



LingvoDoc: Working with Verb Valency

Natalia Kosheliuk

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

August 7, 2022

LingvoDoc: working with verb valency¹

Kosheliuk Natalia

ORCID 0000-0002-5833-7971

Ivannikov Institute for System Programming of the RAS, Moscow
(Russia)

NKoshelyuk@yandex.ru

Abstract. In the article, using the example of the corpora of various Uralic and Altaic languages presented on the LingvoDoc platform for which the parser has been developed, an algorithm of work with the removal of homonymy and subsequent identification of the verb valency is given.

Keywords. LingvoDoc, data mining, linguistics, valency.

1 INTRODUCTION

Valency is the ability of a word to attach dependent words to itself in certain forms. Without these dependent words, the use of the main word is incomplete. Valency are inherent in different parts of speech. In view of the new option introduced on LinvoDoc for working with the *verb valency*, in this article we will focus in detail on this topic.

The words that fill in the valences of the verb are called actants. This term belongs to the French linguist Lucien Tesnière. He was the first to express the ideas that are most common now in modern linguistics: in the center of the sentence is an action, usually denoted by a verb. All other members of the proposal directly or indirectly depend

¹ Supported by Russian Science Foundation, project no. 20-18-00403 'Digital Description of Uralic Languages on the Basis of Big Data'.

on it. Actants and sirconstants are directly dependent on the verb. Actants correspond to the subject and additions in traditional grammar, and sirconstants correspond to circumstances. For example, in the sentence "In the evening John met Jim", the words "John" and "Jim" are actants, "in the evening" is a sirconstant. Actants fill in two valencies of the verb "to meet", the sirconstant "in the evening" does not fill in any valencies, since the expression of this meaning is not mandatory for this verb.

The valencies discussed above are called syntactic, they correspond to certain forms of expression, that is, syntactic actants. In addition to syntactic ones, semantic valency are also distinguished in modern linguistics, that is, those semantic roles that exist in each situation. Semantic valency is the property of a verb to attach to itself any semantic actant. Among the semantic actants, the subject, object, addressee, and instrument are distinguished (etc.).

Below, already having a general idea of what verb valence is, we present a description of the steps for working with *the verb valency option* on the LingvoDoc platform.

2 HOW TO ANALYZE THE VERB VALENCIES ON THE LINGVODOC

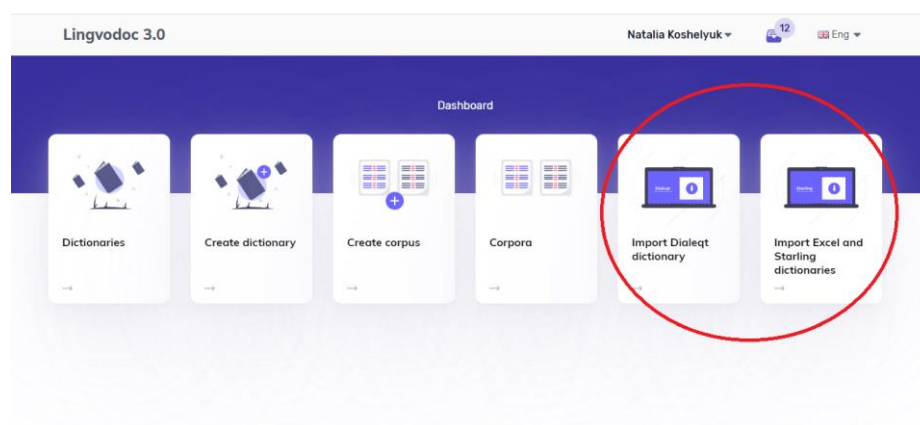
In order to make it possible to qualitatively identify the valency bonds of verbs in the corpus of the language under study, the scientist needs to upload linguistic materials (dictionaries, books, fairytales, news in a literary language or dialects) into one of the corresponding parsers developed by T. Arkhangelsky and remove the homonymy

identified in them. Currently, parsers for Udmurt, Moksha, Erzya, Khanty, Samoyed, Altai, Chulym, and Yakut languages have been implemented on the LingvoDoc platform.

