- utilizing the formal object to describe or construct a set of linguistic things that have a particular attribute
- native speakers have an intuitive capacity to differentiate each language entity with the property (or qualities) in issue from any other linguistic entity that does not have it (practically without ever attempting to verify that empirically)
- that the formal object can be improved so that it contains only those entities that the native speaker assigns the property to, and nothing else, and that the set of entities described (or generated) contains only those entities that are assigned the property or properties in question by the native speaker.

A MSMD linguistic theory, such as Chomsky's transformational grammar, relies on grammaticality as a problematic quality. Even anti-Chomskian linguistic ideas have lately followed the MSMD framework and sought to uncover and/or postulate a set of norms.

For a while it appeared that the MSMD format brought linguistics and computer science near enough that NLP could be directly implemented using the rules and sets of rules given by the former for NLP. This, however, should not lead to the opposite reaction, which has been demonstrated by many NLP specialists and groups, that linguistics is practically worthless for NLP.

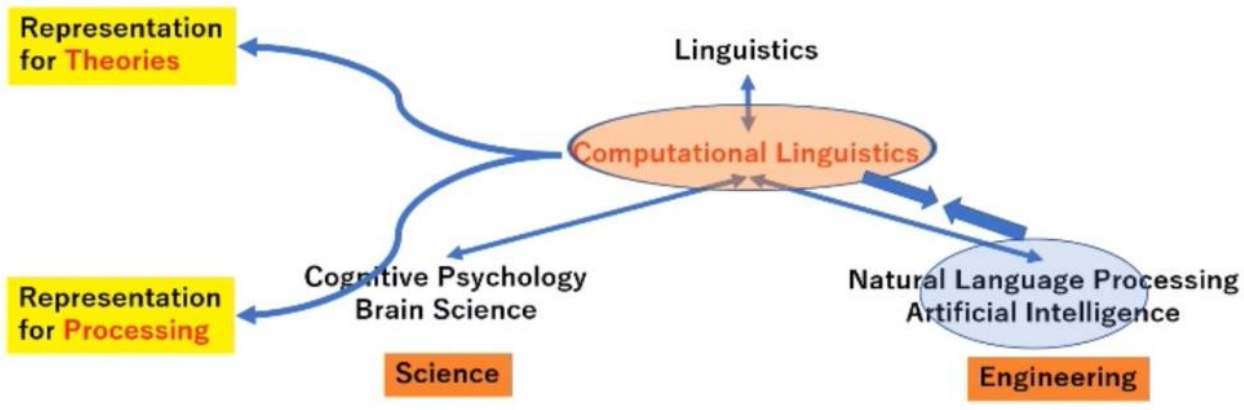

Figure 1: Relation between computational linguistics and natural language processing

Every NLP practitioner knows that describing the morphology, syntax, and semantics of a natural language is a necessary step in the process. Because of this, linguistics already possesses nearly everything an NLP practitioner would need to succeed in this circumstance, even if information isn't in their preferred style or language of comfort. The options to using this resource are outlined in this document:

- difficult to utilise published grammatical descriptions, which are often imprecise and inconvenient
- devoting all of one's time and resources to the use of dictionaries that are only available in one language (bilingual dictionaries are even worse)
- employing one's own (or a colleague's) inherent competency, which inevitably leads to the reinvention of the wheel, and quite frequently, the wheel does not even come off completely round

If linguistics is ignored or not employed, some mix of the aforementioned points is deployed in many projects, and a price is paid for it in terms of both efficiency and quality.

In terms of phonology/phonetics, this level has little bearing on NLP unless an NLP system considers acoustic input/output, which is rare and unrealistic at this stage. Although a sophisticated system could identify acceptable spelling variants, orthography, its textual counterpart, offers nothing as well. Other possibilities include considering spelling as self-correcting codes and creating an algorithm that corrects misspellings to the closest lexical term that is correctly spelt. However, there are numerous word pairings with a distance of one in most languages, including English. As with its oral counterpart, phonotactics, using graphotactics as a code to detect (but not repair) spelling errors is more grounded in reality. Permitted sounds in a language are covered in this section, whereas permissible letters (or other graphemes) in that language are covered in this section. An algorithm based on graphotactics would eliminate strings from the equation. However, this may also be accomplished by searching for a term in the system's vocabulary and not being able to locate it.

