PRINCIPLES OF CREATING AN ELECTRONIC DICTIONARY OF GRAMMATICAL TERMS

Laylo Rakhimovna Raupova*

*Professor,
Doctor of Philology,
Alisher Navo'i Tashkent State University of Uzbek Language and Literature,
Tashkent, UZBEKISTAN
Email id: raupovalaylo@navoiy-uni.uz

ABSTRACT

The development of science and technology, the activation of scientific communication, and the sharp increase in the volume of scientific and technical information have focused the attention of scientists on conducting theoretical, practical and analytical researches on the language of scientific communication, and as a first step in this, creating field terminological dictionaries, improving and automating existing dictionaries. The growing role of science in the world civilization, the increase in the amount of collected, processed and disseminated knowledge showed the necessity of conducting new researches in the field of terminology and terminography. It is known that the creation, systematization, automation of the base of field terms, their reflection in a compatible corpus, which is considered one of the core issues of modern and applied linguistics, and the determination of the position of the terms of their own layer in it, will serve as a source for further research in this direction.

KEYWORDS: *Terminography, Terminological Dictionaries, Electronic Dictionaries, Grammatical Terms, Transcription, Algorithm.*

INTRODUCTION

Despite the centuries-old traditions of creating a dictionary in world lexicography, general theories of terminology organization and description automation show that they are not sufficiently developed. At the same time, there is a lot of work to be done in the field of terminology. Working on terminology includes several stages and conditions. One of them is the inventory of terminology, that is, the collection and description of terms related to a specific field of knowledge or one of its sections. At the same time, all available information about the term - its etymology, frequency of use, various linguistic features will be highlighted.

As you can see, inventory is often the first step in organizing terminology and defining its inter-systemic connections. A large number of specialists are involved in the compilation of terminological dictionaries and necessary and strict criteria are determined for the dictionary. However, in a glossary of field terms, the criteria can be adjusted, taking into account the nature of the proposed terms. One of the important processes in the creation of terminological dictionaries is the editing of terms of terminological activity, that is, the evaluation of the correct

use and compatibility of terms and the replacement of incompatible terms with recommended and standardized forms.

Before explaining the specific features of the borrowed grammatical terms in Uzbek and English, let's briefly touch on the term grammar. It is known that the grammar of the language is a branch of science that reveals how the language is structured, how the words in the language are connected to each other on a constructive basis, the limits of their possibilities of connection in the construction of a sentence, and other specific features.

Grammatical terms as a lexical unit of the language can be formed through internal possibilities or acquired with or without various phonetic phenomena. At this point, it is appropriate to dwell on the important aspects of the phenomenon of word acquisition from one language to another.

Term assimilation is the process of acquiring elements of a grammatical term from one language into another language. As a result of certain linguistic and extralinguistic factors, the acquired unit is directly or indirectly adopted and used in the second language. In the language, as a rule, the assimilation of terms is actively observed in the vocabulary, phonology and grammar layers of the language. The following methods and methods are used in the acquisition of terms.

The issue of terms and their research has attracted the attention of researchers for many years. In scientific sources: one can find notes that terms "do not appear", but are "invented" and "created" [1, p. 24] when they are needed. This opinion of the linguist can be supported, because with the development of science and technology, the need for terms is felt in society. These changes in development contribute to the development of many areas of linguistics, such as onomaseology, terminology, lexicology, lexicography. In other sources: "neither a specific science nor technologies in general, therefore, their terms are not the creation of a specific people. It is emphasized that they are a set of objective existence collected, known, and studied by all the peoples of the world" [2, p. 4]. Specialists in the field try to emphasize that the differences between the concepts of "term", "atama", "istilah" should be taken into account. For example: "It is appropriate to use the word term in a broad sense in relation to words and phrases used in a field or branch of science, and the word term in relation to conditionally given names, including names of geographical objects, toponyms; it should be noted that it is inappropriate to use the word "istilah" in the sense of the term as the name of modern concepts of science and technology, but there is no objection to using this word in texts on historical topics [3, p. 29].

