

## COMMON AND DIFFERENT ASPECTS IN A SET OF COMMENTARIES IN DICTIONARIES AND SEMANTIC TAGS

**Akhmedova Dildora Bakhodirovna**

**A teacher, BSU**

**Hamidova Iroda Olimovna**

**A student, BSU**

***Annotation:** This article discusses dictionaries that can be a source of information for the language corpus, the structure of the dictionary commentary, and the possibilities of this information in tagging corpus units.*

***Key words:** corpus linguistics, lexicography, lexical units, semantic tagging, annotated dictionary, general vocabulary, limited vocabulary, illustrative example.*

### **I. Introduction**

Corpus linguistics is closely related to lexicography because the linguistic unit in the dictionary, whose interpretation serves as a linguo-lexicographic supply for tagging corpus units. Digitization of the text of existing dictionaries in the Uzbek language is the first step in linking the dictionary with the corpus. Uzbek lexicography has come a long way, has a rich treasure that serves as a material for the corps.

According to Professor E. Begmatov, systematicity in the lexicon is not as obvious as in other levels of language. Lexical units are more numerous than phonemes, morphemes in terms of quantity, and have the property of periodic instability. Therefore, it is not possible to identify and study the lexicon on a scale. Today, world linguistics has begun to solve this problem with the help of language corpora and has already achieved results. Qualitative and quantitative inventory of vocabulary, the possibility of comprehensive research has expanded.

The main source for semantic tagging of language units is the explanatory dictionary of the Uzbek language. The dictionary contains more than 80,000 words and phrases in modern Uzbek literary language, terms related to science, technology, art and culture, words used in several dialects, some historical and obsolete terms. The practical application of the words given in the dictionary is proved by examples from the Uzbek literature and press of the XX century. The dictionary is intended for specialists in Uzbek linguistics and Turkic studies,

translators, media workers, teachers and students of higher education institutions, as well as the general public. There is a logical system in the structure of the dictionary article. During our observations, we were once again convinced that its structure consists of the necessary parts. Applying the experience of dictionary commentary structure in semantic tagging of a word gives an effective result. Below we explain the structure of the dictionary article and the possibility of using it in the semantic tagging of language units.

## **II. Analysis**

Auxiliary words formed by the auxiliary word from the main word entered in the dictionary, if there are fixed compounds, they are in the dictionary article formed by the main word, at the end of the dictionary article formed by the main word, artificial words with black letters before the new paragraph, from the dictionary articles then the stable compounds are given as an explanatory unit; they compiled a dictionary article for themselves.

Fixed compounds are also included in the dictionary based on the requirements for the words to be included in the dictionary (mentioned above). That is, common, as well as commonly used fixed compounds with a limited scope, are also common in written speech. These are:

1. Commonly used terms in various fields, both oral and written: *milliy daromad, daromad solig'i, atrof-muhit, ma'muriy jazo, mehnat daftarchasi, qon bosimi, jarima to'pi and etc.*

2. The meaning is not equal to the meaning of the words in the composition (usually considered a compound word): *nazar solmoq, javob bermoq, loyiq ko'rmoq, ozor bermoq and etc.*

3. Compounds that have a figurative meaning in a word: *oqko'ngil, ko'ngli oq, qora ko'ngil, ko'ngli qora and etc.,*

4. Phrasal verbs (phraseology): *ammamning buzog'I, uchiga chiqqan, oyog'ini qo'lga olmoq, toshingni ter and etc.,*

5. Proverbs and sayings that require explanation: *Temirni qizig'ida bosmoq. Sinamagan oting sirdan o'tma. So'ir ko'zga surma. Ko'za kunda emas, kunida sinadi.*

When tagging corpus units, such artificial words and fixed compounds will be in a separate, independent lemma position: *<daromad solig'i>, <ma'muriy jazo>, <mehnat daftarchasi>, <oqko'ngil>, <ammamning buzog'i>, <uchiga chiqqan>.*

If the dictionary commentary is composed of the main word itself, or if the artificial word formed from this keyword contains a fixed compound in which the word is present, the dictionary structured to them may also include the internal commentary.

Keywords (independent words) are given in zero form (main form). Only the names of peoples, tribes, clans, as well as some terms are given in the plural. However, the unit form is also given as a keyword. Body search will include both a

word search engine and a lexeme search engine. Only in the corpus is the main word (lemma) highlighted.

Verbs are given in the form -moq (ishlamoq, kelmoq); this form is conditional: it has no grammatical meaning. When tagging corpus units, all words except the verb group are lemma according to the rule “o’zak + negiz = lemma” (core +base = lemma). Only the words in the verb category are not in the -moq form as in the dictionary, the stem is equated to a lemma. For example, wordform of *otalar* is <ota>, wordform of *qiyinroq* is <qiyin>, verb wordform of *kelyaptiis* <kel>, wordform of *o’qigan* is <o’qi>. It is not correct to lemmatize in the form of *o’qimoq*, *kelmoq*, as in the dictionary. Because the search is not done with the form -moq. There are such similar, different aspects in the dictionary article and body corpus of keyword forms.

