Design of the User's Interface of Virtual Lexicographic Laboratory for Explanatory Dictionary of the Spanish Language

Yevhen Kupriianov[1[0000-0002-0801-1789]],
Iryna Ostapova[2[0000-0001-8221-3277]], Mykyta Yablochkov[2[0000-0003-1175-1603]]

[1] National Technical University "Kharkiv Polytechnic Institute", 61002 Kharkiv, Ukraine
eugeniokuprianov@gmail.com
[2] Ukrainian Lingua-Information Fund, NAS of Ukraine, 03039 Kyiv, Ukraine
irinaostapova@gmail.com, gezartos@gmail.com

Abstract. One of the most effective tools to work with dictionaries in digital environment is virtual lexicographic laboratories (VLL). Unlike electronic dictionaries, they are intended mostly for professional linguists. The paper shares the authors' experience in elaborating the interface of the virtual lexicographic laboratory for Explanatory dictionary of the Spanish language (DLE 23). Using the theory of lexicographic systems a formal model of DLE 23 was elaborated. On the basis of the model the database structure and VLL interface elements were defined. The current version of VLL DLE 23 interface has the following advantages: 1) making an inventory of language units in the dictionary or in a sample; 2) conducting DLE 23-based linguistic researches to reveal lexical-semantic, etymological, grammatical and usage properties of the Spanish language; and 3) building of secondary lexicographic objects or sub-dictionaries on the basis of DLE 23, for example: sub-dictionary of morphemes, homonyms, collocations, etc.

Keywords: User's Interface, Virtual Lexicographic Laboratories, Explanatory Dictionary, Digital Environment, Interface Design.

1 Introduction

For the last two decades, lexicography has undergone significant changes. They relate to the implementation of a wide range of approaches to comprehensive analysis of language vocabulary, creation of integrated lexicographic systems combining different linguistic facts by nature, construction of universal digital lexicographic environments, etc. In this regard, there is a growing interest in the needs and skills of digital dictionary users, which encourages the developers to focus on the user properties of lexicography objects. This interest has led to the fact that most experts now understand the necessity of creating dictionaries tailored to the users' needs and skills [6]. Now the most of printed dictionaries have electronic versions (CDs, online, etc.).

However, there are dictionaries which do not have a printed version, but exist only in digital format [13].

However, some well-known, classic dictionaries that have been traditionally made in paper format are gradually changing to digital format. Among them is Oxford English Dictionary (OED) [7], one of the most reputable academic dictionaries published by Oxford University Press. It traces historical development of the English language (containing all the words that existed or existed in English literary and spoken language since 1150), providing a comprehensive resource for common users and researchers, and describing the usage of its variants around the world.

Another example is "Diccionario de la lengua española" (DLE, Dictionary of the Spanish language) which has been published by Academia Real Española (Spanish Royal Academy) since 1780 [3]. The dictionary went through twenty three editions, changing over time into main reference source of the Spanish language. The most recent, the 23rd saw the light in October 2014. The first electronic version of DLE appeared in CD-ROM at the end of 1995, containing 21st edition and the recent one in 2015, representing 23rd edition (DLE 23) and being available at www.dle.rae.es. Since March 2017, the Academy has been working on 24th edition, which will be digital only [8]. Holding the discussion on further publication of the dictionary in printed form, Academician Alvarez de Miranda gives the following figures: the number of requests to online version of DLE 23 has reached 750 million at the end of 2017, which is an average of 65 million monthly [1].

In our paper, explanatory dictionary is considered as a comprehensive source of information to be used for language researches. Research potential of dictionary is fully developed in digital environment. When working with printed dictionaries, especially multi-volume publications, the researcher has to spend a lot of time for searching, analyzing and summarizing the information he needs. Due to the great amount, elaborated structure and completeness of a lexicographic description such dictionaries are carriers of a huge number of implicitly-defined linguistic, cognitive, logical and other relationships which are difficult or near impossible to be investigated with traditional methods. Also, in paper versions, search is usually limited to headword list only [10].

It is important for digital dictionary to provide access to any structural element of dictionary entry and give an opportunity of selecting particular entries corresponding to the user's presets. In other words, there must be possibility of finding relevant information without knowing the headword. The OED interface offers a powerful tool for researching the vast amount of information in the dictionary. Although OED has extensive capabilities (search by vocabulary classes, categories, thesaurus, etc., advanced search), but it doesn't provide the necessary tools for the researcher. As for DLE 23 online, the interface is limited to headword list only and some filters (e.g. "word", "lemma", "contains", "starts with", "ends with" etc.). To conduct extensive researches the entire text, including its meta-language elements, of a dictionary will be required.

With reference to the above the Ukrainian Lingua-Information Fund has developed and implemented in his dictionary projects a new approach which is called virtual lexicographic laboratories or shortly VLL. In contrast to electronic dictionaries, a VLL is intended for working with the whole dictionary text with clearly defined struc-

ture. It means that the user has a full access to all structural elements of the entry including meta-language information related to a headword (e.g. definition type, availability of word usage examples, headword structure etc.) [11, 12]. The paper deals with the elaboration of the interface for the virtual lexicographic laboratory for the Explanatory dictionary of the Spanish language (VLL DLE 23) and focuses on its research potential.

