Yevhen Kupriianov/Iryna Ostapova/Volodymyr Shyrokov/Mykyta Yablochkov

VIRTUAL LEXICOGRAPHIC LABORATORY AS A LINGUIST ASSISTANT IN CONDUCTING DICTIONARY-BASED RESEARCHES: CASE OF VLL DLE 23

Keywords Virtual lexicographic laboratory; explanatory dictionary; lexicographic text; lexicographic system; dictionary-based researches

Introduction. The virtual lexicographic laboratories (VLL) developed in the Ukrainian Language Information Fund offer effective tools to conduct dictionary-based researches. In this regard, lexicographic text is considered not only as a reference product but a tool assisted basis for creating and updating a dictionary as well as a means of professional communication and transfer of linguistic knowledge. This mainly concerns comprehensive explanatory dictionaries that are characterized by detailed and multi-aspect description of language. Hence, the problem in question is providing these dictionaries with appropriate tools that make it possible to derive from the dictionary text any linguistic information while conducting linguistic researches. The present paper shares the authors' experience in elaborating such tools while carrying out the project of VLL DLE 23, a virtual lexicographic laboratory to perform linguistic researches on the basis of the Spanish dictionary "Diccionario de la lengua Española. 23ª edición" (DLE 23).

To implement the project of VLL DLE 23 the following tasks were to be resolved: 1) building up a formal model of the dictionary (hereinafter referred to as lexicographic system) on the basis of which the VLL DLE 23 database and interface elements are to be developed; 2) implementing parsing technology to convert the structured dictionary text into database format; 3) selecting the format for the database containing all the information elements of the dictionary entry (the choice was made in favor of a document database); and 4) elaboration of the user interface of VLL DLE 23 which provides the access and work with the information elements of the dictionary entries in the database.

Current state of VLL DLE 23 project. The pilot prototype of the VLL DLE 23 allows users to analyze the vocabulary contained in the word list and to examine the entry structure, as well as to provide access to some information elements of the dictionary entries. The main interface (Fig. 1) consists of the menu bar (1), the search panel (2), the wordlist panel (3), the entry area (4) and the area to view dictionary entry text in HTML format (5) to be used for full-text search in the dictionary.
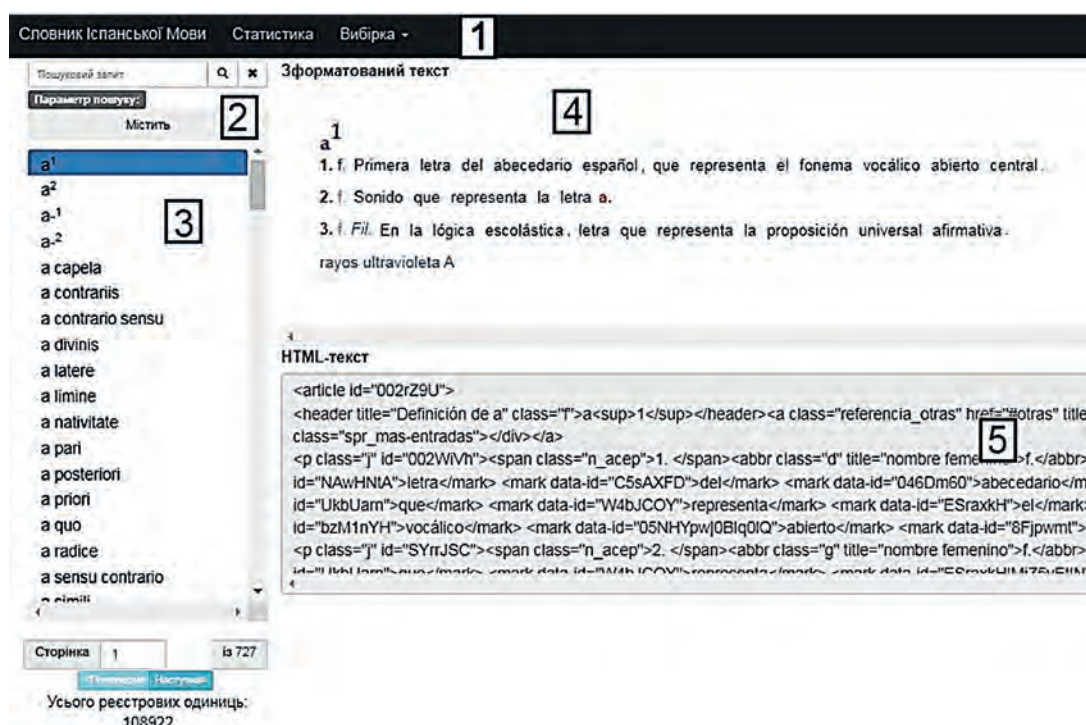
Figure 1: General view of the main window of VLL DLE 23

