



CRITERIA FOR CREATING AN ELECTRONIC DICTIONARY OF GRAMMATICAL TERMS

Shakhnoza Sultanova

Senior teacher, Doctor of philosophy (PhD)
of Tashkent State University of

Uzbek Language and Literature Named after Alisher Navoi

<https://doi.org/10.5281/zenodo.8115591>

Abstract:

In this article, the researcher reflects on the issues related to the criteria for creating an electronic dictionary of grammatical terms of his class. The basis of the article is the clarification of the opinions of Uzbek and foreign scientists in the corpus field, the description of the technological process of the corpus design, the main differences between dictionaries and traditional dictionaries, the ways of forming the base of the electronic dictionary of Uzbek and English grammatical terms are given in detail. In this case, the dictionary database consists of the term in Uzbek language, its explanation, the corresponding grammatical/linguistic tag, the synonym of the term, its symbol, its equivalent in English, the interpretation of the term in English.

Key words: grammatical terms, corpus, own layer, terminology, terminography, educational corpus, electronic dictionary.

The development of theoretical proposals for the organization of language units is an important part of modern linguistic research. Terminology and terminology concepts are of great interest, especially due to recent advances in computing and 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 sufficient grounds to look at terminology not only as a part of vocabulary and lexicology or as a doctrine, but also as a separate independent science of the formation and use of terms. 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. Raising the Uzbek language to the level of "artificial intelligence language" has gained relevance, and this process requires the creation of national and educational corpora. In this regard, N. Abdurakhmonova puts forward the following opinions: "Different computer models of the corpus can be created, its integrity depends on the set goals and tasks, agreed principles, and required audience conditions. Therefore, it is important to understand the concept of corpus. In this regard, it is required to design educational corpora and determine what work needs to be done in practice. 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 must have and be able to use it in communication. In the second aspect, the age and educational institutions of the learners, as well as their gender and general interests or professional ability should also be taken into account as a specific category. After the standards are developed, appropriate texts are

selected. To create educational corpora, it is necessary to bring the texts into electronic form: PDF-> WORD-> TXT [Abdurahmonova N, 2020, 50–58]. Also, when learning Uzbek as a foreign language or a second language, in addition to the criteria mentioned above, to which nation or country one belongs and how long the language has been learned using what methods and the general goals of the learner are also taken into account. If a native speaker views Uzbek as the key to all knowledge and the main means of communication from childhood, he will consider acquiring scientific knowledge as a competence in learning a foreign language in addition to some of the factors mentioned above. Sh.Khamroeva makes the following proposal regarding the corpus of authorship: "Research of language material; scan text; complete formation; creating a corpus" [Khamroyeva Sh, 2018, 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 of printed text (but it will be necessary to correct many errors)" [Savchuk S., 2011, 456] . 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 various tasks (including studying texts written in Russian on the basis of other graphics). Another important issue in the corpus is what parts are taken from the original form of the text and what is left out. 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. They are important in expressing the content of the text, but if left in the corpus, they are difficult to tag. Quotations, personal 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 also be considered before launching the case. Specialists K.F.Meer and I.A.Melchuk indicate the following stages in the technological process of building a corpus [Melchuk I, 1985, 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. Text in electronic form 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, texts received from various sources undergo 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. Designation of a non-standard element, formalization, special textual element (abbreviated name (first name, last name), custom lexeme written in a different alphabet, name given to a picture, caption, citation, list of references, etc.) the same criterion revision based on Obviously, these actions are automatically performed by the text editor program.



When sorting the corpus material, what is the main unit of the corpus, what is its size (how many words should it contain), what source is the written text based on, how much is it, what area of the language does the text in it belong to? A solution to such issues will be found. The first answer to this question was given by R. G. Piatrovsky and his students in the period of 1965-80. They developed principles of text selection for frequency dictionary and linguostatistical research. This problem was also raised in the frequency dictionary under the editorship of L.N. Zasorina. For the first time, factors such as the statistical method of text selection, volume, and quantity are listed in it. The main units of the corpus are word form, root (base, lemma) and sentence. The size of the case to be built is determined based on the purpose of the case. If it aims to study letters, letter combinations, sounds, diphthongs, it does not have to be so big. If the lexical units of the text are compiled in order to study the morphological phenomenon, syntactic, and methodological characteristics, it requires a large volume [Hamroyeva Sh., 2018, 76] . According to S.A. Sharov, during the selection process, choosing a text of any genre (prose, drama, poetry, scientific text, newspaper, magazine material, etc.), which period the text covers (contemporary, 10-year, 50-year-old, and classic text), the text matters such as whether it is only in literary language or whether other sources are included are also important. In this process, the corpus compiler will certainly turn to a linguist, a linguostatistics specialist or the questionnaire method. In the process of creating a corpus, the author, relying on his experience, considers the main factor of the total size of the corpus, the time of publication of the text, the number of texts, the size of the selection of elements, the type of genre to be chosen [Hamroyeva Sh., 2018, 79] . The questionnaire method was used by the creators of "The American Heritage Intermediate Corpus". 5 million words and 22 types of texts for children and teenagers are included in English. A survey was sent to 221 schools in the United States, which determined the