Using the example of the Udmurt corpus, we will show what steps need to be taken to remove homonymy as a preparatory step to determining valency bonds, and subsequently implement *the valency option* itself:

1) Upload the texts necessary for analysis in the usual way via Starling or Excel (Pic. 1);

Pic 1. Import Excel and Starling dictionaries (or use the classic way).



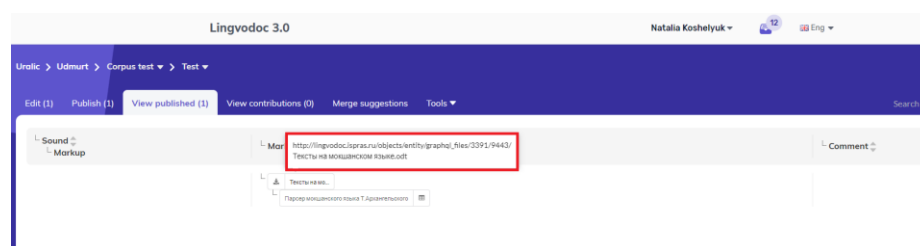
2) Through the search or personal account, select the desired language and go to the parser (Pic. 2);

Pic 2. Language and dictionary selection.



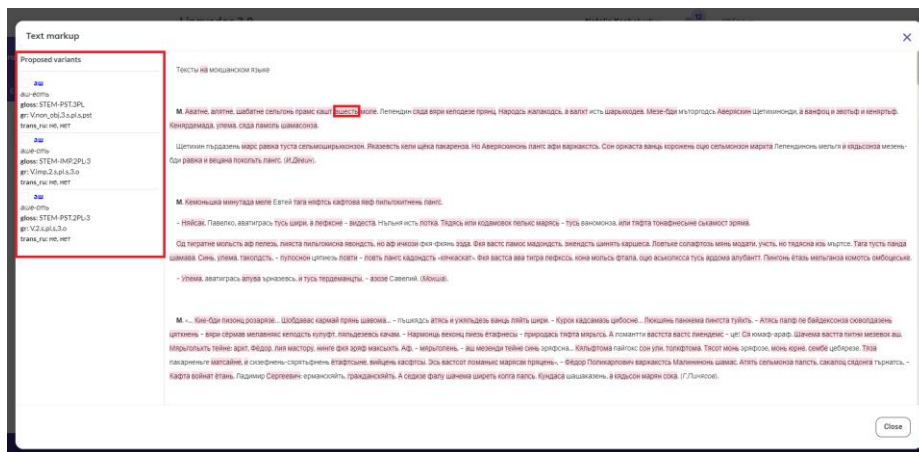
3) In the suggested menu, select one of the downloaded files (Pic. 3);

Pic 3. Opening the text under study in the parser.



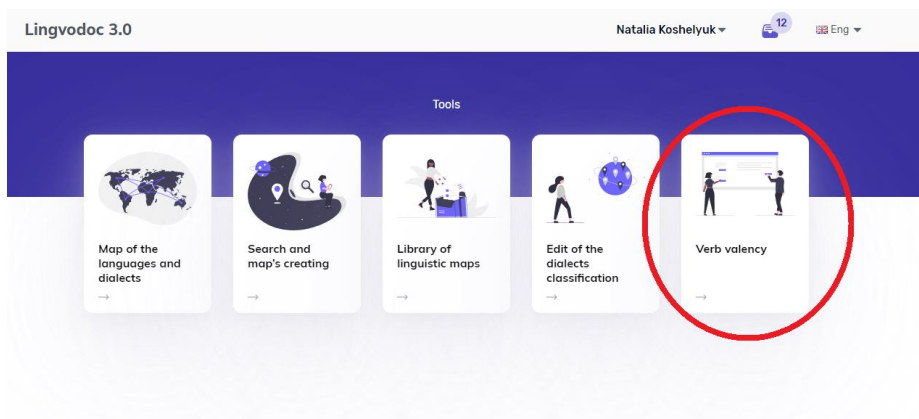
4) In the opened file, the researcher will be able to check each word for the presence of homonymy and remove it (Pic. 4);

Pic 4. Removing homonymy in the text.