2 DLE 23 as an Object of the Research

2.1 Choosing the Object of the Research

Our interest in DLE 23 is arisen by the following reasons: 1) international status of the Spanish language; 2) credibility and academic status of the dictionary; 3) other school of lexicography which differs from other schools, such as Ukrainian and English. In addition, the dictionary in question is of interest to translation lexicography while creating translation systems: Spanish-Ukrainian and Ukrainian-Spanish. In this context, the Russian linguist L. Ščerba states that each language pair needs four dictionaries, one for explanation and one for translation in each direction – one for comprehension and one for production purposes for each speech community [9]. Another and also important reason of choosing DLE 23 is the availability of its digital version that supports HTML5 format. The letter guarantees the authenticity of the dictionary text and allows us to focus our attention on the structure of dictionary entry.

DLE 23 is a fundamental lexicon containing standard vocabulary to be widely used both in Spain and in Latin America. The purpose of the dictionary isn't limited to giving a reference on lexical meaning of language units. Each entry also includes detailed description of their grammatical, syntactic and pragmatic features. It should be noted that the dictionary ideology is based on the conception of the Spanish lexicographer J. Casares. According to this conception: the dictionary isn't a set of alphabetically arranged entries but it's also a tool providing a user with necessary resources for searching appropriate words and phrases he may need during communication process [2].

The headword list covers language units which: 1) belong to Spanish vocabulary and 2) have come from other modern languages (English, German, French etc.). The former ones are in normal font and the letter ones are in italics: "**gato**", "***kilobyte***", "**ojo**", "***software***" etc. The list also includes frequently used abbreviations ("**DNA**", "**ONU**") and acronyms ("**radar**", "**laser**"). Furthermore, the dictionary contains prefixes ("**a-**", "**pre-**", "**contra-**", "**pro-**" etc.), suffixes ("**-aico**", "**-ino**", "**-ivo**" etc.), derivational elements of Greek and Latin origin ("**archi-**", "**hidro-**", "**-ónimo**") and Latin expressions ("***ad hoc***", "***a priori***").

2.2 Lexicographic Description of Spanish Language in DLE 23

The main parts of the dictionary entry, as shown in fig.1, are: a headword with its feminine flection (1); headword information (2); a set of definitions (3); a set of sub-

entries of headword's collocations and set expressions (4); cross-references to other entries (5). A brief description of each element is given below.

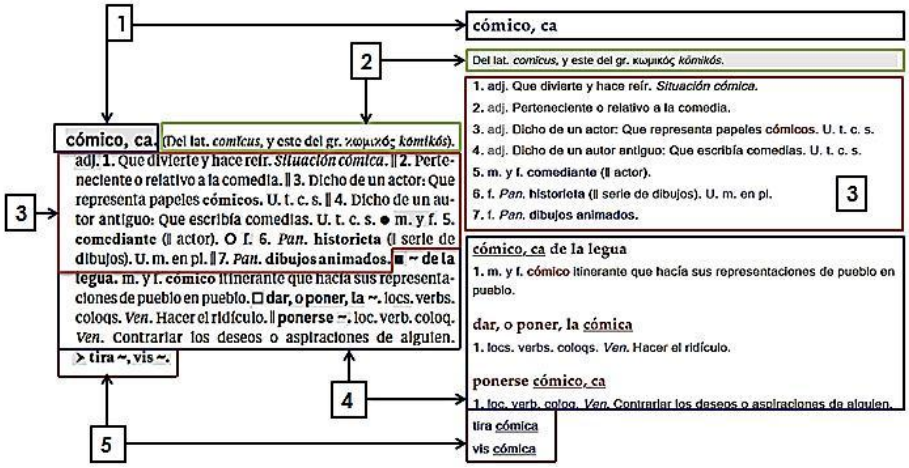

Fig. 1. DLE 23 entry in paper (left) and online version (right).

The first element, lemma, (1) can be of masculine, feminine or both forms. In letter case the masculine form goes first, and then the feminine form, represented by respective ending, follows after. For example: "**alcalde**, **desa**", "**duque**, **quesa**", "**gato**, **ta**" should be read as *alcalde*, *alcaldesa*; *duque*, *duquesa*; *gato*, *gata*.

Headword information (2) includes lemma variants, etymology, word flection and orthography. Some examples of headword information are shown in Table 2. It should be noted that some or all of these information elements aren't always provided in the dictionary. But for building the lexicographic model of DLE 23 they are considered to be present in the entry structure.

Table 1. Headword information in DLE 23

Element	Information provided	Examples (in Spanish)
Lemma variants	Variants to be used in Spain or some regions of Latin America	Tb. **salvavida** en acep. 5, *Cuba*
Etymology	Language of origin, etymon and original meaning	Del ant. *fruente*, y este del lat. *frons*, *frontis*
Word flection	Conjugation model for irregular verbs, superlative degrees of comparison of adjectives etc.	Sup. irreg. **fortísimo**; reg. **fuertísimo.**
Orthography	Writing lemma with capital letter or with accent mark in particular meaning	Escr. con may. inicial en acep. 2

The set of definitions (3) contains the explanation of lexical meanings together with different labels indicating particular limitations on using the word. There are six types of the labels in DLE 23 to denote grammatical class ("adj.", "adv.", "pron."), usage

("iron.", "despect.", "peyor."), style ("coloq.", "estud.", "fest."), domain ("*Comp.*", "*Dep.*", "*Der.*" etc.), region (*"And."*, *"Cat."*, *"Cád."*, *"Vall."*, *"Arg."*, *"Am. Cent."* Etc.) and time period ("ant.", "desus.", "germ.", "p. us."). There are five types of definitions used in DLE 23 to state the meaning of lemmas: standard, by synonym, explanatory, and others not particularly stated by the dictionary makers. A brief characteristic of each definition type is given in Table 3.

