**NOVATEUR PUBLICATIONS**
**JournalNX- A Multidisciplinary Peer Reviewed Journal**
**ISSN No: 2581 - 4230**
**VOLUME 8, ISSUE 12, Dec. -2022**

# LINGUISTIC ISSUES AND METHODS IN COMPUTATIONAL LINGUISTICS

Komoliddin Elbobo ugli
Tashkent University of Information Technologies
"Artificial Intelligence"

Sojida Rayimberdi kizi Ochilova
Tashkent University of Information Technologies Karshi Branch
"Information Technologies"

Elchin Uktamovich Murodovich
Tashkent University of Information Technologies Karshi Branch
"Software Engineering"

**ABSTRACT:**
After we have considered the general nature of the linguistic problems of informatics, their composition, origin, structure and prospects, it becomes clear that specific problems depend on the type of linguistic unit supplied to the computer input. In our consideration, we will adhere to the principle of succession from the smallest linguistic unit to the largest: from phoneme, grapheme, morpheme to word form, word, phrase, statement, sentence, text as a whole. Linguistics issues and methods and applications are discussed in Computational Linguistics.

**Keywords:** Computational linguistics (CL), Artificial intelligence (AI), Morphology, lemmatization, natural language processing (NLP), machine learning (ML).

## INTRODUCTION:

At present, one can see the achievements of the science of computer linguistics in all aspects of human development. Computers and robots specialized for the most complex functions are the result of the development of this science. The result of the research conducted abroad regarding modern technology and computer linguistics can be summarized as follows.

1) an artificial language was created based on natural language processing, which serves as a machine language. Among other things, an artificial language based on English was introduced into the computer to ensure the relationship between man and machine. Artificial language is the main and only communicative language of all created machine programs. became a tool;
2) computer linguistics serves as a basis for the formation of virtual relations;
3) computer linguistics ensured the emergence of translation, which is a scientific achievement;
4) in the field of computer linguistics, programs aimed at solving various linguistic problems have been created and such research is ongoing. They mainly include programs related to text editing, automatic translation, analysis and synthesis, natural language processing, computer lexicography;
5) computer linguistics is developing as a practical science based on its theoretical foundations. Its theoretical foundations are seen in the creation of various models (hypotheses) of the speech process, text learning, and in the development of its theoretical foundations. The practical nature of the science is determined by the creation of translation machines and the organization of its working process. The following three methods and methods are widely used in foreign computer linguistics:

a) logical-mathematical methods: negation, conjunction, disjunction, etc.;
b) theoretical - informative methods;
d) probability-statistical methods.

Computational linguistics (CL) is the application of computer science to the analysis and comprehension of written and spoken language. As an interdisciplinary field, CL combines linguistics with computer science and artificial intelligence (AI) and is concerned with understanding language from a computational perspective. Computers that are linguistically competent help facilitate human interaction with machines and software.

Computational linguistics is used in tools like instant machine translation, speech recognition systems, text-to-speech synthesizers, interactive voice response systems, search engines, text editors and language instruction materials.

**METHODS:**

Morphology is a section of grammar, the main objects of which are the words of natural languages, their significant parts and morphological features. The tasks of morphology, therefore, include the definition of the word as a special linguistic object and the description of its internal structure. Morphology is a branch of linguistics that studies the structure of words and their morphological characteristics.

Computer morphology analyzes and synthesizes words by software. In the most familiar formulation, morphological analysis means the definition of a lemma (the basic, canonical form of a word) and its grammatical characteristics. In the field of automatic data processing, the term "normalization" is also used, denoting the setting of a word or phrase in a canonical form (the grammatical characteristics of the original form are not given). Inverse problem, i.e. setting the lemma into the desired grammatical form is called the generation of the word form.

One important remark should be made right away.

If in traditional linguistics "for a person" the morphology of a word is rightly understood as that and only that which relates to its form - endings, suffixes, affixes, inflections, etc., division into the root and other parts of the word form, then in automatic text processing In natural language, morphological analysis means a procedure, as a result of which, from the form, external design of a word in a text, one can obtain information about the most diverse levels of linguistic structure.

Automatic morphological analysis of a language of a synthetic type, that is, with rich morphology and various forms, begins with the identification of word forms and, if possible, with their grouping into some functional classes. At the same time, an important point is that the difference between word formation and inflection, which plays a big role in the traditional interpretation of the morphological level of a language in linguistics, does not play a special role in computational linguistics (except when such a difference is taken into account in an applied problem), and therefore the difference between word formation and inflection in most cases is erased.

In the early years of work on machine translation, a large number of various algorithms for automatic morphological analysis were proposed for languages of the most diverse structure, differing from each other in “morphology”. To date, the task of morphological analysis, the most complex procedure at the word level, can be considered practically solved, since there are a sufficient number of satisfactorily working algorithms.

Several directions have emerged in the development of morphological analysis. One of them models the classical scheme of analysis by dividing a word form into a stem and a putative ending, and then

checking for compatibility of the ending with the remaining stem. Another direction uses the information contained in the final letter combinations.