Levels of NLP

The 'levels of language' approach to Natural Language Processing is the best way to explain what is actually happening. The synchronic model of language differs from the prior sequential model in that it assumes that human language processing proceeds in a strictly sequential fashion at each stage of development.

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 02 | April - June 2022
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed

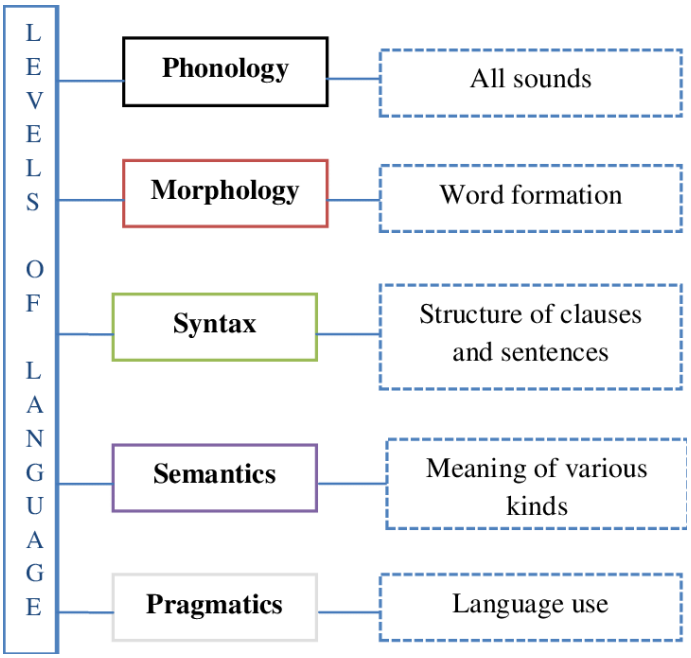

Figure 2: Levels of NLP

According to psycholinguistic research, language processing is far more complex than previously thought. Introspection demonstrates that we regularly employ knowledge gleaned from a higher level of processing to aid in a lower level of analysis. For example, if you know that the paper you're reading is about biology, you'll use that information to interpret a phrase that has several meanings as having a biology context. The following explanation of levels must be provided in a sequential order for the sake of clarity. To put it simply, all levels of language communicate meaning, and since humans have been proved to use all levels of language in order to comprehend, the more capable an NLP system is, the more levels of language it will employ.

A. Phonology:

The interpretation of speech sounds inside and across words is the focus of this level. In phonological analysis, there are really three sorts of rules:

1) Phonetic rules: It's utilised to make words sound better.

2) Phonemic rules: Pronunciation variations can be achieved by using this technique when words are pronounced simultaneously.

3) Prosodic rules: It's a quick way to see if a sentence's emphasis or intonation changes during the course of the text.

Analyzed and converted into a digital signal, sound waves in an NLP system that accepts spoken input are interpreted by various rules or compared to the particular language model being employed.

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 02 | April - June 2022
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed

B. Morphology:

After receiving information, morphology is the first step in the analysis process. It investigates the grammatical state of words by examining how they are broken down into their constituent parts. The primary application of morphology is to identify the many parts of speech in a sentence, as well as the relationships between the various words that make up that phrase. Forsberg's comment below provides some background information on the science of morphology.

The study of the structure and meaning of words in a language is known as morphology. Surface-lexical linkages are described in this framework. A study of the word into its lemma (also known as its dictionary definition) and its grammatical description is known as the lexical form of a word. If you want to be more technical, you might name this process inflectional morphology. The ability to recognize a word's grammatical context relies on being able to identify its part of speech. Irregular verbs, on the other hand, don't follow the same set of rules as regular verbs, hence the complexity of a language increases significantly. Syntactical analysis, which focuses on the grammatical structure of the target language, is informed by the morphological data that was collected earlier.