Table 2. Definition types and their characteristics

Definition type	**Way of explanation**	**Examples (in Spanish)**
Standard	By indicating the broader term and specific differences	Embarcación de estructura cóncava y, generalmente, de grandes dimensiones
Contextual	By illustrating the meaning in context	Dicho de un metal: Frágil, quebradizo, no dúctil ni maleable
By synonym	By another word which is close to lemma by its lexical meaning	**fachento ... jactancioso**
Explanatory	By asking the questions like "for what purpose?", "how is it used?"	U. para reproducir ciertos sonidos
Others	By enumerating the objects (concepts) or characteristics covered by lemma	Lógica, física y metafísica

The definitions may be supplied together with examples to illustrate the usage of the word or represent a pattern to make up a sentence with the headword. Another peculiarity of DLE 23 is indicating additional grammatical or usage features of a headword in a form of the comments, like "U. t. c. s. m." (Usado también como sustantivo masculino – Also used as a noun masculine) or "U. t. en sent. fig." (Usado también en sentido figural – Also used in figurative meaning).

The headword's collocations are described in separate subentries (4) and treated in the same way as the headword itself. In DLE 23 they fall into two types: 1) noun + adjective: *agua bendita*, *agua blanca*, *agua corriente*; and 2) others: a) verbal: *agarrar alguien un agua*, *bañarse en agua rosada*, *beber agua un buque* etc.; b) adverbial: *como agua de mayo*, *como agua para chocolate* etc.; c) adjectival: *de agua y lana*, *de arte y ensayo* etc.

The last element of the entry is the list of the cross-references (5) to other collocations described in the other entries.

3 Implementation of VLL DLE 23 Project

The project of VLL DLE 23 is planned to be implemented in two stages: 1) creating a shortened version of VLL with minimum interface elements to test some technological solutions and 2) developing fully functional application with expanded interface. Currently VLL DLE 23 is at end of the first stage and it demonstrates more capabili-

ties for working with the dictionary in digital environment than the original online version of DLE 23.

The first task was building up a conceptual model which would serve as a basis for elaborating database and interface elements. The conceptual model of the dictionary has been built on the basis of HTML text taken from online version of DLE 23. This text shows much deeper and more transparent mark-up of entry structure than printed version does. With conceptual model it is easy to determine a set of information elements of the dictionary to be accessible in database.

The second task consisted in the development of database and choosing database type for VLL DLE 23. According to our experience, relational databases proved to be inappropriate for developing efficient digital lexicographic systems. In this case the data is stored implicitly as a set of several tables and relationships between them. Operating separate tables as a single object requires building a powerful software infrastructure. Furthermore, the evolutionary potential of such a digital object is limited by database opacity.

Since the dictionary entries are the elements of lexicographic system with a strictly defined structure, it is logical to represent them as classes in object-oriented programming languages, with subsequent processing, editing, and storage in explicit form. This possibility is provided by the so-called NoSQL databases (document type databases). For databases of this type the main element to be stored and processed is a document (object) with a strictly described structure.

The main advantage of NoSQL databases for our project is their capability to store explicitly lexicographic objects without altering their internal structure, which opens up direct access to each element of the lexicographic object and greatly simplifies the possibility of its editing and modification (expansion).

When choosing a specific NoSQL database, we were guided by the following criteria: 1) ease of use; 2) support for transaction mechanisms; 3) support for parallelism; and 4) free of charge for scientific purposes. Taking into consideration these criteria, LiteDB (http://www.litedb.org/) was chosen. This is a relatively simple, free copy of the shareware MongoDB database. An additional advantage of this database is the ease of installation and connection, since LiteDB is implemented as a single library file (dll) and one settings file (xml), rather than a whole software package.

For efficient operation each language unit is put into correspondence with the set of parameters: 1) headword variants; 2) headword structure; 3) headword type; 4) homonymy; 5) number of collocations “Noun + Adjective” and 6) number of collocations of other types. In our opinion, this set of parameters will be enough for selection of the entries and analysis of their structure and content.

The final task was to make a Web application to work with the database of VLL DLE 23. The application based on .Net Core 2.1 technology was created. For easy creation and further editing of interface elements, a set of HTML, CSS templates and JavaScript Bootstrap scripts were used.

4 Method and Technology

The modern digital lexicography has turned to a multidisciplinary field and refers to: a) applied research area originated at the intersection of linguistics and computer science that studies the application of methods and techniques of information science and technologies to creating a wide range of lexicographic systems; and b) a branch of computer industry which is developing rapidly mainly due to the fact that the lexicographic description is one of the efficient ways to obtain and disseminate knowledge [10].

In this context, we consider dictionary text primarily not as a reference system, but as a way of transferring linguistic knowledge. First of all, this statement concerns fundamental lexicons, i.e. big explanatory dictionaries. This sets the problem of providing dictionaries with appropriate tools. Obviously, it can be achieved only in digital environment.

Effective solution of this problem requires general theoretical basis to describe the widest possible range of lexicographic objects. As such, we use the theory of lexicographic systems [11].

The theory is based on a rather universal phenomenological principle, characteristic of any system where information processes take place. These processes are called lexicographic effect in information systems [12].

