



CENTRAL ASIAN JOURNAL OF THEORETICAL AND APPLIED SCIENCES

Volume: 04 Issue: 05 | May 2023 ISSN: 2660-5317
<https://cajotas.centralasianstudies.org>

The Significance of Parallel Corpora in Language Teaching

Shaxribon Musurmankulova
GulDU doktoranti
shmusurmankulova83@gmail.com

Received 4th Mar 2023, Accepted 6th Apr 2023, Online 23th May 2023

Abstract: *The article discusses parallel corpora, the importance of corpora in language education, parallel corpora in the system of computer linguistics, today's attention paid to the creation of computer linguistics and parallel corpora in world linguistics, the importance of parallel corpora in learning linguistics, the use of parallel corpora in the work of teachers. rni, the utility of parallel corpora in finding cross-language translation equivalents is given.*

Keywords: *corpus, parallel corpus, CLARIN infrastructure, concordancer, representative, frequency, archaism, historicism, neologism.*

It is known that today in every field, parallel bodies are used for various practical purposes. Parallel corpora are used to compare linguistic features and their frequencies in two languages. They are also used to explore the similarities and differences between the source and target languages. This allows for systematic, text-based contrastive studies at different levels of analysis. In this way, parallel corpora provide information about languages based on linguistic typological and cultural differences and similarities. Comparative linguistics is closely related to the use of parallel corpora, that is, they allow comparative analysis of two or more languages in translation studies. A parallel corpus can also help translators find translation equivalents between the original source and the target language.

Corpora have value as a resource for improving linguistic research. In addition, corpora serve as a database in the process of learning the social functions of the language, linguistics, and in the process of language education. Corpus contain important and large amounts of information about the language, its historical development, unique phonetic, lexical, and grammatical features. The importance of corpora in language education is explained by the following features:

1. A large amount of linguistic information is collected in corpora.
2. All words of the traditional layer within one language - old layer: archaisms, historicisms; modern layer; new layer words - neologisms are included.
3. As a result of subjecting the corpora to the search system, it is possible to find information quickly and easily.
4. Corpora perform statistical analysis of language units and linguistic processes with the help of concordancers.

5. The corpus of texts performs the function of providing factual material and sources.

The formation of corpus linguistics on a global scale dates back to the 60s of the 20th century. Formation of corpus linguistics is related to machine translation. The following scientists conducted research on the corpus, its types, characteristics, principles of corpus formation: A.N. Baranov [Baranov, 2001], V.P. Zakharov [Zakharov, 2005], Y.V. Nedoshivina [Nedoshivina, 2006], K. Boyarskiy [Boyarskiy, 2013], N. Kozlova [Kozlova, 2013].

The technology of creating a corpus is covered in scientific studies. A. Baranov provided information on the basic concepts of corpus linguistics, text corpus, problem area, database, corpus data storage units, research corpus, illustrative corpora, methods of displaying and storing dynamic and static text corpus [Baranov, 2001].

The following are the parallel corpora available in the world computer linguistics system:

1. English-German translation corpus
2. English-Norwegian Parallel Corpus (ENPC)
3. English-Swedish parallel corpus (ESPC). It was created in 1993. It has now become an important resource for learning English and Swedish. The corpus database contains 64 English texts and translations, 72 Swedish texts and translations. The corpus contains 2.8 million words. The texts are adapted as much as possible in terms of text type, topic, and style, so they can be used as a two-way parallel corpus and as a comparative corpus. Served as a resource for research on epistemic modality and adverbial conjunctions in English and Swedish.
4. International Telecommunication Union Corpus (English-Spanish)
5. The Intersect Parallel Corpus (English-French)
6. Multilingual parallel corpus (Danish, English, French, German, Greek, Italian, Finnish, Portuguese, Spanish, Swedish texts).

The CLARIN infrastructure provides access to 87 parallel corpora, most of which are available from national databases, as well as text downloads through concordancers such as Corp, Corpuscle and KonText. The CLARIN infrastructure contains 47 bilingual corpora, mainly European language pairs, but also Hindi, Tamil and Vietnamese. 40 corpora in the system are multilingual, and 5 contain texts in more than 50 languages. Almost half of the corpora base is almost 100% translationally adapted, which allows for easy comparative research.

The formation of corpora in world linguistics creates the basis for the transformation of languages into the language of the Internet, and the rapid and wide transmission of information. There are several types of corpora, among which parallel corpora occupy an important place in the exchange of information between languages.

A corpus is a language resource consisting of a large and systematized set of texts. In corpus linguistics, they are used to perform statistical analyses, to test views, linguistic phenomena or theoretical rules within a specific language or a specific section of the language.

Corpus linguistics is a branch of computational linguistics that develops general principles of building and using linguistic corpora (text corpora) using computer technologies. A linguistic or language corpus of texts is a machine-readable, combined, structured, marked, philologically perfect set of linguistic data designed to solve specific language problems.

Corpus are of two types according to the language of expression of the text: corpus of texts in one language; corpus of parallel texts.

A corpus of parallel texts is an electronic representation of works of art, manuals, mass media, various documents in two or more languages. For example, Gigaword corpora: covers English, Arabic, Chinese languages and consists of 2 billion words. Aquis ommunautaire is the world's largest corpus of parallel texts.

Currently, parallel corpora have been created that reflect the text features of English, German, Japanese, Finnish, and Slovak languages with Russian [Rakhmonova, 2020:30].

Corpus should be based on natural language data. The corpus should be representative, that is, it should contain elements of different speech styles.

A parallel corpus also helps in finding cross-linguistic translation equivalents. It provides information about the frequency of words and combinations.

Parallel corpora are widely used in teachers' activities. With the help of the parallel corpus, they can identify frequent linguistic phenomena in the language, enrich their knowledge of the language, design educational materials, and perform analysis on the original source during the teaching process. Parallel corpora are actively used, especially in the process of language learning.

Parallel corpora are of practical importance in teaching Uzbek as a foreign language, learning foreign languages, translating literary sources in Uzbek and other languages. This corpus serves as a material and linguistic support in the formation of the national corpus, in the creation of educational corpuses. The use of parallel corpora is also important for discovering and learning subtle aspects of meaning in the language.

References

1. Баранов А.Н. Введение в прикладную лингвистику. – М.: Эдиториал УРСС, 2001.
2. Боярский К. Введение в компьютерную лингвистику. – Санкт-Петербург, 2013.
3. Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие. – Санкт-Петербург, 2005.
4. Козлова Н.В. Лингвистические корпуса: определение основных понятий и типология. Вестник НГУ. Серия: Лингвистическая и межкультурная коммуникация. 2013.Т.11. Выпуск 1.
5. Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – Санкт-Петербург. – 2006.
6. Рахмонова А. Ўзбек тили миллий корпусини яратишдаги компьютер усуллари. Филол.фан. фалс.д-ри. ...дисс. – Тошкент, 2020.