- **Syntax**

The goal of syntactic analysis is to establish the role of each word in a sentence and organize this data into a structure that can be more easily manipulated for subsequent research. The study of meaning in words and phrases is called semantics.

a) Grammar:

A noun phrase, a verb phrase, and, in certain situations, a prepositional phrase make-up a statement in English. As the name suggests, a noun phrase is used to refer to something that can be summed up in one word. As well as the noun itself, this phrase may have articles and adjectives and/or a verb phrase contained inside it. There may be an embedded noun phrase in a verb phrase that denotes an activity. An adverb or prepositional phrase is used to characterize a noun or a verb. Verbs, nouns, adjectives, adverbs, conjunctions, pronouns, and articles are some of the most common components of speech in natural languages.

b) Parsing:

Syntactic structure is represented in a tree form by the process of parsing a phrase The noun phrase "The green book" and the verb phrase "is sitting on the desk" combine to form the statement: "The green book is sitting on the desk." Rather than starting with the whole phrase, the sentence tree would begin with a noun phrase and go from there. The articles, adjectives, and nouns would then be labelled. Parsing is the process of determining the validity of a sentence in light of the grammatical rules of a given language.

C. Semantics:

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 02 | April - June 2022
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed

Objects and behaviours that are described in a sentence are depicted using adjectives, adverbs, and propositions to form a picture of the scene. This procedure collects data essential to the pragmatic analysis, which aims to discover the user's intended meaning.

D. Pragmatics:

An utterance in a human language is "analysed for its true meaning through disambiguation and contextualization" in the field of pragmatics. This is performed through the use of several disambiguation techniques to discover and resolve any observed ambiguities by the system.

- **Ambiguity:**

"The difficulty that an utterance in a human language might have more than one conceivable interpretation" is referred to as "ambiguity."

Types of Ambiguity:

- When there is more than one way to parse a statement, this is called syntactic ambiguity. "The branch with the red leaf was pulled by him." The phrase "with the red leaf" may be understood as a prepositional phrase describing the action rather than the branch, meaning that he utilised the red leaf to raise the branch as part of the embedding noun phrase describing the branch.
- For example, in the line "He picked up the branch with the crimson leaf," there is semantic ambiguity since there is more than one alternative interpretation. If a red leaf was used to raise the branch, it may imply that the individual in question pulled a red-leafed branch.
- When terms like "it," "he," and "them" are used to allude to something without specifically naming it, the outcome is referential ambiguity. For example, in the line "The interface delivered the peripheral device data which caused it to fail," it might refer to either the peripheral device or data, making a definitive determination impossible.
- Local ambiguity arises when a segment of a statement is ambiguous, but when the sentence as a whole is evaluated, the ambiguity may be resolved. ambiguity in the phrase "is colder than" is seen in the statement "this hall is cooler than the room."

Literature review

(Tyagi, 2021) In accordance with the study done it can be clearly stated that NLP is much superior techniques employed in comparison to other ways as NLP is having capacity for recognizing the text and speech additionally and other method such as text mining just deals with the evaluation of text quality. For NLP system you are having less knowledge requirement of abilities like NLTK, competency in neural networks whereas in text mining you are having requirement of information about different elements such as cosine similarity or feature hashing, text processing, like Perl or python. Knowledge of statistical approach is another extremely crucial feature in text mining. In line with the study of recent 10 years it can be clearly proven that NLP is being utilised more in comparison to other statistical approaches as it is more practicable, easier to use and less knowledge aspect needed.

(Pandey & Rajput, 2020) In comparison to other approaches to information technology, NLP is a relatively new area of research and application. However, there have been enough successes to date that suggest that NLP-based information access technologies will continue to be a major area of research and development in information systems now and far into the future. Speech technologies, notably Text-to-Speech synthesis and Automatic Speech Recognition, benefit from cutting-edge Natural Language Processing approaches. In three and a half minutes. This shows the importance of NLP in the processing of the input text to be synthesized, which is shown. When the signal-processing modules create utterances that sound natural, they are directly linked to how well the prior text-processing modules performed. Complementary usage of NLP is very important in ASR. A specified set of grammatical rules is assumed to be followed in order to simplify the process of voice recognition. As a result of its NLP-enhanced features, it can, nevertheless, be improved. In order to benefit from this specialized information, this article reviews the primary methodologies presented in language model adaptation.