We consider the lexicographic system (L-system) as a special informational (semiotic and semantic) system, in which a lexicographic effect (or a certain combination of lexicographic effects) is induced. The formal representation in simplified form is as follows:

$$\left\{D, Q, I_0^Q(D), V\left(I_0^Q(D)\right), \beta, \sigma[\beta], Red\left[V\left(I_0^Q(D)\right)\right]\right\} \quad (1)$$

The elements of the model get the following interpretation according to the theory of lexicographic systems:

- D is a modeled object (area);
- Q denotes lexicographic effect which induces a class of relatively stable information entities;
- $I_0^Q(D) = \{x_i\}$ refers to a class of elementary information units in relation to lexicographic effect Q;
- $V(I_0(D))$ designates a set of descriptions (interpretations) of elementary information units;
- β is a subset of structural elements composing $V(I_0(D))$;
- $\sigma[\beta]$ denotes a separate substructure generated by a σ-operator within β;
- $Red[V(I^Q(D)]$ indicates a process of a recursive reduction that decomposes the structure of lexicographic system into its fine elements.

We consider the dictionary as a lexicographic system of special type where $I_0^Q(D)$ is a set of headwords. By $V(I_0(D)) = \{V(x_i)\}$ we mean a set of entry texts, $V(x_i)$ is the entry describing the headword (lemma) x_i. The subset of structural elements (β) are entry text fragments containing a piece of lexicographic description. If we apply $\sigma[\beta]$ to

$V(x_i)$, we shall get all β-elements, i.e. lexicographic descriptions, related to the headword x_i only.

One of the main aspects in the definition of an L-system as an information system of a special type is the concept of its architecture. We use ANSI/X3/SPARK architecture consisting of three levels of data representation: conceptual, internal and external. Conceptual model (conceptual level of representation) of the subject area is a semiotic, semantic model in which the concepts of the subject area are integrated in an unambiguous, final and consistent way. Internal model (the internal level of presentation) specifies the types, structures and formats of presentation, storage and manipulation of data, algorithmic base and software environment in which the conceptual model is implemented. External model (external level of presentation) reflects the end users' views (and, therefore, applied programmers) on the subject area. The model implements a set of tools that enable a user to manipulate the data represented at the internal level. One conceptual model may correspond to several internal and external models.

As it was mentioned above, the Ukrainian Lingua-Information Fund has developed the software systems to support creating, maintaining and functioning of the dictionaries in digital environment named virtual lexicographic laboratories (VLL). Moreover, VLLs permit corporate work on large lexicographic projects by dictionary makers living in different cities and even different counties but having equal access to the dictionary and working tools. The first VLL was created in the Ukrainian Lingua-Information Fund in 2001 and is being used to create and maintain the multivolume explanatory Dictionary of the Ukrainian language. At present, the Ukrainian Lingua-Information Fund has designed over forty virtual lexicographic laboratories for lexicographic systems of various types (http://lcorp.ulif.org.ua) [12].

The advantages of VLL include the following: almost unlimited potential for the integration of various linguistic facts in one object, ability to reflect language dynamics, efficiency of navigation through the structural elements, possibility for computational experiments. The possibility of multiple use of once formed lexicographic structures and arrays by many professionals (linguists, linguistic technologists and publishers) provided by the digital environment is also of great importance.

All entry elements represented in conceptual model are displayed in database structure. Any VLL provides direct access to each structural element of the lexicographic system and offers the possibility to construct different index schemes. As a rule lexicographic database of a dictionary displays only basic structures. Further expansion is determined by two factors: the allocation of more subtle structural elements of the basic structures and the introduction new parameters of a dictionary entry. The increasing of parameters of lexicographic system sets two tasks for a computer tool system: to create a form for the effective representation of parameters and to develop an interface circuits to work with them [10].

The virtual laboratories have been initially developed to support dictionary-making process in digital environment. They don't fit for comprehensive analysis of dictionary text due to the lack of respective tools. Therefore the problem is to build up a

VLL provided with tools to perform dictionary text analysis and linguistic researches on the basis of the dictionary.

5 Conceptual Model of DLE 23

The lexicographic data model is used as a conceptual model, in which the structural elements of the dictionary entry and the relations between them are fixed. All selected elements of the dictionary entry are displayed on the structure of a computer database, and that provides both direct access to each of them and ability to build a variety of index schemes. The electronic text and the interface of online version gave us necessary material for constructing the model.

In dictionary structure we can distinguish a set of headwords $W = \{x\}$ which serve as identifiers of corresponding entries $V(x)$. As we have already mentioned, the list of the headwords includes morphemes, words, abbreviations and collocations. For convenience, all language units that compose the list will be considered as headwords.

In its turn, the structure of each entry $V(x)$ is decomposed into left part $L(x)$ which contains headword characteristics, and right part $P(x)$ where the semantics of a headword x is given. As the object of lexicographic description covers the headwords of two types (words and collocations) it would be logical to represent the formal model of DLE 23 in the following way:

$$V(x) \equiv V^{Lex}(x) \cup \left[\cup_i^{n(x)} \cup_j^{m(i)} V_i^{jFras}(x)\right], \quad (2)$$

Here $V^{Lex}(x)$ is a lexicographic description of the headword x; $V_i^{jFras}(x)$ is a description of the j-th collocation of i-th type; $m(i)$ is the number of phrases of i-th type, and $n(x)$ is the number of collocation types in the dictionary entry $V(x)$. As it was mentioned in section 2, there are two types of collocations: noun + adjective and others, i.e. with other parts of speech. Both lexicographic descriptions V^{Lex} and V_i^{jFras} is put in correspondence with a basic structure composed of left part and right part, respectively:

$$V \equiv (L_0; P_0), \quad (3)$$

In case of $V = V^{Lex}(x)$, L_0 refers to entry headword with its lexicographic description. As for $V = V_i^{Fras}(x)$, L_0 denotes a collocation with its description. The right part P_0 for headword and collocation is the same by its structure.