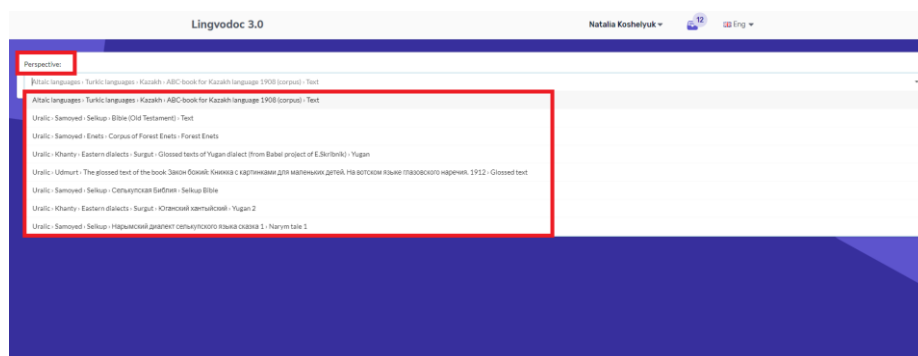
5) Further, after the implementation of the procedure for removing homonymy, the researcher can proceed to the analysis of the valency of verbs. To do this, go to Tools -> Verb Valency (Pic. 5);

Pic 5. Verb valency.



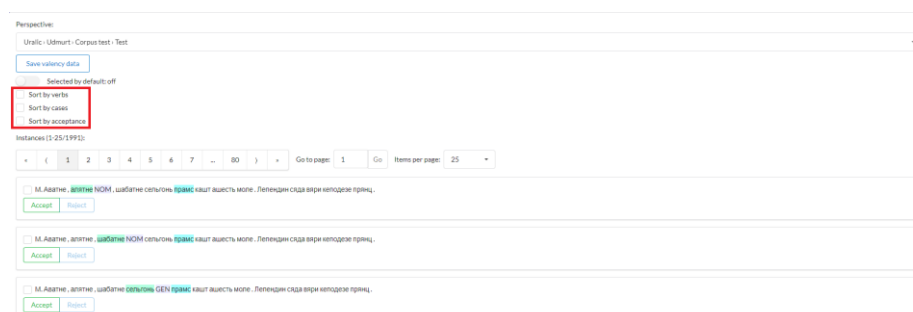
6) In the window that opens, you can select the language corpus on which you plan to conduct the analysis (Pic. 6);

Pic 6. Choosing the right corpus of the texts.



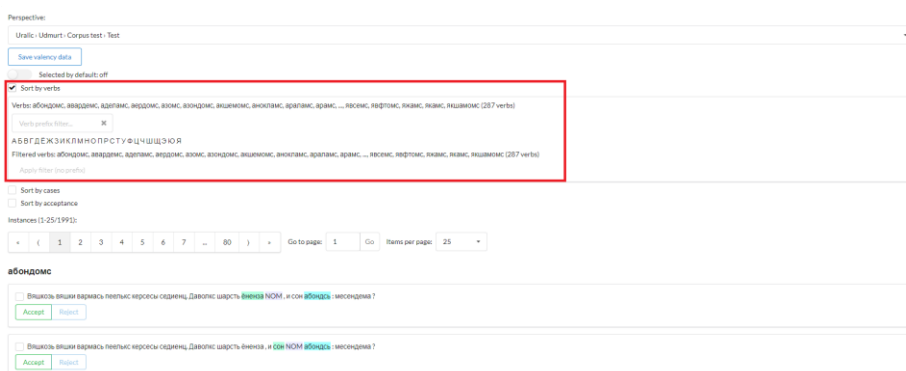
7) Valency analysis can be carried out by sorting by word, case, frequency of occurrence of syntactic connection in the text, or by all at once (Pic. 7).

Pic 7. Sorting options.



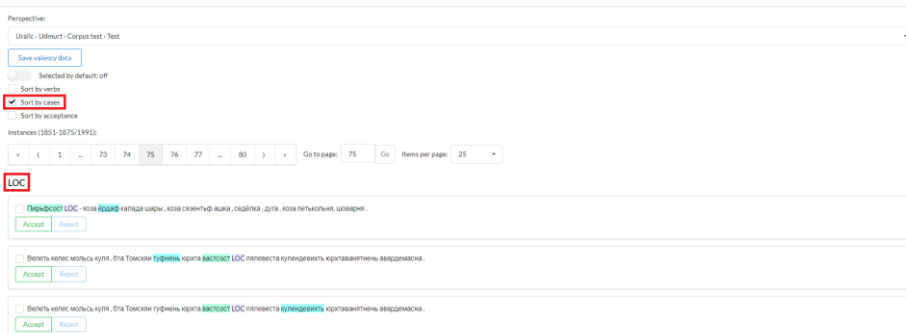
8) Sorting by a specific word involves selecting sentences where it appears. Here it is possible to correlate contexts and valencies characteristic of a particular word (Pic. 8);

Pic 8. Sorting by a specific word.