Termin	Izohi	Tegi	Sino nimi	Ram zi	Ingliz tili	Inglizcha izohi
Undov gap	His-hayajon (shodlik, qo‘rquv, g‘azab, ajablanish, achinish, do‘q, yalinish, hayrat va h.k.) bilan aytilgan gap	nima? Ot turkumi, turdosh ot, qo‘shma ot, yasama ot, birlik son, bosh kelishik		N+N =N	Exclamation sentence	An exclamatory sentence, also known as an exclamation sentence or an exclamative clause, is a statement that expresses strong emotion.
Matn	-Mazmun va grammatik jihatdan bog‘langan, bir mavzuni ifodalovchi gaplar yig‘indisi (8-sinf) -O‘zaro va tarkiban bog‘langan gaplar ketma-ketligidan iborat yozma yoki og‘zaki shakldagi yaxlit birlik (6-sinf)	Nima ? Ot turkumi, turdosh ot, sodda ot, tub ot, birlik son, bosh kelishik		N (noun)	Text	Text is usually a written form of communication information, which is a non-interactive nature.
Mavzu	Matndagi asosiy fikrning nima haqida ekanligi. Shundan kelib chiqqan holda matnning sarlavhasi ham bo‘lishi mumkin	nima? Ot turkumi, turdoshot, soddaot, tub ot, birlik son, bosh kelishik		N (noun)	Theme , topic	In linguistics, the topic, or theme, of a sentence is what is being talked about, and the comment (rheme or focus) is



						what is being said about the topic.
--	--	--	--	--	--	-------------------------------------

appropriateness of the choice of text. After studying the results of the survey, a list of 19,000 books was compiled. Based on this, 1045 texts with 500 word forms each were selected. It should be noted that in the process of designing the corpus, the selection of material (text), sorting, and its technical adaptation to the corpus is the most important stage [Hamroyeva Sh., 2018, 82].

Below, we will talk about the ways of forming the database of the electronic dictionary of grammatical terms of Uzbek and English languages, as well as the analyzes carried out. In the appendix of the work, there are 34 linguistic terms in Uzbek, their explanation, equivalent in English, explanation in English, and a dictionary of synonyms of the term (see: appendix). The structure of the proposed dictionary base will be as follows:

As can be seen from the table, the dictionary database consists of a term in Uzbek, its explanation, a grammatical/linguistic tag, a synonym of the term, a symbol, an equivalent in English, and an interpretation of the term in English.

According to the general structure, the dictionary consists of the following components:

1. Dictionary interface - a window for using the dictionary, performing a search, and obtaining general information about the dictionary.
2. Search window - a field that allows the user to search for a word/term.
3. Results window - a window that displays information about the searched term.
4. Dictionary data warehouse - a warehouse that stores electronic dictionary data. It is in tabular form and appears to the user as a dictionary article rather than a tabular form.

In short, the integration of the electronic dictionary with artificial intelligence creates the basis for easy and convenient learning of Uzbek language terms by different peoples at the international level. Also, reflecting the words characteristic of our national terminology in this dictionary can have a positive effect on the terminology of the neighboring nations, and can be a resource for use. For example, the existence of national terms such as "thick" and "thin" sounds, "thick" and "thin" words, and "pressure" will be shown in a separate column as the wealth of our language and will be updated. It is known from the world experience that the creation of an electronic dictionary of grammatical terms is an important and gradual work. In this case, it is appropriate to research the Uzbek language on the example of the grammatical terms of its own layer, such as nationalization of the term as much as possible, tagging based on the criteria of natural language processing, dividing it into lemmas, annotating, balancing and matching terms that can enter into a hierarchical relationship with each other.

References:

- Abdurahmonova N.Z. Zamonaviy korpuslarning kompyuter modellari // O'zbekistonda xorijiy tillar. -2020. - № 1(30). - 50-58 б.
- Хамройева Ш. О'zbek tili mualliflik korpusini tuzishning lingvistik asoslari: f.f.d. (PhD) diss. avtoref. - Qarshi, 2018. - 456 б.
- Мельчук И.А. Порядок слов при автоматическом синтезе русского слова (предварительные сообщения) / Научно-техническая информация, 1985. - № 12. - 12-

36 c; Meyer C.F. English Corpus Linguistics. – Cambridge: Cambridge University Press, 2002. – URL: <https://pdfs.semanticscholar.org/c775/4bfab1d0f7>

Савчук С.О. Корпус современной русской прессы: из опыта создания и использования // Труды Международной конференции “Корпусная лингвистика – 2011”. – СПб.: Санкт-Петербургский государственный университет, 2011. – 149–154 с.

Хамроева Ш. Муаллифлик корпусини тузишнинг лингвистик асослари. Филология фанлари бўйича фалсафа доктори диссертацияси. – Бухоро, 2018. – 168 б. – Б. 82.

Rubleva O.S. Slovo v elektronnom slovare (s positsii polzovatelya elektronnyimi risurasami): author reference. Dis. sugar Philol. Nauk. Tver: Tverskoy gosudarstvennyy universitet, 2010. – 20 p.

Komarova Z.I. Methodology and methodology terminograficheskoy semantizatsii // Vocabulum et vocabularium. - Kharkiv, 1995. Vyp. 2. - p. 43-8. - S. 43-44.

Durnovo N.N. Grammaticheskyy dictionary. - SPb.: Izd-vo Frenkel, 1924. - 155p.

Hojiev A. Explanatory dictionary of linguistic terms. - Tashkent: "National Encyclopedia of Uzbekistan" state scientific institute 158 p.