AMERICAN Journal of Language, Literacy and Learning in STEM Education
Volume 4, Issue 06, 2026 ISSN (E): 2993-2769

The Importance of Corpus Linguistics in Computational Linguistics

N. Norpolotova
SamDCHTI Foundation Ph.D.

Abstract: *This article highlights the importance of corpus linguistics in computational linguistics. It also provides general information about what corpus linguistics does. For example, its role in the automatic analysis and processing of natural language, the creation of electronic dictionaries, and the development of automatic translations are discussed. This reflects the interconnectedness of corpus linguistics and the importance of these disciplines in modern linguistics. The role of these areas in the creation of scientific computers and applications is also expressed.*

Keywords: *Language Theory, Linguistics, Natural Language, Corpus Annotation, Corpus, Corpus Linguistics, Computational Linguistics, Leech*

1. Introduction

As we all know, in recent times, new modern disciplines have emerged in the field of information technology and linguistics. One of them is computational linguistics, which is the largest of the modern branches of linguistics today. It is the processing and analysis of natural language, as well as the implementation of modeling. Computational linguistics is closely related to artificial intelligence, computer technology, and databases. One of the important foundations of this science is corpus linguistics, which deals with primary texts and studies language from a practical perspective. The importance of corpus linguistics in the development of computational linguistics is that it is an important source for automatic text analysis, speech amplification, and linguistic models.

XXI century: In the age of information and technology, it is difficult to imagine all spheres of life without a computer. That is why they are used in almost all parts of the world. Gradually, a time has come when, if human society does not use these information technologies, it will feel as if it is cut off from the world. Of course, this should be associated first of all with language. Just as there is no society without language, no modern language can exist without society. First of all, it is worth mentioning what language is, the concept of language theory. The President of the Republic of Uzbekistan Sh.M.Mirziyoyev[1] speaks separately about respecting the native language at each spiritual and educational conference.

Language[2,3,4] is one of the greatest human values. Language is a means of national communication and communication, but also a source of thought. The relationship between language and society, language and thought, language and speech, its importance in society and the emergence and development of an important person, and many of its developments play a key role. The practice and study of language appeared in the 5th century BC, and it first advanced in India, Greece, Rome, Egypt, Mesopotamia, China, Transoxiana, and the Arabian Peninsula. The first writings also appeared in these countries. No matter how much we talk about language, there is no end to it. Language embodies such help.

2. Methodology

This study employed a qualitative and descriptive research approach to investigate the importance of corpus linguistics in the development of computational linguistics. The research was based on a comprehensive review and analysis of scientific literature, including textbooks, monographs, journal articles, dissertations, and international publications related to linguistics, corpus linguistics, and natural language processing. Relevant sources published by both national and international scholars were selected to provide theoretical and practical insights into the relationship between corpus linguistics and computational linguistics.

The methods of comparative analysis, content analysis, and descriptive interpretation were applied throughout the study. Particular attention was given to the role of language corpora in text processing, corpus annotation, machine translation, morphological and syntactic analysis, and the development of linguistic databases. In addition, existing corpus-based approaches and statistical methods used in computational linguistics were examined. The collected data were systematically analyzed to identify the contribution of corpus linguistics to modern language technologies and its significance in contemporary linguistic research.

3. Results and Discussion

We want to talk about the term linguistics. Linguistics, or linguistics: Greek lingua - “language”, logos - “science”, is a multifaceted and complex science that studies language in its inextricable connection with other sciences, both scientific and practical, from its emergence to the present stage of development. Language is a tool for society, plays a cultural role in the spiritual and educational work of man. Linguistics did not exist as a science before, it was part of the philosophy of use. By the 19th century, it began to take shape as an independent science. Modern[5] directions of linguistics, namely sociolinguistics, ethnolinguistics, ethnolinguistics.’arbdavistika. During this period, attention was paid to the emotional impact of language and the emotional-expressive properties of language were revealed. Linguistics[6] social nature of language[7,8] social nature of language. A science that studies language in depth is coming. Currently, there are students in the field of independent language work, and they are called:

1. Theory of Linguistics
2. General Linguistics
3. Introduction to Linguistics
4. History of Linguistics

These disciplines are used by the most effective readers in their work. The basis, language, and additional disciplines in other areas are currently being studied and processed in institutes and universities as the main methodological tools of the Uzbek language. It is also known that without a set of these disciplines, it is impossible to study the direction of linguistics in depth and thoroughly. One of the foundations of this is that until today's language was formed, it was studied based on different views and assumptions. Therefore, it is considered unacceptable to give a single definition to the language. The above disciplines [9] help the student to study all the methods of linguistics and develop into a perfect philologist. In world linguistics [10] A system has a certain system of language and its systematicity was proven and determined by Humboldt. The term computer linguistics, which is inextricably linked with linguistics, is currently receiving [11,12]a lot of attention. It is effective in linguistics. There are many sections in Uzbek linguistics. These include phonetics, syntax, morphology, lexicology, orthography, and orthoepy. Each of the three sections has its own rules, but there is no model of these and no legal rules on how to regulate them.

There are also models of linguistic departments, but there are no departments that deeply reveal the internal parts of the document. Computational linguistics is engaged in obtaining models that

reveal the most important internal parts of linguistics. The main purpose of computational linguistics [13] is to study these 3 main departments. The term computational linguistics, like other terms, is called by different names in science. Ruslan Mitkov's[14] program, the term "Computational Linguistics", was developed by David Hyde in the 1960s. After that, this term began to be published in American journals.