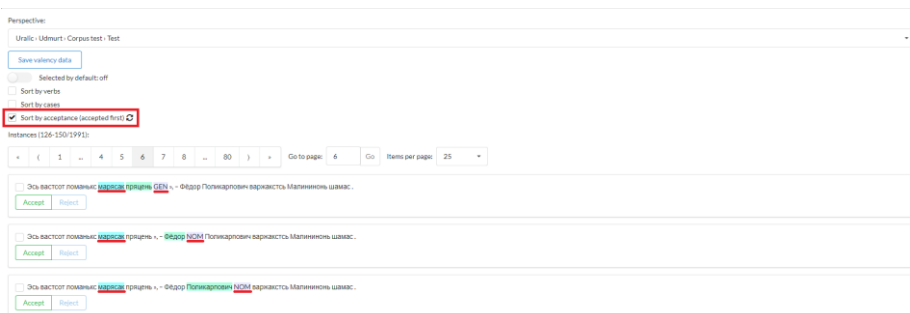
9) The choice "by case" sorts sentences in which potential connections are expressed in one case or another: all sentences with verbs are given only in the locative, nominative, dative, and so on (Pic. 9);

Pic 9. Sorting by case.



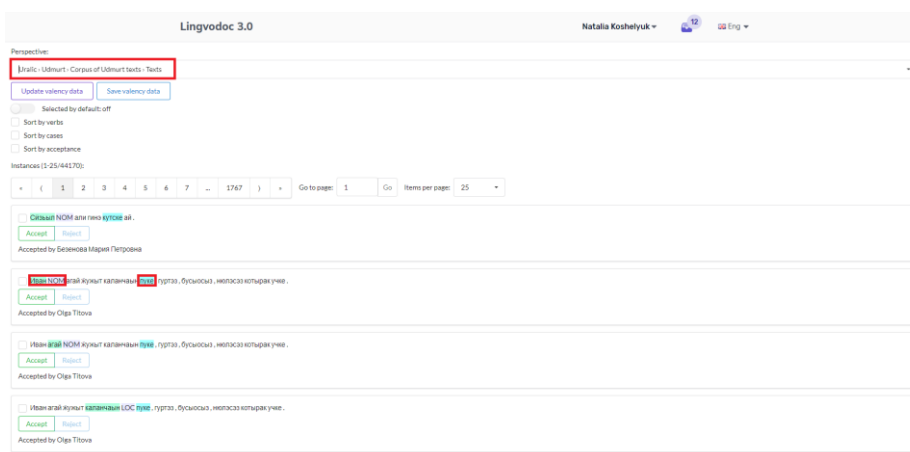
10) Sorting by acceptance demonstrates which cases prevail in the studied material and with which words they are associated (Pic. 10);

Pic 10. Sorting by acceptance.



11) In each of the sentences, the researcher can check whether the valency of the word has been correctly identified and approve it using the appropriate buttons. It is possible to carry out such manipulation with each sentence in the analyzed corpus. The accepted valence is colored with a green marker (Pic. 11).

Pic 11. Showing the verb valencies.



CONCLUSION

The field of phenomena related to valencies is very extensive and not everything in this area has been studied in detail by linguists. This is not least due to the fact that so far the main research has been carried out on a small amount of material, which limits scientists in achieving a significant scientific result. Analyzing the valence of verbs based on big data available on the LingvoDoc platform, in view of the possibility of adding and placing tens and hundreds of thousands of lexemes there, with the help of *the verb valency algorithm*, it allows identify new and sometimes unexpected patterns.

REFERENCES

1. LingvoDoc Homepage, <http://lingvodoc.ispras.ru/>. Last accessed: 24/07/2022;
2. Academic. Encyclopedia (valency of verbs), <https://dic.academic.ru/dic.nsf/ruwiki/313556>. Last accessed: 03.08.2022.