Morphological classes of words are divided into two types:

1) fundamentally changing classes that characterize the system
fundamental changes;

2) inflectional classes of words.

There are the following types of morphological analysis:

- morphological analysis with a dictionary of basics;
- morphological analysis with a dictionary of word forms;
- morphological analysis by the method of logical multiplication;
- morphological analysis without a dictionary, using tables.

Due to the fact that in the vast majority of languages there are changes in the form of words due to grammar rules, lemmatization algorithms are necessary for the automatic dictionary to work, i.e., bringing different textual forms of a word to its canonical, either registered in the dictionary as such, or — when all word forms are included into an automatic dictionary of word forms - designated as the main, initial, carrying the main composition of information.

The operations that form the lemmatization are, in principle, few how do they differ from automatic morphological analysis. The difference lies in the final product and purpose. If the operational morphological analysis of an automated type aims to give the greatest possible information about a word, regardless of the level of the language, i.e., morphological, syntactic, semantic, lexical, etc., then lemmatization, with all the practically repetitive set of tools for its implementation, is intended only to establish a connection between the analyzed word form and one of its canonical representations, fixed in the source dictionary with all the necessary information.

Tech software companies, such as Microsoft, typically hire computational linguists to work on natural language processing (NLP), helping programmers to create voice user interfaces that enable humans to communicate with computing devices as if they were another person. A computational linguist is required to have expertise in machine learning (ML), deep learning, AI, cognitive computing and neuroscience**.**

The term computational linguistics is also very closely linked to NLP, and these two terms are often used interchangeably.

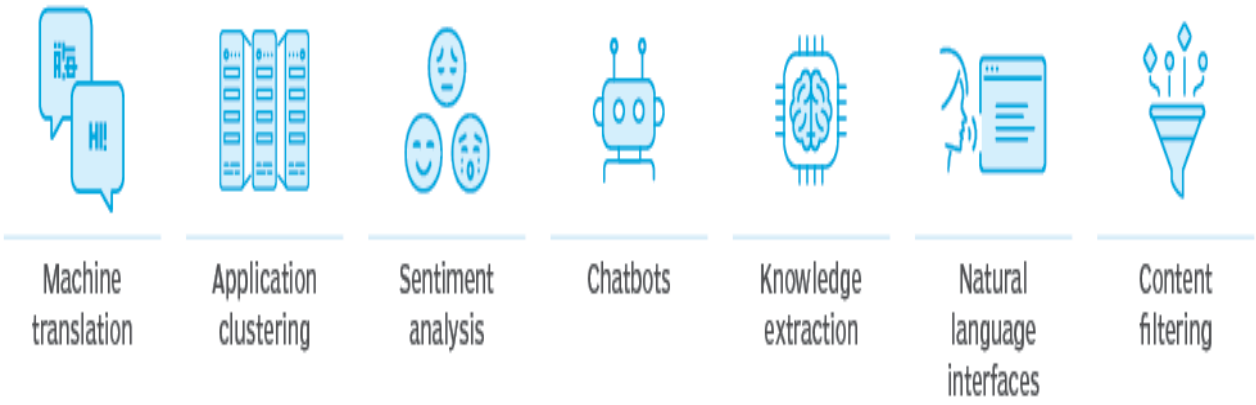


**Fig.1. Applications of comutatioanl linguistics.**

Goals of computational linguistics

Business goals of computational linguistics include the following:

- Create grammatical and semantic frameworks for characterizing languages.
- Translate text from one language to another.
- Retrieve text that relates to a specific topic.
- Analyze text or spoken language for context, sentiment or other affective qualities.
- Answer questions, including those that require inference and descriptive or discursive answers.
- Summarize text.
- Build dialogue agents capable of completing complex tasks such as making a purchase, planning a trip or scheduling maintenance.
- Create chatbots capable of passing the Turing Test.

**CL vs. NLP**

Computational linguistics and natural language processing are similar concepts, as both fields require formal training in computer science, linguistics and machine learning. Both use the same tools, such as machine learning and AI, to accomplish their goals, and many NLP tasks need an understanding or interpretation of language.

Where NLP deals with the ability of a computer program to understand human language as it is spoken and written, CL focuses on the computational description of languages as a system. Computational linguistics also leans more toward linguistics and answering linguistic questions with computational tools; NLP, on the other hand, involves the application of processing language.

**Applications of Computational Linguistics**

Most work in computational linguistics -- which has both theoretical and applied elements -- is aimed at improving the relationship between computers and basic language. It involves building artifacts that can be used to process and produce language. Building such artifacts requires data scientists to analyze massive amounts of written and spoken language in both structured and unstructured formats.

Applications of CL typically include the following:

- **Machine translation**. This is the process of using AI to translate one human language to another.

- **Application clustering.** This is the process of turning multiple computer servers into a cluster.
- **Chatbots.** These software or computer programs simulate human conversation or *chatter* through text or voice interactions.
- **Knowledge extraction.** This is the creation of knowledge from structured and unstructured text.
- **Natural language interfaces.** These are computer-human interfaces where words, phrases or clauses act as user interface controls.
- **Content filtering.** This process blocks various language-based web content from reaching end users.