Based on the DLE 23 text analysis [4, 5], we distinguish the following parameters for L_0 component: *RR* (headword with gender forms or collocation with its variants), *DUPL* (regional variant), *ETYM* (etymology), *MORPHO* (word flection), *ORTHO* (orthography) and *UNCRT* (uncertain). We introduce another parameter $PARHW_i$ (i-th parameter of a headword) for any parameter regardless of its type. *UNCRT* refers to a type of lexicographic description which can't be identified by formal markers but it can be regarded as an information element of headword. Each parameter is represented in our model by text line (authentic original text).

The mandatory parameter is *RR* which describes a headword with its gender forms (or collocation with its variants), and the others are optional. Our model doesn't impose any restrictions on the number of parameters. Moreover, we assume that the entry may include several parameters of the same type. Each parameter is a text line of particular structure that carries an element of lexicographic description. The examples of parameter contents are given below:

Example 1.
RR
bikini.
DUPL
Tb. biquini.

Example 2.
RR
poeta, tisa.
ETYM
Del lat. *poēta*, y este del gr. ποιητής *poiētḗs;* para la forma f., cf. fr. mediev. *poétisse*.
MORPHO
Para el f., u. t. la forma *poeta.*

Example 3.
RR
inmaculado, da.
ETYM
Del lat. *immaculātus.*
ORTHO
Escr. con may. inicial en acep. 2.

Example 4.
RR
ONG.
UNCRT
Sigla de *organización no gubernamental.*

The right part (P_0) is described by *MNGN* (meaning number), *REM* (block of labels), *DEF* (definition), *ED* (encyclopedic data), *COMM* (comment) and *IL* (illustration). All these parameters denote the information elements which compose the text of the right part.

Let us consider the structure of the label block (*REM*). The text of the block is subdivided sequentially into smaller fragments, each one representing a label of a certain type: *REM-GR* (grammar); *REM-PR* (pragmatics); *REM-ST* (stylistics); *REM-SF* (domain); *REM-REG* (geographic region); *REM-WHU* (where-used). Grammar remark *REM-GR* is mandatory remark, and it goes first and determines the part of speech the headword belongs to.

As a rule, lexical meaning in entry text is described by the component *DEF*. The comments *COMM* correlate with definitions. Each definition and each comment can

be accompanied by its illustrations *IL*. Each lexical meaning can be described by several *DEF*, *COMM* and *IL*. Let us demonstrate the decomposition of text into fragments for headword *cómico*:

Example 5.

MNGN
1.
REM
adj.
DEF
Que divierte y hace reír.
IL
Situación cómica.

MNGN
2.
REM
adj.
DEF
Perteneciente o relativo a la comedia.

MNGN
3.
REM
adj.
DEF
Dicho de un actor: Que representa papeles cómicos.
COM
U. t. c. s.

The implementation of conceptual model in a form of computer database will allow the selection of dictionary entries according to their profile. For example, the user will have options for selecting entries with a certain number of meaning; without illustrations; with or without comments; entries with two (or more) definitions; meaning with two (or more) comments etc.

6 Results and Discussions

6.1 Description of VLL DLE 23 Interface

VLL DLE 23 is available at https://services.ulif.org.ua:44359 and accessible using the user's log-in and password. Since the application is at the development stage, the interface language is Ukrainian but in final version English and Spanish will be also added. The main window (Fig. 2) consists of the following interface elements necessary to perform research works:

- Menu bar;

- Search panel;
- Word list panel;
- Entry text box.

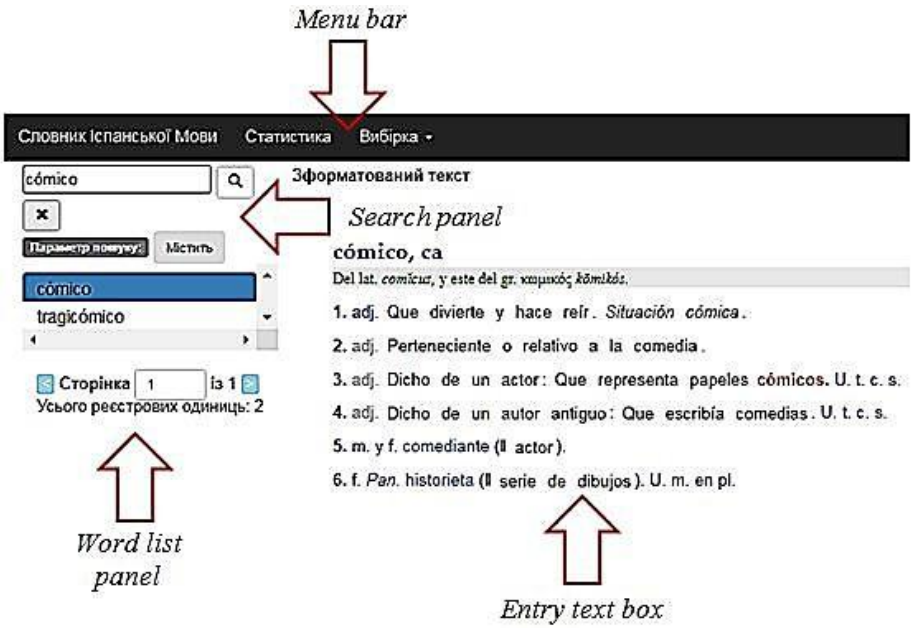

Fig. 2. Main window of VLL DLE 23

The headword list panel is composed of the list of alphabetically arranged lemmas and navigation bar to flick through the list. For convenience the headword list has been split into the pages each one consisting of 150 elements. The user can use "Forward" and "Back" buttons or enter a page number in text field to get to appropriate part of the list.