Transcription is an artificial (transcriptio) type of writing used to accurately express the pronunciation of speech sounds; the method of accurately reflecting the sound qualities of speech fragments (sound, syllable, word) in writing also implies transcription. Transcription is divided into two: scientific transcription and practical transcription. Scientific transcription is divided into phonetic and phonemic transcriptions according to the type of meaning. Phonetic transcription provides an accurate representation of spoken speech with all its features in writing and is used in a dictionary of a foreign language (for example, English), in linguistics textbooks and speech guides, in recording a living language (dialectological notes, phonetics textbooks, etc.). The main principle of phonetic transcription is that each pronounced sound should be recorded separately in the record. Phonetic transcription is usually created on the basis of an existing alphabet, by adding to its superimposed symbols or an arbitrary invented system of symbols that represent certain articulations of the speech organs. Phonemic transcription represents each word according to its phoneme content. According to the principle of this

transcription, each phoneme, regardless of its position, is always represented by the same symbol. Phonemic transcription uses fewer symbols than phonetic transcription, because the number of phonemes is always less than the number of variants. Practical transcription reflects nouns, terms, special names and similar untranslatable words from other languages through the means of specific national alphabets and thus incorporates them into the printed texts of the receiving language. Practical transcription does not go beyond the character limits of the recipient language alphabet, but allows for some unusual uses of characters. One of the main requirements for practical transcription is to preserve the pronunciation of the received unit as accurately as possible. At the same time, practical transcription should preserve the morpheme structure of the word, its graphic features, etc., and ensure the easy assimilation of a particular word [4].

Above, we mentioned the idea that the Greek language was taken as the basis for the formation of English grammar. Most of the grammatical terms of the Uzbek language are taken from the English or Russian language through an intermediary language, and below we can give an example of the terms *Lexicology, Lexeme, Syntax*, which were acquired using the transcription method.

Transliteration (**Trans** and **Latin. litera** - letter) is a letter-by-letter representation of texts written on the basis of one graphic system (alphabet) or individual words with the means of another graphic system (alphabet) [5]. This method is also actively used in reflecting grammatical terms. For example: predicate, transitive, etc.

Translation is the re-creation of a text in one language in another language, unlike transcription and transliteration; it is the most actively used method in the acquisition of grammatical terms. It is based on the semantic field of the word. Most of the terms originally derived from Latin are translated from one language to another. For example: we can see this in the example of the following Uzbek, Russian and English terms.

TABLE 1 SOME GRAMMATICAL TERMS IN UZBEK, RUSSIAN AND ENGLISH

Uzbek	Russian	English
So'z	Слово	Word
Gap	Предложение	Sentence
Ot	Имясуществительное	Noun
Olmosh	Местоимение	Pronoun
Sifat	Прилагательное	Adjective
Fe'l	Глагол	Verb

At a time when fields and science directions such as computer linguistics, corpus linguistics, language digitization and automation, natural language processing are developing in world linguistics, creating the necessary conditions for the development of the Uzbek language, creating national, educational, authorial and parallel corpora of the Uzbek language, enriching and requires improvement. The electronic dictionary of grammatical terms of the Uzbek language also serves as one of the important basic elements of the educational corpus of the Uzbek language. In this regard, we will focus on the important conditions and principles for creating electronic dictionaries as an example of creating an electronic dictionary of Uzbek grammar terms.

Until recently, electronic lexicography was considered one of the least developed areas of linguistics in terms of sophistication. The main reason for this is that it took a lot of time and labor to collect dictionary articles, collect index cards, and extract the necessary linguistic materials from written sources to create traditional paper dictionaries. Unlike traditional lexicography, modern lexicography is becoming one of the most developed and promising fields in terms of creating active electronic dictionaries that allow collecting, storing and processing the necessary information in a reliable, compact volume. It is known that traditional dictionaries cannot continuously and consistently reflect the changes occurring in the living language, in particular, new words, terms, obsolete terms, and the scope of use or expansion of a particular term. But electronic dictionaries allow you to enter new information on a daily basis and improve the database of electronic dictionaries. Also, electronic dictionaries provide great convenience to the user as they have a fast and efficient search system. To add a new element to the base, each word is compared with the words in the base. If the word is in the database, it adds its frequency to the frequency of the word in the database; otherwise, if the new word is not found in the database, it is added from the end of the file with this frequency of the word. Its algorithm and program are as follows:

TABLE 2 ALGORITHM FREQUENCY OF WORD PROGRAM
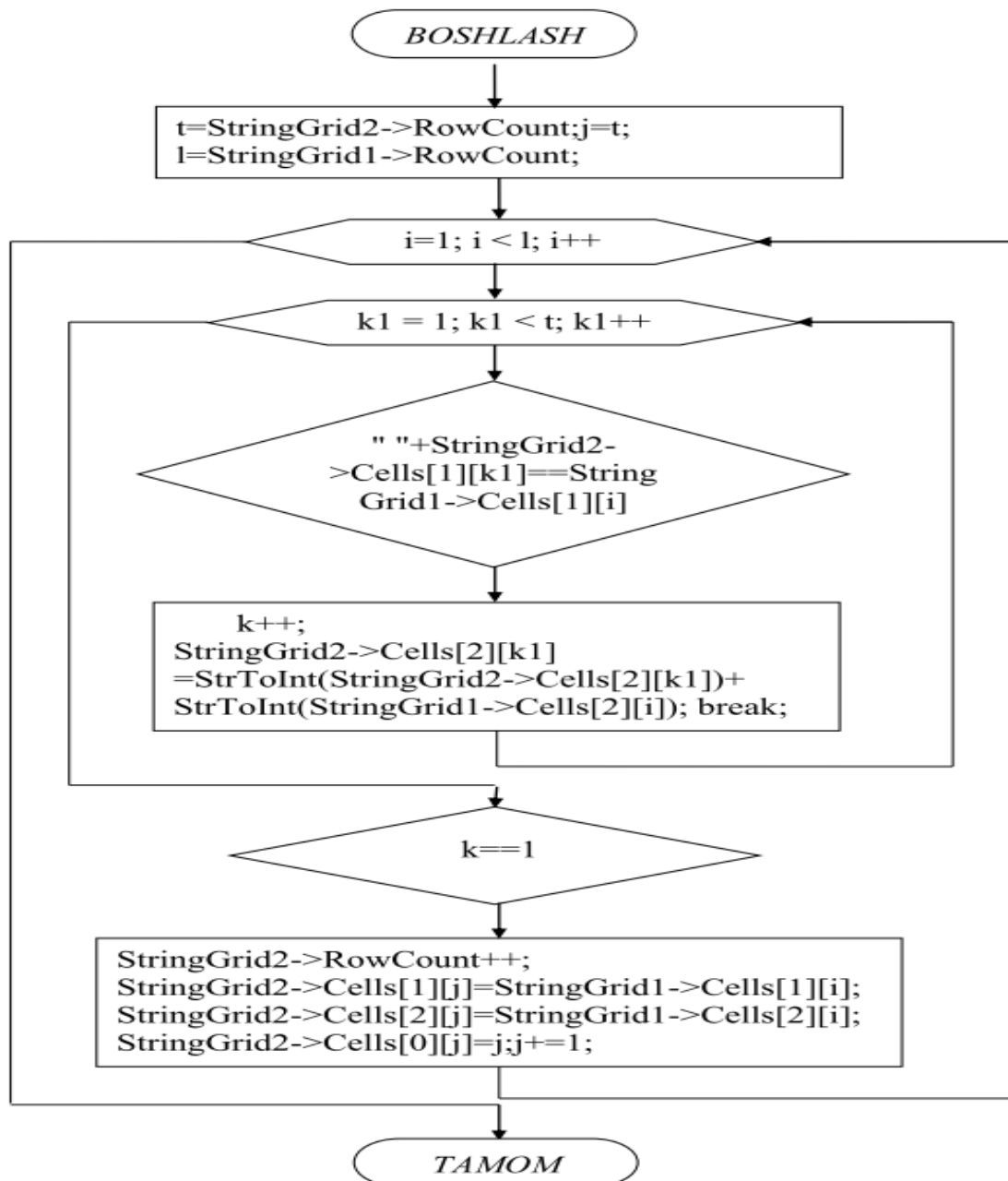
An electronic dictionary is an analogue of a paper dictionary equipped with a convenient search system. An electronic dictionary is defined as an electronic information source that is compatible with a traditional paper dictionary. In the electronic version, the information can be called from any specially defined program by pointing to a word or a group of words, which leads to the visualization of the necessary fragment of the corresponding dictionary. Unlike traditional dictionaries, an electronic dictionary can contain video and animation fragments, sound, music, etc. along with text and graphic images can be implemented on the basis of hypertext or hypermedia.