As the VLL development is still underway, the language of the interface remains Ukrainian. However, in final version the users will be able to switch between English, Spanish and Ukrainian. With the current interface the users can: 1) generate statistics on the whole dictionary or specific sample of the dictionary entries; 2) perform linguistic researches aimed at studying lexical meanings, etymology, grammar and usage peculiarities of the Spanish language units; and 3) build sub-dictionaries on the basis of DLE 23, for example: sub-dictionary of morphemes, homonyms, collocations, etc.

The menu bar contains two main tools: "Статистика" (Statistics) and "Вибірка" (Sample). The first automatically gets indicators for a separate sample or the entire dictionary. The second one offers parameters to form the samples of DLE 23 entries:

- **Headword parameters:** lemma, masculine, feminine, regional variant, not defined;
- **Language level of the headword:** word, collocation, morpheme, not defined;
- **Headword origination:** borrowing, contraction, acronym, not defined;
- **Homonym:** numerical value.

For example, we can get a sample of homonymous foreign words in the modern Spanish language. To do this, the following parameters are to be selected:
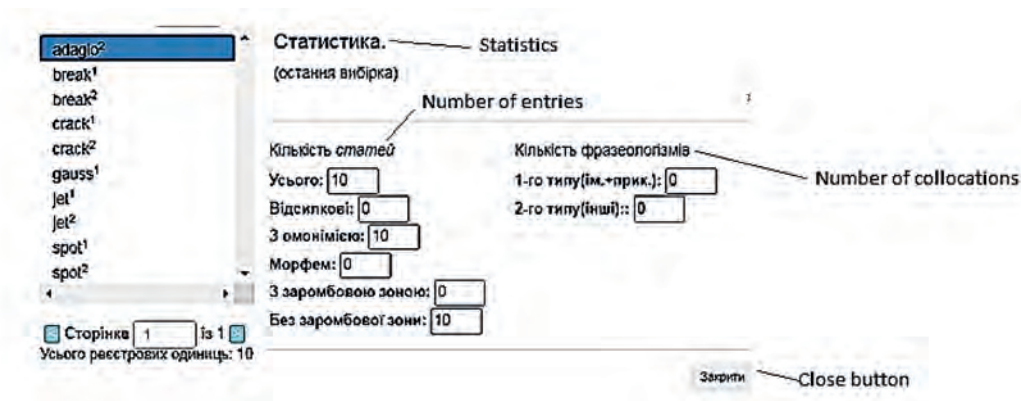
- **Language level of the headword:** word;
- **Headword origination:** borrowing;
- **Homonym:** ≥ 1.

The results are shown in Figure 2 (on the left).
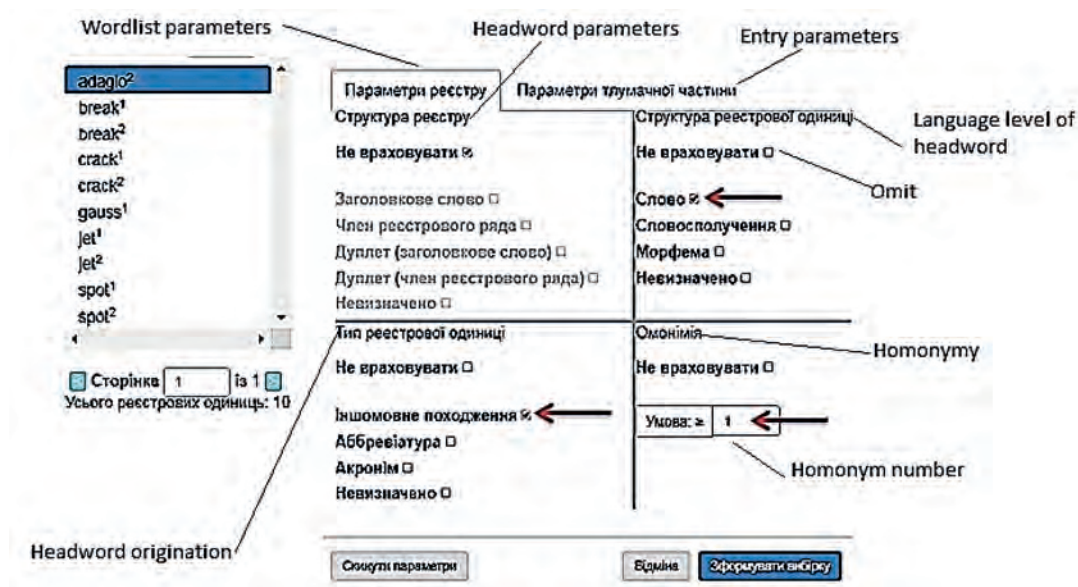
Figure 2: List of the entries with homonymic words of foreign origin (left) and parameters selected (right)