Entry text box displays dictionary entries in HTML format. The way of entry representation is the same as in the online version. The main window has also a text box (not shown in fig. 2) to view HTML text of the entry selected or copy text fragments for full-text search.

The interface of VLL DLE 23 allows the following modes to work with the dictionary:

- Headword list;
- Entry profile;
- Full-text search.

Headword List. According to the standards of Ukrainian Lingua-Information Fund, the list of the headwords should represent all the words even those which haven't been initially included by dictionary makers. In contrast to original DLE 23, the

headword list of VLL DLE 23 is completed by feminine forms and regional variants of the headwords. Regardless of the working mode, the information on the number of headwords is always displayed. The current version runs to 106323 units.

The headword can be selected either by clicking on it on the list, or entering a sequence of characters that match exactly match exactly a word in search. The search filters enhance quick search in cases when the user doesn't know the correct spelling of the headword. To enter diacritic symbols the virtual keyboard (fig. 3) can be also used. The symbol "–́" has been introduced to search for the morphemes bearing graphical accent. For example, the morphemes *-cóla*, *-éo* or *-fágo* must be typed as *–́cola*, *–́eo*, *–́fago*.

Fig. 3. Search field with virtual keyboard.

Entry Profile. This mode is good for making a sample of entries which satisfy the parameters of entry elements represented in DLE 23 conceptual model. In current version the choice is limited to headword and some definition & collocation parameters. In future all the entry elements will be available to work with through the interface.

The mode is activated by clicking the menu "Sample" after which the dialog box appears. The dialog box has two tabs "Headword" and "Entry". In first the user can choose headword parameters by which the entries are to be selected:

- Headword variants: lemma, masculine, feminine, regional variant, not defined.
- Headword structure: word, collocation, morpheme, not defined.
- Headword type: foreign word, abbreviation, acronym, not defined.
- Homonymy: yes (≥1) / no (0).

The second tab is intended for selecting the entries which correspond to definition / collocation parameters:

- Number of definitions: numerical value and additional options (>, ≥, = ≤, <).
- Number of collocations "Noun + Adj.": numerical value and additional options (>, ≥, = ≤, <).
- Number of collocations of other types: numerical value and additional options (>, ≥, = ≤, <).
- Number of cross-reference: numerical value and additional options (>, ≥, = ≤, <).

The entries are possible to be selected by headword and definition / collocation parameters at the same time. The figure 4 shows the sample of the entries containing homonymic polysemous morphemes. The sample corresponds to the following parameters:

1. Headword structure: morpheme;
2. Homonymy: ≥1;
3. Number of definitions: >1.

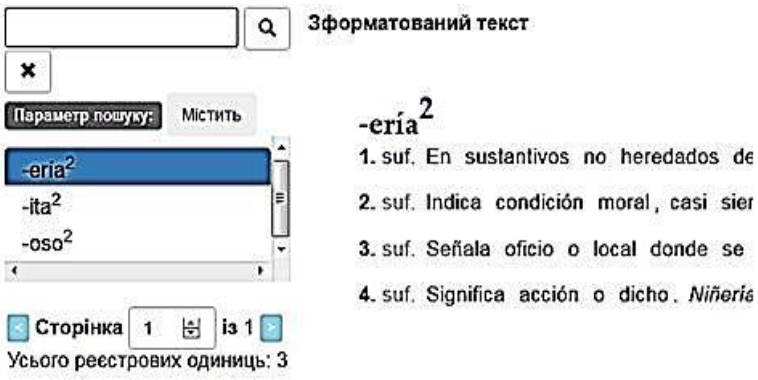

Fig. 4. Selection of homonymic polysemous morphemes.

Each of these parameters can be discarded by checking respective checkbox in “Sample” dialog box. As for homonymy, polysemy and collocations quantitative parameters are provided: amount (to be entered by user) and conditions to limit the amount. As it was mentioned above, statistical calculations are made for each sample. In this case the sample is made up by 3 elements.

Full-Text Search. This mode is required when it is necessary to select entries by specific meta-language elements of DLE 23: labels of different kind, characters that make up additional comments (“U. m. en sent. fig.”), meta-language markers (“Tb.”, “Voz”) etc. Furthermore, the text string for search may include both the text of a dictionary entry and the elements of HTML code (“<abbr title="Usado solo en infinitivo y en imperativo">”) taken from the text box in bottom of the main window. The figure 5 shows the way of selecting the verbs marked as “U. solo en infinit.” (Used only in infinitive).

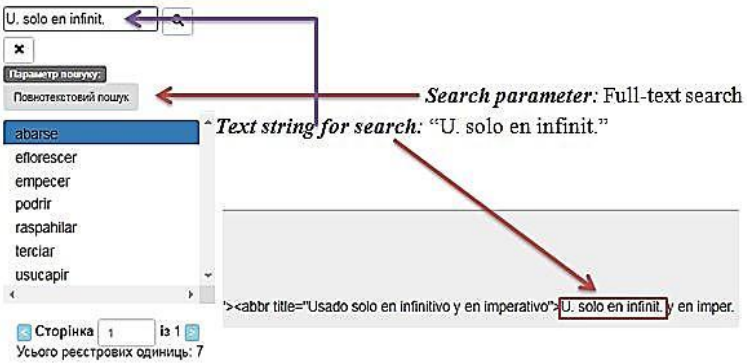

Fig. 5. List of the entries which contain the verbs marked as "Usado solo en infinitivo".