In a dictionary article, if a form of a word has acquired a meaning and function other than that of a particular form and function, that form is given as the keyword; each meaning, function is recorded. In the corpus, this is a much more difficult issue: grammatical / morphological homonymy.

If the keyword is used independently of a word or in a fixed phrase, it is followed by two dots (:) and then given by that word. For example:

*LOM-MIM: lom-mim demaslik...*

*NIL: bo’yoqchining nili...*

To obtain such units as a corpus unit, the fixed form is defined as a lemma: like<*lom-mim demaslik*>.

Separate artificial words and fixed compounds are also given mainly in the form of the above principle, i.e., in the case of a noun, in the main consonant, and in the form of a verb. But some fixed (phraseological) compounds are used only in a certain form in speech, and they are given in the same form in the dictionary. For example, expressions like *ilonning yog’ini yalagan*, *yog’ tushsa yalaguday*, *ilon po’st tashlaydi* are used only in this form; given in the same form in the dictionary.

The post-key component of a dictionary article is information about the etymology of the key word. Etymological information is given to cognate foreign words. In this case, the language to which the word belongs (with a conditional abbreviation), the spelling and meaning of the word in that language are noted.

In the semantic tagging of language units, this parameter should also be taken into account and a brief information about the etymology of the word should be included in the linguistic base.

The next part of the dictionary article after the etymological information is the grammatical, methodological and reference notes given before the main word, its meaning, which are given in the form of abbreviations.

Defining and recording the meaning (meanings) of explanatory units (keywords and fixed expressions) is one of the most basic, permanent parts of a dictionary article. The value and success of the dictionary is also determined by the

degree to which this work is done.

The most important part (component) of a dictionary article is a commentary on the meaning of interpretive units. It has been said above that none of the meanings of the word recorded in the dictionary should be left unexplained. But not every meaning of the word mentioned in the dictionary needs to be interpreted here. It is not even possible to do so in a glossary. Because if this is done, the dictionary will be filled with repetitive comments.

Language corporations differ from dictionaries in this respect by the breadth and infinity of their possibilities. There is no need to refer to another word as in the dictionary (in order to avoid repetition) in language corpora. Artificial intelligence has a large volume of memory, paper is not wasted, the volume can be expanded as desired. This is no problem for the user as millions of words are only a few megabytes. Why do you need a link from word to word in the dictionary? Macalan, there are many synonymous words in the language. If the same meaning is explained in each of them, the repetitive comments in the dictionary will increase. To prevent this, it is achieved in two different ways to ensure that every meaning of the word recorded in the dictionary has an explanation:

1) to interpret the meaning of the word by definition, explanation, that is, to interpret the meaning;

2) referring the meaning of a word to the meaning of another word.

The last component of the dictionary article is an illustrative example given after the explanation of the meaning of the explanatory unit. An illustrative example, although given after the interpretation of each meaning, ensures that the interpretive unit (including the keyword) is included according to the stated requirements, that the grammatical and stylistic markings are correct, that the meaning of the interpretive unit is correctly defined; is given to confirm the correctness of the interpretation given to a meaning.

#### **References:**

1. Begmatov E. Lexical layers of modern Uzbek literary language. - Tashkent: Fan, 1995.

2. Annotated dictionary of the Uzbek language (edited by A. Madvaliev). 5 volumes. - Tashkent: State Scientific Publishing House "National Encyclopedia of Uzbekistan", 2000-2006.

3. Khamroeva Sh. Linguistic bases of creation of the Uzbek language author's corpus: dissertation of Doctor of Philosophy (PhD)- Bukhara, 2018.

4. Akhmedova D. Tag groups in a semantic markup system. "Oriental art and culture" scientific methodical journal. - Kokand, 2020. №3.– B.440-444.

5. <http://www.ruscorpora.ru/reqsem.html>

6. <http://www.natcorp.ox.ac.uk/>

7. Rajabboyevna A. N. Individualization in education and methods of improving teaching the english language //International Journal of Psychosocial Rehabilitation. – 2020. – T. 24. – №. 1. – C. 91-96.
8. Alimova N., Radjabova M. THE ROLE AND IMPORTANCE OF INDIVIDUAL EDUCATION IN THE SYSTEM OF ORGANIZATION //Theoretical & Applied Science. – 2020. – №. 4. – C. 401-404.
9. Alimova N. R. IMPROVING TECHNOLOGY OF INDIVIDUALIZATION ON EDUCATION FOR STUDENTS OF TECHNICAL SPECIALTIES IN TEACHING ENGLISH //Theoretical & Applied Science. – 2019. – №. 12. – C. 352-355
10. Rajabboyevna A. N. Organization of English Lessons Based on Advanced Pedagogical Techniques //International Journal on Integrated Education. – T. 3. – №. 3. – C. 41-45.
11. Alimova N. POSSIBILITIES FOR INDIVIDUALIZING TEACHING FOREIGN LANGUAGES IN A NON-LANGUAGE UNIVERSITY //European Journal of Research and Reflection in Educational Sciences Vol. – 2019. – T. 7. – №. 12.
12. Nozima A. Content And Methods Of Individualization Of Teaching Activities //International Journal of Progressive Sciences and Technologies. – 2021. – T. 25. – №. 1. – C. 50-53.