Approaches and methods of computational linguistics

There have been many different approaches and methods of computational linguistics since its beginning in the 1950s. Examples of some CL approaches include the following:

- The **corpus-based** approach, which is based on the language as it is practically used.
- The **comprehension** approach, which enables the NLP engine to interpret naturally written commands in a simple rule-governed environment.
- The **developmental** approach, which adopts the language acquisition strategy of a child -- acquiring language over time. The developmental process has a statistical approach to studying language and does not take grammatical structure into account.
- The **structural** approach, which takes a theoretical approach to the structure of a language. This approach uses large samples of a language run through CL models so it can gain a better understanding of the underlying language structures.
- The **production** approach, which focuses on a CL model to produce text. This has been done in a number of ways, including the construction of algorithms that produce text based on example texts from humans.
- The **text-based interactive** approach, in which text from a human is used to generate a response by an algorithm. A computer is able to recognize different patterns and reply based on user input and specified keywords.
- The **speech-based interactive** approach, which works similarly to the text-based approach, but the user input is made through speech recognition. The user's speech input is recognized as sound waves and is interpreted as patterns by the CL system.

**History of Computational Linguistics**

Although the concept of computational linguistics is often associated with AI, CL predates AI's development, according to the Association for Computational Linguistics. One of the first instances of CL came from an attempt to translate text from Russian to English. The thought was that computers could make systematic calculations faster and more accurately than a person, so it would not take long to process a language. However, the complexities found in languages were underestimated, taking much more time and effort to develop a working program.

Two programs were developed in the early 1970s that had more complicated syntax and semantic mapping rules. SHRDLU was a primary language parser developed in 1971 by computer scientist Terry Winograd at MIT. SHRDLU combined human linguistic models with reasoning methods. This was a major accomplishment for natural language processing research.

Translating languages was a difficult task before this, as the system had to understand grammar and the syntax in which words were used. Since then, strategies to implement CL began moving away from

procedural approaches to ones that were more linguistic, understandable and modular. In the late 1980s, computing processing power increased, which led to a shift to statistical methods when considering CL. This is also around the time when corpus-based statistical approaches were developed. Modern CL relies on many of the same tools and processes as NLP. These systems may use a variety of tools, including AI, ML, deep learning and cognitive computing. As an example, GPT-3, or the third-generation Generative Pre-trained Transformer, is a neural network machine learning model that produces text based on user input. It was released by OpenAI in 2020 and was trained using internet data to generate any type of text. The program requires a small amount of input text to generate large relevant volumes of text. GPT-3 is a model with over 175 billion machine learning parameters. Compared to the largest trained language model before this, Microsoft's Turing-NLG model only had 17 billion parameters.

**Summary**

Currently, the process of computerization is taking place at different levels in the countries of the world. Regardless of the level of progress, humanity has come to understand the incomparable role of information technology in society. Computerization of all areas of human activity is an important task of society and a factor of social development today. difficult The field of computer science emerged as a result of this need. Computational linguistics is the effective use of computers and issues related to linguistics (acquiring information style, gaining knowledge about the functional scope of the language, teaching languages, evaluating knowledge, editing and analyzing texts, translating from one language to another) by means of a computer. It involves defining the ways of solving, increasing the level of computer literacy, teaching logically correct and consistent thinking, forming theoretical knowledge and forming skills related to practical application in certain directions.

**References**

1. Авербух К. Я. Общая теория термина. Иваново, Ивановский госуниверситет, 2004. 251 с
2. Вельская И. К. Язык человека и машина. М.: Изд-во МГУ, 1969. Т. 1.408 с. Т. 2. 250 с.
3. Диалог-95. Труды международного семинара по компьютерной лингвистике и ее приложениям. Казань, 31 мая — 4 июня 1995. 362 с.
4. Зеленков Ю. Г. Морфологический анализ в системах автоматической обработки научно-технической информации. Канд. дис. М.: ВИНИТИ, 1988. 145 с.
5. Маковский М. М. Лингвистическая комбинаторика. М.: Наука, 1988. 232 с.
6. Рябцева Н. К. Лингвистическое моделирование естественного интеллекта и представление знаний. В кн: Проблемы прикладной лингвистики 2001. М.: РАН, 2002. С. 228-252.
7. Шуклин Д. Е. Морфологический и синтаксический разбор текстов как конечный автомат, реализованный семантической нейронной сетью, имеющей структуру синхронизированного линейного дерева. В кн: Новые информационные технологии. М.: МГИЭИМ, 2002. С. 74-85.
8. Jurafsky D., Martin J.H.Speech and Language Processing: An introduction to natural language processing, computational linguistics and speech recognition (2007).
9. Daniel Jurafskiy& James H.Martin. “Speech and Language Processing(2008).
10. S.Yu and L.Deng “ Automatic Speech Recognition: A Deep Learning Approach” (Publisher: Springer),2014.

11. Mairesse F .(2011).Conroling user perceptions of linguistic style: Trainable generation of personality traits.Comutational linguistics.
12. Марчук Ю.Н. Компьютерная лингвистика. – М.: АТС: Восток и Запад, 2007.