

## NATIONAL FOUNDATIONS OF THE CORPUS OF PHARMACEUTICAL TERMS.

<sup>1</sup>Rakhmanova Azizakhan Abdugafurovna,<sup>2</sup>Abdullaeva Albina Ergashevna,<sup>3</sup>Tursunoy Bakhtiyarovna Akhmedova, <sup>4</sup>Matyusupova Shakarjon Bekmamatovna

<sup>1</sup>Assistant professor of the Department of Uzbek language and literature, Faculty of Industrial Pharmacy, Tashkent Pharmaceutical Institute, Tashkent, Uzbekistan.

<sup>2</sup>Head of the Department of Uzbek language and literature, Faculty of Industrial Pharmacy, Tashkent Pharmaceutical Institute, Tashkent, Uzbekistan.

<sup>3</sup>Senior teacher of the Department of Uzbek Language and Literature, Tashkent Pharmaceutical Institute.

<sup>4</sup>Senior teacher of the Department of Uzbek Language and Literature, Tashkent Pharmaceutical Institute.

E-mail address: [Khudoyqulova-1982@mail.ru](mailto:Khudoyqulova-1982@mail.ru)

**Annotation:** There are discussed the creation of national bases of Uzbek pharmaceutical terms, specific stages and principles of corpus formation. The requirements for the development of computer methods for the creation of the database of the national corpus of the Uzbek language were studied. Corpus formation methods are improved by studying the experience of world corpus linguistics, and the structure and characteristics of existing corpora, the foundations of the creation of the national corpus of the Uzbek language, corpus linguistics research, corpus creation technology, tools, and computer methods.

**Keywords:** linguistics, pharmaceutical term, corpus, the corpus of texts, the corpus of parallel texts, concordance, the corpus of authorship, methodology, creative style, phonetic, lexical, grammatical features, lexicology, lexicography.

**Introduction:** Corpus linguistics is a new direction in linguistics that has existed for a little more than half a century and engaged in the creation and use of cases to solve a variety of linguistic tasks. These tasks naturally include the task of learning the language, native or foreign. National corpora play a special role in this. Such projects are supported at the national level, leading corpus linguists are involved in their creation countries, and this usually contributes to their high quality and popularity. Enclosures allow quickly and effectively check the features of the use of an unfamiliar word, the compatibility of words or grammar on a large and living language material.

A corpus is a set of texts subjected to a search program for the purpose of determining the characteristics of language units, a set of written or spoken texts in natural language stored in electronic form, placed in a computerized search system on the basis of software. Linguistic corpora are an undeniable tool for language research and solution of practical tasks [Mengliev, 2020: 2].

A corpus of texts reflects the vocabulary of a particular language. A text corpus is a large database represented by text or supersyntactic entities. The vocabulary of national corpora is not synchronic, but includes lexical units based on diachronic development. This makes it possible to analyze the phonetic, lexical, and grammatical characteristics of the general development stages of the language, to determine the size of the vocabulary, and to determine the principles of development.

Nowadays, understanding and mastering the semantics of historical, archaic, dialectal units, slang, and slang at the lexical level is somewhat complicated. Because lexemes belonging to these consumer circles are not regularly used. Historical, archaic words are found in the text of classical sources, dialectal words, slang words are actively used in oral speech. Dictionaries are a means of mastering the content of lexical units in

this semantic field. The problem is that in the lexicography of the Uzbek language, perfect dictionaries have not been created that fully cover historical, archaic, dialectal units, slang. The formation of the corpus of texts ensures the elimination of the noted problems, as well as the knowledge and mastery of all layers of the lexical level.

While classifying linguistic corpora based on various parameters, Sh.Hamroeva distinguishes monolingual, bilingual, multilingual corpora according to their parallelism. He also describes parallel corpora as follows: "Since the main task of the corpus is to collect language material, to serve as a basis for research, we can find answers to all our questions no matter what aspect of linguistics we apply to it. Naturally, this also applies to literature education.

The national corpus also includes an internal corpus, which are directed to a specific purpose. With the help of the corpus of works of art, it is possible to study the language features of the works of a certain period or the same creator, and to carry out a comparative analysis of the text of the work using the parallel corpus [2,9].