Z. Komarova says that it is necessary to pay attention to the following parameters when creating a terminological dictionary [6, pp. 43-44]: 1) criteria for choosing a dictionary; 2) the issue of using illustrative materials; 3) justification of the pometta system; 4) improvement of basic semanticization (explanation) tools; 5) structure of the dictionary article.

In another case, V. Selegei considers the electronic dictionary to be a lexicographic object that can implement many productive ideas that are not required in paper dictionaries for various reasons. The researcher calls such antinomies of traditional paper dictionaries:

1) The larger the size of the dictionary, the more complete the description of lexical meanings, the more difficult it is to use;

2) The more complete and deep the description of lexical meanings, the less appropriate the dictionary is to the current linguistic and cultural situation;

3) The more interesting the lexicographic concept of the dictionary, the narrower its lexical base.

Experiences from the process of data automation show that the conversion of conventional traditional dictionaries into electronic form is not enough to create an active lexicographic system. It is necessary to develop special technologies that allow optimization in the Uzbek language, to use the opportunities provided by the computer in relation to the Uzbek language. For example, hypertext technology is very useful for improving the quality of the electronic dictionary, speeding up the work of users, and optimizing the search for the necessary information. As you know, there are many types of dictionaries and they differ from each other. These differences are determined by how complex and multifaceted the lexicographical definition of an element is as a specific vocabulary unit. Creating electronic dictionaries is a very complex process. In addition to linguistic rules about lemma, its meanings and usage, grammatical and phonetic features, knowledge of vocabulary building techniques is also an important part of this process. Expressing the semantic meaning of the words in the dictionary, presenting a specific dictionary in a form divided into meaningful parts, giving stylistic, grammatical and phonetic comments or tags to lexemes and their semantics, attaching idiomatic and phrase logical combinations corresponding to the given lexeme are among the important principles encountered in creating an electronic dictionary. It focuses on translation in multilingual dictionaries and interpretation in monolingual dictionaries. In the process of creating language-related software products, electronic dictionaries are often used, and this situation requires frequent changes in the semantic-structural state of the electronic dictionary.

To create an electronic dictionary of Uzbek grammar terms, it is recommended to perform the following basic steps.

1. Forming a database of grammatical terms adapted to the grammar of the Uzbek language.

2. To determine the method by which the formed base of grammatical terms was mastered and to determine the final recommended grammatical form of the new term through it.

3. Determining the position of the acquired grammatical term according to its form and meaning.

4. Creating an explanation or translation of the acquired grammatical term taking into account etymology, semantics and phonology;

5. Passing through stages of sorting, lemmatization, parsing, and tagging of grammatical terms that have been mastered in an appropriate manner, which is intended for the purpose of creating an electronic dictionary.

6. Determining the appropriate valency of the acquired grammatical terms.

7. To ensure that the acquired grammatical terms are reflected in the electronic dictionary in a reliable, concise and suitable form for the search system of the electronic dictionary.

8. Creation of the intended software of the database of grammatical terms and its reflection in the corresponding corpus.

The development of theoretical proposals for the organization of language units is an important part of modern linguistic research, because "linguistic theory is necessary not only to describe facts, but also to determine how to determine them." [7, pp. 247-253] Concepts of terms and terminology are of great interest, especially due to recent advances in computing, textual terminology, and the rapid development of its applications. This study aims to review them and formulate relevant concepts relevant to today's problems. By synthesizing and generalizing the existing terms, an attempt is made to form an appropriate comprehensive definition. It can be concluded that there are enough grounds to look at terminology not only as a part or doctrine of vocabulary and lexicology, but also as a separate independent science related to the formation and use of terms [17]. An important task is to discuss the research methods of terminology as a science, general scientific methods, and issues related to the natural sciences and traditional linguistics.

The process of creating the concept of a term is long and varied. On the one hand, this is due to the fact that the concept is quite complicated. On the other hand, defining the concept of "concept" itself is difficult. Among the different approaches to this concept, we consider it from the international standard ISO 1087-1: 2000, 3.2.4, as a unit of knowledge created as a result of a specific combination of properties.

