



Levels of Natural Language Processing

Nadezhda Borisova, Elena Karashtranova

Department of Informatics, South-West University "Neofit Rilski"

Abstract

Undoubtedly, the Internet plays a role in contemporary society's day-to-day activities. This worldwide system offers billions of users a wide range of information resources and services. Books, news, a range of reports, scientific studies, and other sorts of material are continuously produced. A system must be able to process natural language in order to retrieve a significant amount of the data kept on computers. The interaction between computers and human languages, or natural languages, is the focus of the field of artificial intelligence and linguistics known as natural language processing.

Introduction

The interaction between computers and human languages, or natural languages, is the focus of the field of artificial intelligence and linguistics known as natural language processing. It is a very active area of research because of the wide range of information stored in natural language and accessible via computers. Liddy (2001) offered the following definition for NLP: "Natural Language Processing is a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis to achieve human-like language processing for a range of tasks or applications".

For proposed definition are further detailed several elements:

- By *range of computational techniques* is meant the multitude of existing methods or techniques that can be chosen to implement different types of linguistic analysis.
- *Naturally occurring texts* can originate from any language practiced by humans to communicate with one another.
- The *levels of linguistic analysis* refer to the different combinations of linguistic analysis utilized by various NLP systems.
- For a *range of tasks or applications* is meant that NLP is meant for accomplishing a particular task, which is utilized in systems for Information Retrieval (IR), Machine Translation (MT), Question-Answering, etc.

Levels of natural language processing

Two distinct focuses in Natural Language Processing are pointed out – natural language understanding (NLU) and natural language generation (NLG). The task of NLU is comparable to the role of reader->listener, and its purpose is to understand and create a meaningful representation of input data as natural language. The NLG, also referred to as text generation, represents a software process that produces natural language output.

Natural language processing is involved in the following stages (Figure 1):

- **Morphology** deals with the constituent morphemes in a word, which are the minimal meaning-bearing units in a language. The word's morphological structure may include elements such as prefixes, root, suffixes, circumfixes, and infixes, which can modify its meaning. The NLP systems likewise aim to gain and represent a word's meaning by recognizing the meaning of its morphemes.
- The **lexical analysis** aims to interpret the meaning of words within their context. This process contributes to detecting the relationship between individual words in the sentence and involves breaking down the text into meaningful elements such as words, paragraphs, and sentences. The word's potential parts of speech are also assigned in lexical analysis.
- **Syntactic** or Syntax analysis, also called "parsing", is used to analyze the words in a sentence conforming to formal grammar rules. Grammatical rules are applied to categorize and group the words into phrases. It involves identifying the parts of speech, determining the relationships between them (such as subject-verb agreement), and constructing a parse tree to represent the sentence's hierarchical structure.
- **Semantic** analysis is a complex process of understanding and interpreting the meaning of words, phrases, and sentence structure in the way humans do. That is the reason it still needs to be fully solved. It extracts and examines the exact meaning of text or dictionary definition by mapping the domain's syntactic structures and objects. Semantic analysis may include the semantic disambiguation of words that have multiple senses. Semantic disambiguation permits the one sense of polysemous words to be selected and included in the semantic representation of the sentence.

- In the phase of **discourse integration**, the meaning of the current sentence is determined by the meaning of the sentences that precede it. The meaning of the current sentence, in turn, impacts the meaning of the following sentence. At this level, one of the most common types of discourse processing is Anaphora Resolution. Anaphora resolution is the process of identifying the link between the anaphor (i.e., repeated reference, most commonly in the form of a pronoun) and the previous mention of the entity to which they refer.
- **Pragmatic Analysis** is the last level of NLP and deals with the purposeful use of language in real-world situations for extracting extra meaning by analyzing the context. It uses a set of rules that describe cooperative dialogues, relying on knowledge of general principles of human communication to help to discover the intended effect. For Example, "Open the window" is interpreted as a request instead of an order when used in a sentence.

Phonology is one additional level performed before Morphological Analysis and deals with sounds of language and their transforming into recognizable morphemes and words. After that transformation, the same levels of processing mentioned above on written text are utilized.

