PROBLEMS AND SOLUTIONS: INDIAN DIACHRONIC CORPORA

Nozimjon Bobojon ugli Ataboev
Doctor of philosophy (PhD) in philology, associate professor
Dean of the Faculty of Foreign Languages of BukhSU
E-mail: n.b.ataboyev@buxdu.uz

Annotation: The following paper informs about the theoretical data regarding the achievements of English corpus linguistics and the problems of Indian corpus application, creation and multilingualism.

Key words: Corpus linguistics, diachronic corpus, multilingualism, language history, etymology

Introduction

At the beginning of the XIX century, as a result of the formation of theoretical principles of corpus linguistics, the first example – the Brown corpus, was created. As the field of linguistics, which deals with the creation and application of language corpora, the interaction of corpus linguistics with computer technology was intensified in the 1960s. Automatic digitization of the intellectual wealth of mankind on the basis of the electronic corpus is one of the urgent tasks of modern linguistics today. An electronic linguistic database, which is a practical expression of a corpus analysis method, represents a language or a specific part of a language, generalizes written and spoken texts, as well as shows in concrete numbers for the statistic analysis of the development of a language, lexical layer enrichment and its present status.

While the quality of linguistic research in world linguistics is based on statistical analysis, the effective result of the practice of creating language corpus in the field allows saving time spent in the process of text analysis. The primary feature of the corpus is the concordant identification of thousands examples of linguistic units use, their properties in the structure of speech. As science enters the third decade of the 21st century and pushes the need of applying computers into practice forward, the scientific-theoretical study of corpus linguistics is gaining relevance in all areas.

Main part.

Studies have shown that the practice of creating a national language corpus for India, which consists of a multilingual population, causes a number of complications, since the constitution of this country recognizes 2 national and 22 local languages as state languages. In our opinion, the languages in the territory of this country complement each other under the influence of interaction, communication and a single media space, causing the reduction of each other's vocabulary. In our opinion, the creation of a multilingual corpus for this nation could preserve the current diversity of languages.

Indian linguist N. Dash agrees with the above points and says: "As a multilingual country, India is a language giant. It maintains a large number of existing "living" languages of various ethnic and linguistic communities. Due to the lack of corpora, these languages suffer from the fact that they cannot achieve their technological progress. Building corpora can accelerate language learning to improve literacy, preserve endangered Indian languages, and protect languages that have lost their importance from the encroachment of English. Corpora are the only linguistic bases that can help them survive the battle of linguistic imperialism. In addition, it is characterized by the fact that it can provide statistically reliable data to restore lost position" [1; P 1]. The burning of the scientist is that in the era of globalization, if research is not properly oriented, in the current development of

languages, due to the increase in the flow of popularized media text elements in other languages, weak languages will disappear or will be condemned to absorb many foreign languages, such as the Mandarin dialect of the Chinese language. This is exactly the issue that is raised in our research. After all, based on the materials that are the objects of research, if we conclude, there are languages that are ahead of us in development and production, and taking into account the fact that their influence is expanding in the prism of the media space, the practice of creating a diachronic corpus of the Uzbek language is also necessary. It is the demand of the time to set it up.

Without rejecting the above points, the study of language development means the analysis of the etymology of words, phrases and, in general, any linguistic unit that has entered the lexical layer. Etymological studies can in some sense be included in ethnographic studies. The practice of using language corpora in this field is evident in the works of K. Harrington. By the scientist, the role of corpus linguistics in the study of ethnography of a closed society is to analyze the interaction of immigrants in a refugee reception center in Ireland to determine the instinctive resourcefulness of people who are faced with the problem of communicating in the absence of a common language or culture [2]. Using three years of ethnographic observation and an innovative mix of applied methodologies, particularly corpus linguistics and conversation analysis, the following was achieved:

- Based on a corpus of 98,000 words;

- It is based on learning the use of English in the process of interaction with residents and English-speaking staff of the center;

- It was found that constructions such as speech community, communicative competence and interlinguistic interference are becoming more difficult and the purity of one language is relatively declining.

We believe that any society has a need for communication with another group of people. As this communication approaches, it affects language development in the following stages and sequence:

1. Social views harmonize with each other;

2. Religious faith and secular ideological rapprochement occurs;

3. Mutual sharing of scientific, technical and household achievements and inventions;

4. Dressing, customs, ceremonies, in a word, interference of cultures;

5. As a result of the above, the lexical layer changes and enriches under the influence of language interference.

In general, the process of interference of languages in the above mentioned five-stage sequence is accelerating today. Because the factors of the current process of communication are fundamentally different from the previous person-person and society-society categories and are being built on the basis of new, modern forms of information exchange. In our opinion, the acceleration of communication in this media field is reducing the scope of the purity and diversity of cultures and languages. Therefore, as long as the study of language development, which is the object of research, is not studied statistically based on the prism of corpora, the practice of preserving the current state of languages and leaving them as a basis for future research will not emerge. And this cannot be improved with any excuse.

According to Swedish scientist M. Kytö, electronic historical corpus and corpus methodology are research methods aimed at studying and evaluating the current state of languages and allowing to consider linguistic changes that may occur in the future. With this, the scientist emphasizes that within the wide range of corpus linguistic methodology, historical corpus linguistics has emerged as a vibrant field that has significantly increased the attractiveness for the study of language history and change. Indeed, the research process of evidence-based historical linguistics would not have been completed

without the methodology and new impetus of corpus linguistics. In an era of rapidly changing life and research, increasing competition for academic careers and opportunities for young scientists, there is no doubt that the methodologically easy field has a future. Historical corpora and other electronic resources have also made the study of language history attractive: working on them engages students individually and interactively [3; P 417].

CONCLUSION

In our opinion, it is important to create corpora of this type in the case of the English language as well. After all, if the number of manuscripts and published literature decreases at the same time when materials are being digitized, and if they are summarized in corpora in electronic form, any linguist conducting any linguistic research can come to a certain conclusion about the present day based on the history of the language.

REFERENCES:

1. Dash N.S. Language Corpora: Present Indian Need. - Indian Statistical Institute, Kolkata. 2004. – 1-6 pp. available at https://www.researchgate.net/publication/2938590_Language_Corpora_Present_Indian_Need

2. Harrington, K. The Role of Corpus Linguistics in the Ethnography of a Closed Community: Survival Communication. // International Journal of Corpus Linguistics, 24 (4). – 2018. – pp. 541-547. ISSN 1384-6655

3. Kytö M. Corpora and historical linguistics // RBLA, Belo Horizonte, v. 11, n. 2. – Uppsala University, Uppsala: Sweden, 2011 – pp. 417-457.