The safety of languages is ensured by the fact that they have an alphabet, a large number of speakers, and the status of a state language, as well as the fact that it is the language of the information and communication spheres and the Internet [8, p. 96]. All languages are classified in terms of survival rate as follows:

1. Supported languages: *English, Chinese, Spanish, German, Japanese, French, Arabic*.

2. Languages in need of caution: *Belorussian, Yakut, Chechen, Tuva languages*.

3. Languages approaching the danger of death: *Chukot, Karelian, Kalmyk, Udmurd*languages.

4. Endangered languages*: Aleut, Ters-Saam, Itelmen* languages.

5. Dying languages: *Ain language, Yuka (Gir) language, Ubih language*.

Therefore, raising the Uzbek language to the level of "artificial intelligence language" becomes urgent, and this process requires the creation of national and educational corpora. Therefore, it is important to have a proper understanding of the corpus concept. In this regard, it is required to design educational corpora and determine what work needs to be done in practice [15]. First of all, it is necessary to determine for whom the teaching of the Uzbek language is intended. For example, if the Uzbek language is taught as a mother tongue, its graded criteria must be decided

in advance, that is, how many words the learners should have and be able to use it in communication. In the second aspect, the age of learners and in which educational institutions they receive their education, as well as their gender and general interests or professional ability, should also be taken into account as a specific category. After the standards have been developed, appropriate texts are selected.

In order to create educational corpora, the texts should be brought into electronic form: PDF-> WORD-> TXT [9, pp. 50-58]. Also, in learning Uzbek as a foreign language or a second language, in addition to the criteria mentioned above, the nationality or country one belongs to, how long the language has been studied, and the general goals of the learner are also taken into account. If a native speaker considers Uzbek language as the key to all knowledge and the main means of communication from childhood, he accepts having scientific knowledge as a competence in learning a foreign language in addition to some of the factors mentioned above. Regarding the technological process of creating a corpus, V.V. Rikov lists the main requirements: "Corpus user (individual, group, linguistic community); the logical purpose of the corpus; the size of the database used in building the corpus and its real, necessary level; method of using the text (fragment, whole, or both)" [10]. Sh.Khamroeva makes the following proposal regarding the corpus of authorship: "Research of language material; scan text; complete formation; creating a corpus" [11, p. 456] S.O. Savchuk, a scientist who worked on the texts of the National Corpus of the Russian Language, divides the technological stages of corpus creation into the following process: "Re-introduction of texts; use of texts available in electronic form; scanning the printed text (but this will require a lot of debugging)" [13, pp. 149–154]. V.P. Zakharov and S.Y. Bogdanova cite the issue of chronology as an important aspect of the corpus design process. For example, what should be understood by the modern corpus of the language? It is natural that the chronological limit of the corpus is different in different genres. The corpus is designed for wide public use and for performing a variety of tasks (including studying texts written in Russian on the basis of other graphics). In the corpus, what parts of the original form of the text are taken and what is left out is another important issue. Since the pictures contained in the text do not belong to the language material, it is important to exclude them from the text included in the corpus, and to process the tables according to the corpus [16]. They are important in representing the content of the text, but if left in the corpus, they are difficult to tag. Quotations, quotations, self-identified units (terms), units of measurement also require special attention. If the listed issues are solved on the basis of a certain principle at the design stage, some of them are solved in the process of creating a corpus, using the corpus. At the same time, feedback with the user should be considered before launching the case. Specialists K.F.Meer and I.A.Melchuk show the following steps in the technological process of creating a case [14, pp. 12–36]:

1. Ensure that the text is included in the corpus according to the specified source.

2. Text processing in the form of automatic reading. The electronic text to be included in the corpus can be obtained by various means: scanned, manually entered, exchange, author's copy, gift, Internet, original layouts provided by publishers to the corpus compiler, etc.

3. Analysis, preliminary text processing. At this stage, the texts received from various sources go through the stages of philological verification and editing.

4. Conversion, graphematic analysis. Some texts repeatedly go through the first machine processing, where the recoding process takes place, non-textual parts (pictures, tables) are deleted or changed. Hyphenation in the text, borders (in MS-DOC texts) are canceled, hyphens, other characters are made identical. Graphematic analysis is the process of dividing the corpus text into parts (words, conjunctions) and removing non-textual elements.