(Kalyanathaya et al., 2019) Processing of Natural Language includes the following stages: lexical (structure) analysis, parsing and semantic analysis; integration of conversation; and pragmatic analysis. Speech Recognition, OCR, Machine Translation, and Chatbots are just few of the well-known applications of NLP. More recently, methods for machine learning have been employed to interpret Natural Language input by looking at millions of human-written samples of words, phrases, and paragraphs. It is via this study that the training algorithms develop a better knowledge of the "context" of the human language. Algorithms based on machine learning and deep learning are frequently employed in NLP framework development and routine task execution.

(Young et al., 2018) To get the most out of enormous amounts of computing and data, deep learning provides a low-cost alternative. Various deep models have become the current state-of-the-art approaches for NLP challenges thanks to distributed representation. In contemporary deep learning research for NLP, supervised learning is the most common method. There are numerous real-world situations where we need powerful semi-supervised or unsupervised ways to analyse data. Methods like zero-shot learning should be used when there is a dearth of labelled data for some classes or when a new class appears while testing the model. Deep learning-based NLP research is still in its infancy, but we expect it to focus on making greater use of unlabeled data. More and better model designs are likely to continue this trend in the near future. Reinforcement learning technologies, such as conversation systems, are likely to be used in more NLP applications in the future. Research on multimodal learning is also expected to increase in the future, as language is often based on (or connected with) other signals in the real-world.

(Gelbukh, 2011) Translating should be made easier by the universals of the language, one would imagine It's true that translating into a language that's not spoken by humans, such an artificial language or a language spoken by extraterrestrials, will be more difficult. There are a few practical applications for most universals (e.g., the universal stating that each natural language uses three levels of entities, sound, word, and sentence). The shift from a general linguistic contribution to translation to a specific linguistic contribution to machine translation is mostly driven by the selection function of the translator. When translating from one language to another, translations are frequently rated on a scale that ranges from excellent to barely adequate.

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 02 | April - June 2022
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed

(Grosz, 2010) For the most part, the output of a computerized text processing system has a different format from the text input that was used to create it. Translation of ambiguous natural language queries and texts into unambiguous internal representations, on which matching and retrieval may take place, is the primary purpose of natural language text processing systems.

Conclusion

NLP's possible societal impact was examined in this position paper, along with how practitioners might mitigate it. The problem of exposure can only be addressed via rigorous study design, and the problem of dual-use can only be addressed by a community effort.

With this study, we aim to draw attention to ethical issues in data collection, experimental design, and the evaluation of our systems' possible applications while also kicking off a field conversation about the method's flaws and shortcomings.

References

Gelbukh, A. (2011). Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6608 LNCS*(PART 1). https://doi.org/10.1007/978-3-642-19400-9

Grosz, B. J. (1982). Natural language processing. *Artificial Intelligence*, *19*(2), 131–136. https://doi.org/10.1016/0004-3702(82)90032-7

Kalyanathaya, K. P., Akila, D., & Rajesh, P. (2019). Advances in natural language processing –a survey of current research trends, development tools and industry applications. *International Journal of Recent Technology and Engineering*, *7*(5), 199–201.

Pandey, V. K., & Rajput, P. (2020). Review on natural language processing. *Journal of Critical Reviews*, *7*(10), 1170–1174. https://doi.org/10.31838/jcr.07.10.230

Tyagi, A. (2021). A Review Study of Natural Language Processing Techniques for Text Mining. *International Journal of Engineering Research & Technology (Ijert)*, *10*(09), 586–589. www.ijert.org

Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing [Review Article]. *IEEE Computational Intelligence Magazine*, *13*(3), 55–75. https://doi.org/10.1109/MCI.2018.2840738