There are 2 types of bilingual corpora: text corpora that are translations of each other; a bilingual text corpus on the same topic. The first type of corpus is called "parallel corpus" and is used to study different aspects of a given translation. In turn, the parallel body is further divided into 2 types. This is an aligned (aligned), not aligned (not aligned) case. The term "aligned" means that there is a clear relationship between translation units in the corpus that requires each other. From such a corpus, it is found how this or that sentence is translated. This type of corpus is important for the translator because it contains a unique resource - "translation memory".

An unabated corpus is also called a "comparative" corpus. Aligning the text with its translation is to indicate which unit it corresponds to in the translation. The adjustment can be done automatically or manually. The first method is easy, but there are many errors. For example, in the process of translation, a simple sentence can be given in the form of a compound sentence. In this case, it will be difficult to determine which construction is original. As an example of a multilingual "adapted" corpus, we can cite the European Union's AcquisCommunaire database. The cost of a parallel corpus is determined by its size and the number of languages. AcquisCommunaire is the largest parallel corpus in the world. Two of its most important aspects are the free use of the corpus and the presence of rare language pairs such as Maltese-Estonian and Slovenian.

Experts define the importance of these corpora as follows: typical translation methods, creating transformation; study automatic translation system statistics; creation of monolingual and multilingual dictionaries; study, evaluation of data storage and transmission programs; automatic verification of the correctness of the translation; facilitating the translator's work through the wide range of equivalent options.

The second type of corpus is called "translation corpora" and is important for studying the expression of the same idea in different languages.

**Materials and Methods:** In addition, bilingual corpora are considered a very useful tool for graduate work for a student of linguistics, as they can provide some linguistic observational results (study of translation theory or evaluation of a translation product). Our observation of parallel corpora revealed the following information about The International Corpus of English as a corpus spanning several parallel corpora: "The International Corpus of English (ICE) is a collaborative project of several dozen universities, each using millions of words. It consists of 20 parallel inner casings covering the state. It is possible to study the peculiarity of the countries where English is the state language or the second language (Austria, Canada, New Zealand, etc.). Excellent software was specially developed for the analysis of this corpus.

Parsing unrestricted text is useful for many language technology applications but requires parsing methods that are both robust and efficient. MaltParser is a language-independent system for data-driven dependency parsing that can be used to induce a parser for a new language from a treebank sample in a simple yet flexible manner. Experimental evaluation confirms that MaltParser can achieve robust, efficient and accurate parsing for a wide range of languages without language-specific enhancements and with rather limited amounts of training data [3].

Two main directions can be distinguished in this process. They include the development of non-pharmaceutical information retrieval languages and the creation of pharmaceutical information retrieval languages that incorporate pharmaceutical terms, among other words. This means the means of unifying the description, classification and search of documents, which include types of classification of terms,

abbreviations, signs of terms, etc.

In the monograph "Pharmaceutical information issues" by R. M. Rinyajko and B. L. Rarnovsky, the existence of the field of pharmacy in information search languages such as Universal Ten Classification, Library-bibliographic classification - library-bibliographic classification was considered. The authors note that there are no special (network) tables for pharmacy in the universal ten classification, besides, the interpretation of pharmacy as a branch of pharmacology is a serious error in the classification structure of the universal ten classification. In the library-bibliography classification, pharmaceutical terminology is more widespread than the universal ten classification, as well as in different departments. This makes it more difficult to use search languages for indexing pharmaceutical publications[4].

Pharmaceutical terminology is also included in the linguistic support of automated information systems in fields such as chemistry and chemical technologies, biological sciences, and medical-biological production.

Thus, in non-pharmaceutical information-search languages containing pharmaceutical terminology, it is possible to observe classification relationships that do not correspond to those accepted in the field of pharmacy.

A number of works in the field of publications devoted to pharmaceutical information retrieval languages are not only focused on automated information systems, but are intended for a wider field of application.