With "Statistics" (Fig. 3) we get the following figures for the above sample:

Figure 3: Statistics for the sample of homonymous words of foreign origin
- Total number of the entries: 10;
- Referential entries: 0;
- Homonyms: 10;
- Entries with collocations: 0;
- Entries without collocations: 0.

Another example is getting a sample of Spanish collocations composed of foreign words (Fig. 4.).
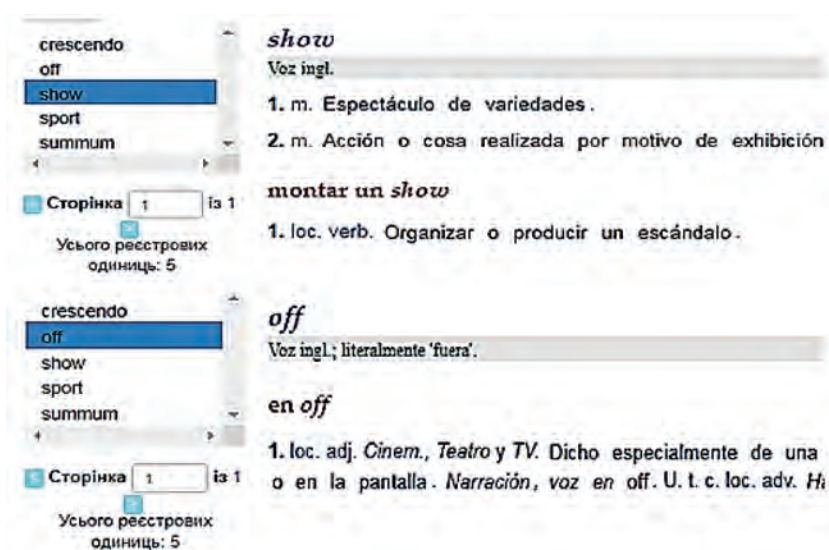
Figure 4: Sample of dictionary entries containing Spanish collocations composed of foreign words

Future development of VLL DLE 23 project. The next version will ensure the access to all the information elements of the entry provided by the dictionary lexicographic system. The users will be able to specify the content of any information element. Further extension of the functionality will be determined by the request for solving linguistic problems.

References

Diccionario de la lengua española (2022): https://dle.rae.es/ (last access: 23-03-2022).

Kupriianov, Y./Ostapova, I./Yablochkov, M. (2020): Design of the user's interface of virtual lexicographic laboratory for explanatory dictionary of the Spanish language. In: Lytvyn, S. (ed.): Proceedings of the 4th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2020), Lviv, 23–24 April 2020. Volume I: Main Conference. Lviv. http://ceur-ws.org/Vol-2604/paper10.pdf (last access: 23-03-2022).

Kupriianov, Y./Shyrokov, V./Ostapova, I./Yablochkov, M. (2021): Digital toolkit to develop research potential of explanatory dictionary (case of Spanish language dictionary). In: Lytvyn, S. (ed.): Proceedings of the 5th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2021), Kharkiv, 22–23 April 2021. Volume I: Main Conference. Kharkiv. http://ceur-ws.org/Vol-2870/paper28.pdf. (last access: 23-03-2022).

Stanković, R./Stijović R./Duško V./Krstev C./Sabo O. (2018): The dictionary of the Serbian academy: from the text to the lexical database. In: Krek, S./Jaka C./Gorjanc, V./Kosem, I. (eds.): Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts, Ljubljana, 17–21 July 2018. Ljubljana, p. 941. https://euralex.org/publications/the-dictionary-of-the-serbian-academy-from-the-text-to-the-lexical-database/ (last access: 23-03-2022).

Contact information

Yevhen Kupriianov
National Technical University "Kharkiv Polytechnic Institute"
eugeniokuprianov@gmail.com

Iryna Ostapova
Ukrainian Lingua-Information Fund, NAS of Ukraine
eugeniokuprianov@gmail.com

Volodymyr Shyrokov
Ukrainian Lingua-Information Fund, NAS of Ukraine
vshirokov48@gmail.com

Mykyta Yablochkov
Ukrainian Lingua-Information Fund, NAS of Ukraine
gezartos@gmail.com