5. Determining a non-standard element, formalizing it, revising a special textual element (abbreviated name (first name, last name), custom lexeme written in a different alphabet, name given to a picture, annotation, quotation, bibliography, etc.) based on the same criteria. Obviously, these actions are automatically performed by the text editor program.

CONCLUSION

In conclusion, it should be said that by creating an electronic dictionary of grammatical terms for the acquisition of the Uzbek language and reflecting it in a suitable corpus, it is possible to help the scientific research conducted in various fields of Uzbek linguistics to a new level, to achieve improvement in terms of quality and time, students of the Uzbek language, practical projects in the field of linguistics. Implementation, it will be possible to provide wide opportunities and facilities for local and foreign specialists.

REFERENCES

1. VinokurG.O. On some phenomena of word formation in Russian technical terminology. // Proceedings of MIFLI. Volume 5. Collection of articles on linguistics. – Moscow. 1939. – p. 24.

2. Doniyorov R. From the history of scientific and technical terms of the Uzbek language. – Tashkent: Science, 1973. – p. 4.

3. MadvalievA. Issues of Uzbek terminology and lexicography. – Tashkent: National Encyclopedia of Uzbekistan, 2017. – p. 29.

4. https://uz.wikipedia.org/wiki/Transkripsiya

5. https://uz.wikipedia.org/wiki/Transliteratsiya

6. Komarova Z.I. Methodology and technique of terminographicsemantization // Vocabulum et vocabularium. – Harkov. 1995. Issue. 2. – pp. 43-44.

7. Khursanov, N. I. (2022). Types of Semantic Transference in Dramatic Discourse (on the Example of Uzbek and English Texts). ACADEMICIA: An International Multidisciplinary Research Journal. Year: 2022, Volume: 12, Issue: 4. First page :(152). Last page :(158). *Online ISSN*, 2249-7137.

8. Mengliyev B., Tashmirzayeva Sh., Atoyeva S., Majidova S., Mannopova D. Mother tongue / Textbook for the 10th grade of general secondary schools, part 2. – Tashkent: "Teacher" NMIU, 2020. – p. 96.

9. Abdurahmanova N.Z. Computer models of modern corpora // Foreign languages in Uzbekistan. – Tashkent. 2020. – No. 1(30). – pp. 50-58.

10. Rykov V.V. Course lecture on corpus linguistics. – URL: http://rykov-cl.narod.ru/c.html.

11. Khamroyeva Sh. Linguistic foundations of the composition of the author's corpus of the Uzbek language: Abstract of Dissertation PhD. –Karshi. 2018. –p. 456.

12. Kursanov, N. (2022). The Social Characterization of the Speech of the Protagonists of Dramatic Works. *Computer Linguistics: Problems, Solutions, Prospects*, *1*(1).

13. Savchuk S.O. Corpus of modern Russian press: from the experience of creation and use // Proceedings of the International Conference "Corpus Linguistics – 2011". – St. Petersburg: St. Petersburg State University, 2011. – pp. 149–154.

14. Melchuk I.A. Word order in automatic synthesis of the Russian word (preliminary messages) / Scientific and technical information, 1985. - No. 12. – pp. 12–36; Meyer C.F. English Corpus Linguistics. – Cambridge: Cambridge University Press, 2002. – URL: https://pdfs.semanticscholar.org/c775/4bfab1d0f7

15. Raupova, L., Botirova, A., Musulmanova, N., Kadyrova, H., & Safarova, G. (2020). Logical and Grammatical Relations in Word Categories: The Factor of Difference and Incarnation. *InternationalJournalofPsychosocialRehabilitation*, *24*(04).

16. Khursanov, N.I. (2022) Linguopragmatic Characteristics of Stylistic Units Used in Uzbek and English Dramas. Science Department of the National Information Agency of Uzbekistan (Electronic Journal). – Tashkent. № 4. Issue 1. – pp. 149-155.

17. Raupova, R. L. R., & Elov, B. B. (2020). The Experience of Bologna in Reforming Educational System: The Forming of a Credit-Module System in the Educational Process. *Solid State Technology*, *63*(2).