In next subsections of the paper we consider the application of the interface for different tasks: getting statistics data on language material, conducting linguistic researches and creating sub-dictionaries on the basis of DLE 23.

6.2 Statistics

In each mode, statistical information is generated for each sample by clicking menu "Statistics". A general view of statistics window is shown in fig.6.

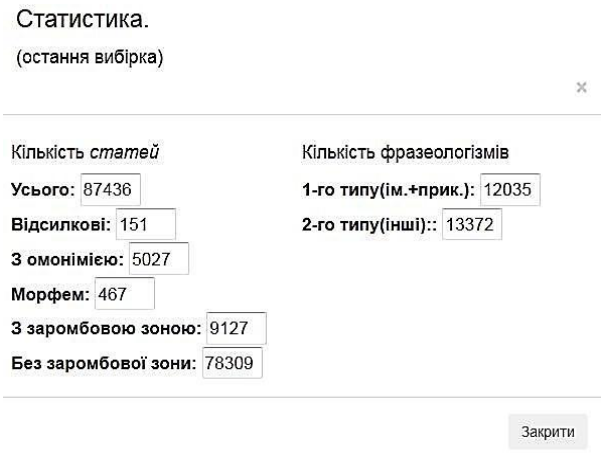

Fig. 6. General view of "Statistics" window.

For example, we can get quantitative characteristics of all language units composing DLE 23:

- Total number of entries, including referential entries: 87436;
- Referential entries: 151;
- Homonyms: 5027;

- Morphemes: 467;
- Entries with collocations: 9127;
- Entries without collocations: 78309;
- Collocations "Noun + Adjective": 12035;
- Collocations of other types: 13372.

6.3 Linguistic Researches

The interface of VLL DLE 23 allows conducting linguistic researches on the entire text of the dictionary. On their basis the user can make certain conclusions regarding lexical-semantic, etymological, grammatical and usage peculiarities of Spanish language units. The linguistic information to be drawn from DLE 23 text using the interface is shown in Table 4.

Table 3. Linguistic information to be get from DLE 23.

Aspect of research	**Linguistic information**
Semantics	Ability of words to have single or multiple meanings
	Semantic changes in word formation
	Semantic relationships, in particular synonymy and homonymy
	Semantic of derived words such as diminutives, augmentatives etc.
	Words belonging to a particular semantic field
Etymology	The role of other languages in the formation of Spanish vocabulary
	The way the unit from another language came to Spanish (directly or through intermediary language)
	Changes in the form and meaning of etymons in course of adaptation to Spanish
	Information about specific changes in etymons, e.g. apheresis, syncope, apocope etc.
Grammar	Grammar categories and grammar classes
	Conjugation model for verbs, availability of double or defective paradigms
	Syntactic functions of lemma, in particular prepositions, conjunctions
	Dependability of lexical meaning on grammatical meaning
	Ability of a lemma to form collocations
	Additional grammar characteristics of lemmas
Usage	The influence of social environment, domain, style and other factors on using lemmas

Let us show a concrete example of linguistic research by means of our interface. The task is to get a set of lemmas which are hyponyms to the word *embarcación* (ship,

vessel). To achieve this goal it will be necessary to build up a sample of DLE 23 entries in which lemmas are explained by indicating the broader term *embarcación* in their definitions. In this case both tools "Sample" and "Full-text search" will be required. So, the sequence of steps is as follows:

1. In the dialog box "Sample", tick the checkbox "Word" (tab "Headword parameters") and click the button "Make sample";
2. Select "Full-text search" in the combo box "Search parameters";
3. Enter the word "Embarcación" in the search box and then click the search button.

As a result, we get a sample of 109 entries, the lemmas of which are hyponyms to the word *embarcación*: *aljibe* (cistern), *almadía* (plank boat), *barca* (cockle boat), *barcón* (warship), *barquía* (rowboat), etc. In definitions we can outline components of lexical meaning by which the lemmas differ: 1) purpose; 2) design; 3) characteristics such as shape and size; 4) geographic region; and 5) time period. It should be noted that some definitions may represent several components. The results obtained, in our opinion, are possible to be used in compiling glossaries to lessons, vocabulary quizzes and exercises to memorize new words.

6.4 Sub-dictionaries

The samples created by means of VLL DLE 23 can be considered as sub-dictionaries to be built on the basis of DLE 23. The table 5 shows the sub-dictionaries with their volume and language units they describe. The volume has been calculated using "Statistics" tool.

Table 4. Sub-dictionaries built on the basis of DLE 23.

Sub-dictionary name	**Volume**	**Examples of headwords**
Dictionary of morphemes	586	-ante, -ar; centi-, circun-
Dictionary of homonyms	5027	chorizo[1], chorizo[2], za
Dictionary of Latin expressions	172	ab initio, vox populi
Dictionary of acronyms and abbreviations	130	ARN, ONG, laser, radar
Dictionary of foreign words	303	acid, amateur, body
Dictionary of foreign collocations	17	reality show, pop art
Dictionary of collocations "Noun + Adjective"	3284	agente comercial
Dictionary of set expressions	2943	ahora bien, ahora mismo
Dictionary of words of common gender	19011	abierto, ta; abuelo, la
Dictionary of monosemantic words	53663	abajera; chocón, na
Dictionary of polysemantic words	104304	abad, -desa; agua

By activating the parameters of headword and entry at the same time in the dialog box "Sample", the user can get a combined sub-dictionary. Using "Statistics" tool we can count all the units contained in the sub-dictionary.