In the development of automated information systems about drugs, much attention is paid to the issue of linguistic support. In drug information, there are many areas of terminology (pharmaceutical, chemical, pharmacological, medical and other terms), a large number of controlled vocabularies - drug substances, chemical compounds, drug forms and names, pharmaceutical analysis electronic classifiers, rubricators, reflecting the terminology and nomenclature of methods, methods of administration, instructions for use and contraindications, pharmacological effects, dosages, forms of release, packaging, packages, etc., also requires the creation of thesauri[5].

The analysis of the works described above shows that they mainly cover general pharmaceutical terminology and the terminology and nomenclature of drugs.

The results of the study of official documents and scientific literature in the field allow drawing the following conclusions.

In the 90s of the last century, significant qualitative changes in the pharmaceutical sector of our republic affected the pharmaceutical lexicon. This was manifested in the emergence of a large number of new terms and phrases that are not adapted to the field of pharmacy in our country, in the presence of terms from foreign languages (mainly anglicisms) even when there is an Uzbek alternative to the terms that are common in the Russian language. On the positive side, along with the expansion of the vocabulary of the pharmaceutical lexicon, such situations have caused serious problems related to professional terminology.

Despite the fact that in the 70s and 80s of the 20th century, the leading scientists of our country carried out scientific work on pharmaceutical terminology, later this direction of research was practically stopped. Scientific articles on the interpretation and analysis of this or that term are rarely published in journals. Due to the lack of terminological studies aimed at a specific goal, carried out in accordance with separate, scientifically based rules, information chaos has arisen, and it has become difficult to develop legislation and departmental (normative) documents.

Of course, it is not enough to study the scientific publications related to terminology, to make a professional evaluation of the essence of the concepts expressed by terms only in the standardization of pharmaceutical terminology. In this case, it is absolutely necessary to use the systematic approach and methodology of terminological research, which should form the scientific basis and foundation of standardization, regulation, standardization.

This process can become more complicated if terminological units are used in communication without regulating the existing national terminology and without critical analysis of international standards.

Thus, terminological research in the field of pharmacy should be a priority direction of pharmaceutical sciences at the present time and should be conducted on a regular basis.

**Result and Discussion:**The analysis of the literature and other materials related to pharmaceutical terminology showed that pharmacy, as a field of scientific and practical knowledge, should create and develop its own terminological phrases in combination with appropriate language units. They are studied and

analyzed in detail by specialists in pharmaceutical and philological fields.

However, despite the achievements in the research of pharmaceutical terminology at a certain level, which appeared primarily due to the efforts of philologists working in pharmaceutical and medical higher education organizations, the disparity in the development of sciences related to the field of pharmaceuticals remains. It can be explained that such a situation has arisen primarily due to insufficiently developed theoretical and methodological foundations of terminological activity in the field of pharmacy.

In accordance with the main goal of the scientific work and on the basis of a systematic approach, a research program including three stages, intermediate goals and planned results was formed.

They are:

- regular analysis of the development of terminological activities in the field of pharmacy;
- development and linguistic research of the scientific-theoretical bases of pharmaceutical terminology;
- development and analysis of the scientific-methodical bases of terminological activities in the field of pharmacy.

The purpose of the first step will be to justify and select the first directions of this research. Therefore, there are many unsolved terminological problems in the field of pharmacy, and it is impossible to solve all of them within the framework of one study. The result of this process is the development of a research program.

The second stage of the scientific work is explained by the fact that it is aimed at studying the necessity and possibility of the formation of pharmaceutical terminology, a new scientific and practical science direction. For this, it is necessary to pay attention to the following steps of research. For example:

- as a result of the study of external and internal factors of the lexicon of the field and the constant analysis of the existing problems of pharmaceutical terminology, the foundation and formation of the pharmaceutical terminology concernation;
- it is desirable to analyze and research the main and basic terms and phrases of the pharmaceutical industry, as well as to describe them, to analyze the names of drugs and their constituent phrases. In this sense, it is appropriate to form separate terminological systems “Drugs” and “Pharmacy” based on the received information. As a result of the second step, it is planned to form a terminological system and terms related to pharmacy as a single scientific-theoretical base required for further terminological research[5].