D. Jurafsky [15] divided the technological processes of the science of computational linguistics into separate works. The first stage falls on the 1940s-1950s. The second stage falls on the 1970s-1983s. During these periods, the founders of the science, as well as the problems and solutions of language and computers, were studied. From this, accurate and thorough models of all the processes of language entering the computer language were developed. In the subsequent period, a new direction in science, popularizing in this process, began to be the focus of everyone's attention. Now it is impossible to imagine the world of language without modern trends. , one of the most important, it can be safely said that the emergence and development of this science came about in the interaction of various sciences and fields, and without the connection of this field, a new science would not have existed. In Uzbek, computer linguistics[16] It is necessary to separately mention the works of S. Rizayev and S. Muhammedova, who conducted linguostatistical results. Computer linguistics was established in Uzbekistan in 2001 as a mathematical statistics discipline by prof. A. Pulatov introduced it, and led many products in the implementation of the famous product. In addition, many more scientific examples can be cited in this area. It can also be said that in the 80-90s of the global age of the 21st century, computer linguistics and natural language processing achieved great success.

It is difficult to imagine the science of computational linguistics without corpus linguistics. Tools have been created to perform various operations on the corpus. Corpus linguistics also has different views and theories. The famous English scientist Douglas Bieber dwells separately on the features of the corpus. It has different interpretations as a science. Some sources consider it a part of computational linguistics, some of its application, and in other sources as a separate independent science. The 18th century is the era of studying corpus linguistics. There are also electronic corpora, which have many stages. What is a corpus? This term can be defined. A corpus is a collection of oral and written texts stored in a computer database. When writing a corpus of texts, various standards and practices are used, as well as field classifications. Corpus linguistics can be said to be the main direction of computer linguistics, since there are programs that process text, which turn it into a corpus. A corpus has the property of systematically studying a language, and there are corpus systems for various languages. In corpus linguistics, there is also the concept of corpus annotation. These annotations are consistent with the Leech rule There are also stages of linguistic annotation of a corpus. There are three methods in computer linguistics:

a) Statistical method;
b) Modeling method;
c) Effective corpus analysis method;

The corpus analysis method uses such areas of computer linguistics as machine translation, morphological analysis, syntactic analysis. A number of modern areas of linguistics, including translation theory, psycholinguistics, dialectology, sociolinguistics, and psycholinguistics, also use corpus-based stylistics. Today, corpus-based teaching is widely used in world linguistics.

4. Conclusion

In conclusion, it can be said that corpus linguistics is one of the most important areas of computational linguistics. This area helps to study language units based on large volumes of texts and analyze their properties on the basis of accurate statistical data. While linguistics studies the theory of the study of language materials, corpus linguistics provides a practical approach to studying them.

It is worth mentioning that the most important factor in the development of corpus linguistics is the development of production. In the future, the further development of Uzbek language corpora, their enrichment based on new sources and the development of world standards are one of the important roles facing these disciplines.

References

[1] President of the Republic of Uzbekistan, "On the authority of the Uzbek language as a state language and measures to fundamentally enhance its use," Decree No. PF-5850, Oct. 21, 2019. [Online]. Available: https://lex.uz/docs/4561730

[2] Z. Kholmanova, Theory of Linguistics. Tashkent, Uzbekistan: Shafoat Nur Fayz, 2020.

[3] S. A. Rustamiy, "Developmental views on linguistics in the science of medieval Islam," D.Sc. dissertation abstract, Tashkent, Uzbekistan, 2018.

[4] U. Sanakulov and A. Turobov, Theory of Linguistics. Samarkand, Uzbekistan: SamSU, 2019.

[5] N. Unakulov, Theory of Linguistics. Tashkent, Uzbekistan: Barkamol Fayz Media, 2016.

[6] R. Rasulov, General Linguistics. Tashkent, Uzbekistan: Science and Technology, 2017.

[7] A. Nurmonov, Selected Works, vol. 2: History of Linguistic Data. Tashkent, Uzbekistan, 2012.

[8] O. Mamirov, "Modeling methods in computer linguistics," Philology Horizons Journal, vol. 1, pp. 1–7, 2020.

[9] N. Z. Abdurakhmonova, Computer Linguistics. Tashkent, Uzbekistan: Nodirabegim, 2021.

[10] R. Mitkov, Ed., The Oxford Handbook of Computational Linguistics. Oxford, U.K.: Oxford Univ. Press, 2003.

[11] D. Jurafsky and J. H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Upper Saddle River, NJ, USA: Prentice Hall, 2007, pp. 9–14.

[12] N. Z. Abdurakhmonova, "Problems of processing texts on a computer," Graduation Qualification Work, National University of Uzbekistan, Tashkent, Uzbekistan, 2009.

[13] N. Z. Abdurakhmonova, Machine Translation Software Support. Tashkent, Uzbekistan: Muharrir, 2018.

[14] B. Heine and H. Narrog, Eds., The Oxford Handbook of Linguistic Analysis. Oxford, U.K.: Oxford Univ. Press, 2015.

[15] G. Leech, "Corpus-based language change and usage analysis," in The Oxford Handbook of Linguistic Analysis, B. Heine and H. Narrog, Eds. Oxford, U.K.: Oxford Univ. Press, 2015, pp. 193–210.

[16] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. Cambridge, MA, USA: MIT Press, 1999.