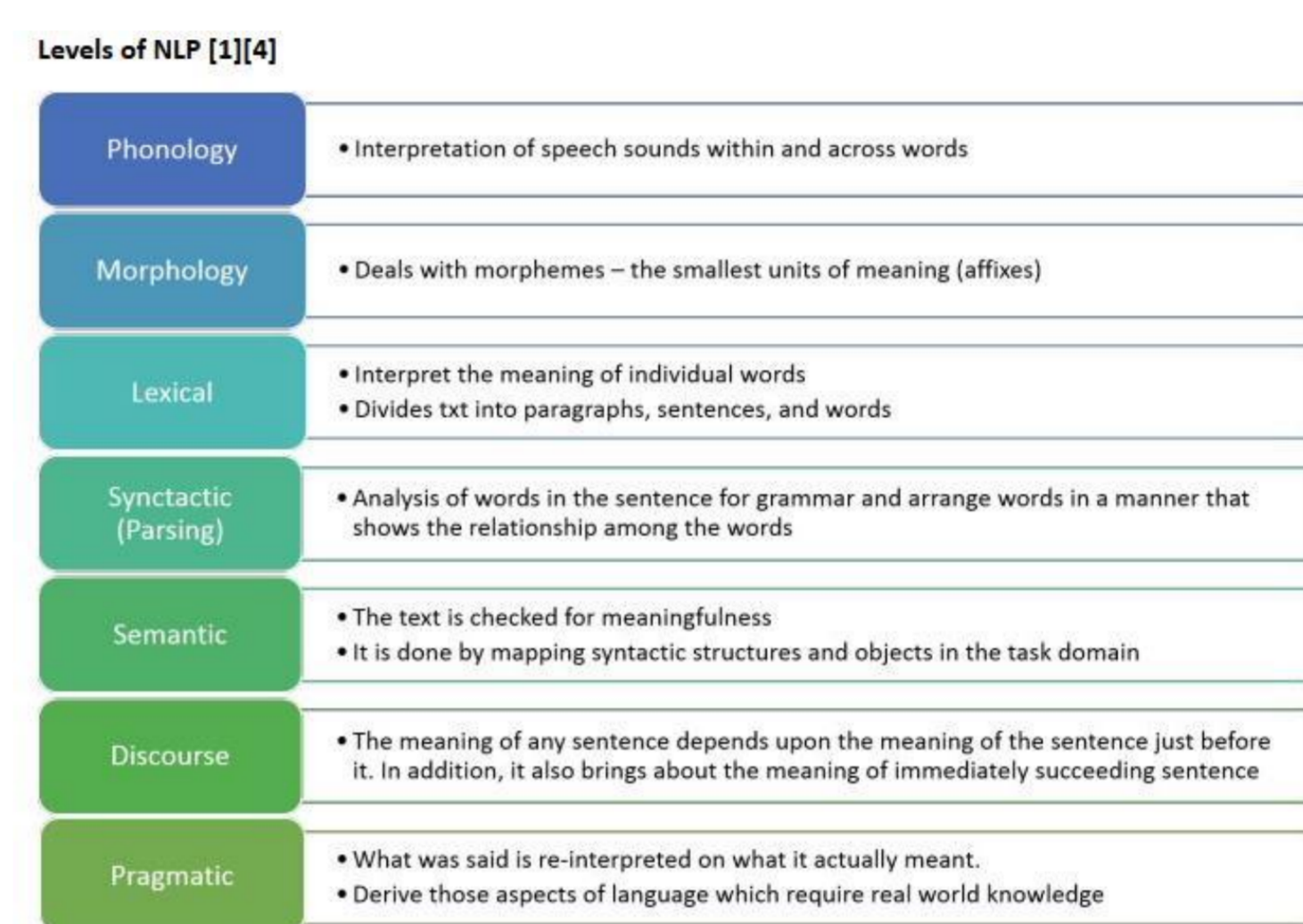


Figure 1. Different levels of NLP (Source: Forbesindia – <https://bit.ly/3AfYNws>)

Conclusion

In recent years, the volume of unstructured data has increased and created new challenges for information extraction techniques. The important information included in this large volume of unstructured data needs to be extracted in appropriately structured forms.

Natural language processing is an area of artificial intelligence that gives machines the ability to extract meaning from data represented in human languages. The lower levels of NLP concentrate on smaller units of analysis (such as morphemes, words, and sentences) and have been more detailly researched. In addition, these units are rule-governed, which can be more achievable to implement. These are the major reasons for the tendency in NLP systems to implement mostly modules for processing at the lower levels.

Acknowledgments: This work was supported by Project No. RP-A7/22 – "Data mining of students' behavior in distance e-learning systems: a case study LMS Blackboard," South-West University "N. Rilski," Blagoevgrad, Bulgaria

Bibliography

1. Liddy, E.D. (2001). *Natural Language Processing*. In *Encyclopedia of Library and Information Science*, 2nd Ed. NY: Marcel Decker, Inc.
2. Chopra, A., Prashar, A., & Sain, C. (2013). *Natural Language Processing*. *International Journal of Technology Enhancements and Emerging Engineering Research*, 1, 131-134
3. Khyani, D., Siddhartha, B. S., Niveditha, N. M., & Divya, B. M. (2021). *An interpretation of lemmatization and stemming in natural language processing*. *Journal of University of Shanghai for Science and Technology*, 22(10), 350-357
4. Guohuan, L., Hao, Z. & W. Honghui, (2009) *Research on Methods of Semantic Disambiguation about Natural Language Processing*, *International Conference on Wireless Networks and Information Systems*, Shanghai, China, 347-349, doi: 10.1109/WNIS.2009.21.
5. Webber, B., Egg, M., & Kordoni, V. (2012). *Discourse structure and language technology*. *Natural Language Engineering*, 18(4), 437-490
6. Klosowski, P. (2018), *Deep Learning for Natural Language Processing and Language Modelling*, *Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, Poznan, Poland, 223-228, doi: 10.23919/SPA.2018.8563389
7. Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). *Natural language processing: State of the art, current trends and challenges*. *Multimedia tools and applications*, 82(3), 3713-3744.
8. Singh, S. (2018). *Natural language processing for information extraction*. *arXiv preprint arXiv:1807.02383*.