The main goal of the third step is to develop and research the scientific and methodological basis of electronic terminological activity in the field of pharmacy. This goal is achieved by developing the forms of terminological support of experts in the field of pharmacy, as well as by developing criteria and technologies for regulating pharmaceutical terminology and nomenclature. Carrying out this research, first of all, from the functional-semantic point of view, it is possible to study the methodical-methodological principles, criteria and specific lexical-semantic features of working with pharmaceutical terms. Therefore, methods of researching the etymological and word-forming, i.e., derivational aspects of the professional lexicon are traditionally developed by philologists. As a result of the implementation of the third step, it is envisaged to create a system of electronic official documents that includes pharmaceutical terminological units and nomenclature regulated in the selected field of pharmaceutical knowledge. The general systematicity of the research is provided by the consistent use of a systematic approach to the implementation of the second and third steps of the work.

The study of scientific publications on terminology, the professional assessment of the content and essence of terms expressed only in terms of standardization and unification of pharmaceutical terminology will not be enough. In this case, it is necessary to use the systematic approach and methodology of terminological research, which should form the scientific basis and foundation of standardization, regulation, and standardization.

On the basis of such a general purpose and a systematic approach, it is extremely important to form research programs that take into account the following goals and planned results:

- regular analysis of the development of terminological activities in the field of pharmacy;
- developing and linguistically researching the scientific-theoretical bases of pharmaceutical terminology;
- development and analysis of the scientific-methodical bases of terminological activities in the field of pharmacy.

At the end of the research, the creation of the scientific-theoretical and methodological foundations of

pharmaceutical terminology as a direction of the pharmaceutical industry ensures the achievement of the intended main goal[6].

**Conclusions:** The most important thing when creating dictionaries of electronic industry terms is to consider internal structure that is more or less common in printed dictionaries. Morphological structure is very important in creating electronic dictionaries, including electronic corpora[7,8]. Subordinating the large volume of the entire language to the national corpus requires a complex process of determining the general laws of the language, in particular, morphological forms, and bringing them into the form of integrated models. As a result, full-text search can distinguish, for example, translations, usage examples, and reviews, which greatly improves its capabilities from a user's perspective.

In short, a corpus-based approach to language education is the need of the hour. The corpus of terms that we offer is a continuation of corpus linguistics research carried out in our country as a small study on creating an educational corpus.

#### References:

1. Менглиев Б. Ўзбек тилининг ҳаммиллий корпусиярати ишилозим. <https://kun.uz/>
2. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: филол. фанлари бўйича фалсафа доктори диссертацияси. – Бухоро, 2018. – Б. 29, 15. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: филол. фанлари бўйича фалсафа доктори диссертацияси. – Бухоро, 2018. – Б. 39, 52.
3. Joakim Nivre, Johan Hall, Jens Nilsson, Atanas Chanev, Gülşen Eryigit, Sandra Kübler, Svetoslav Marinov, Erwin Marsi. “MaltParser: A language-independent system for data-driven dependency parsing”. Natural Language Engineering 95-135 P, 2007/6, Cambridge University Press
4. Пиняжко Р.М., Парновский Б.Л. Вопросы фармацевтической информации. – М.: Медицина, 1979. – 152 с
5. Шобабаева А.Р. Разработка основ лингвистического обеспечения автоматизированной информационно-поисковой системы (АИПС) “Лекарство”. Дис. ... канд. фармац. наук. – М., 1987. – 147 с.
6. D.K. Khudoyqulova., O.S.Akhmedov “Diachronic study of pharmaceutical terms” Journal of Critical reviews. Volume 7, Issue 2, may. 2020. P 117-119
7. A. A.Rakhmanova, M.T.Abdurakhmanova, Z. T.Xolmanova “Representation of linguistic issues in corpus” Journal of Critical reviews. Volume 7, Issue 2, may. 2020. P 121-127
8. Course “Corpus Linguistics” (A.B. Kutuzov) License Creative commons Attribution Share-Alike 3.0 Unported (Electronic resource) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf
9. Baranov A.N. Introduction to applied linguistics. - М.: Editorial URSS, 2001.-S.61.