For example, we want to make a sub-dictionary of monosemantic words which have the collocations “Noun + Adjective”. In the dialog box “Sample” the following options must be selected:

1. In the tab “Headword parameters”, tick the checkbox “Word” (“Headword structure” panel);
2. In the tab “Entry parameters” select the number of definitions “= 1”, number of collocations Noun + Adjective “≥ 1” and number of collocations of other types “= 0”.

The sub-dictionary has in total 509 monosemantic headwords, out of which 476 are non-homonymic and 43 are homonymic. The total amount of the collocations “noun + adjective” is 724.

7 Conclusions and Future Works

DLE 23, like other fundamental lexicons, is an exhaustive source of information on the lexical-grammatical and lexical-semantic properties of Spanish language units. Therefore, it can be useful not only to ordinary users as a reference book, but also to linguists as a means for studying the language. For convenient use of the dictionary for research works, a virtual lexicographic laboratory VLL DLE 23 providing access not only to the world list, but also to the text of dictionary entries was developed. Unlike the online version of DLE 23, the virtual lexicographic laboratory has the following advantages:

- Practically unlimited potential for integrating various linguistic facts in one object;
- Ability to reflect language dynamics;
- Possibility of selecting linguistic information from dictionary text by using sample parameters (in current version the number of parameters is limited to lemma parameters and some definition / collocation parameters);
- Availability of three working modes: headword list, dictionary profile and full-text search.

The virtual lexicographic laboratory provides the users with the tools necessary for studying grammatical, semantic, pragmatic, and other features of the Spanish linguage units. The current version of VLL DLE 23 allows conducting the following researches on the basis of DLE 23:

- Statistical calculations both on the entire dictionary and on a separate sample of dictionary entries;
- Studying Spanish vocabulary and dictionary entry texts to extract various linguistic facts recorded in DLE 23 (semantics, etymology, grammar, word usage);
- Building secondary lexicographic objects (or sub-dictionaries) that describe the specific lexical composition of the Spanish language.

Using VLE DLE 23 tools, it is possible to extract linguistic information, which may serve as a basis for compiling learner's dictionaries, preparing didactic materials and tests for those who study Spanish.

In final version of VLL DLE 23 the structural profile of dictionary entry will be determined by all the structural elements of the conceptual model. The user will be able to select the entries by indicating obligatory presence or absence of a structural element. Additionally the user will have the possibility of specifying specific content of the structural elements. The final stage of our elaboration will also include testing of the fully developed user's interface with representatives and real users.

References

1. Cadenaser.com.: La RAE regala ejemplares del diccionario en papel porque nadie los compra, https://cadenaser.com/ser/2018/07/04/cultura/1530712888_864828.html, last accessed 11.04.20.
2. Casares J.: Nuevo concepto del diccionario. Editorial CSIC, Madrid (1992).
3. Diccionario de la lengua española, https://dle.rae.es/, last accessed 2020/02/23.
4. Kupriianov, Y., Akopiants, N.: Developing Linguistic Research Tools for Virtual Lexicographic Laboratory of the Spanish Language Explanatory Dictionary. In: Lytvyn, V., Sharonova, N., Hamon, T., Cherednichenko, O., Grabar, N., Kowalska Styczen, A., Vysotska, V. (Eds.) Computational Linguistics and Intelligent Systems. Proc. 3rd Int. Conf. COLINS 2019, Vol. I: Main Conference. Kharkiv, Ukraine, April 18-19, 2019, pp. 43 – 52, CEUR-WS.org, http://ceur-ws.org/Vol-2362/, last accessed 2020/02/23.
5. Kupriianov, Y. Lexicographic system of the Spanish language: Phenomenology of integral description. Ukrainian Lingua-Information Fund, Kyiv (2018).
6. Lew, R: Online dictionary skills. In: Kosem, I., Kallas, J., Gantar, P., Krek, S., Langemets, M., Tuulik, M. (eds.): Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference, 17-19 October 2013, pp. 16–31, Tallinn, Estonia (2013).
7. Oxford English Dictionary, https://www.oed.com/, last accessed 2020/02/23.
8. Rae.es.: El nuevo diccionario académico será digital y más panhispánico, https://www.rae.es/noticias/el-nuevo-diccionario-academico-sera-digital-y-mas-panhispanico, last accessed 2020/02/23.
9. Ščerba, L.: Towards a General Theory of Lexicography. In: R. R. K. Hartmann (ed.), Lexicography. Critical Concepts, Vol. 3. pp. 11–50. Routledge, New York ([1940] 2003).
10. Shyrokov, V.: Computer lexicography, Kyiv (2011).
11. Shyrokov, V (Ed.): Computer linguistic studies: Proceedings of the Ukrainian Lingua-Information Fund NAS of Ukraine. Vol. 1: Research paradigm and basic language information structures. Ukrainian Lingua-Information Fund, Kyiv (2018).
12. Shyrokov, V (Ed.): Computer linguistic studies: Proceedings of the Ukrainian Lingua-Information Fund NAS of Ukraine. Vol. 5: Virtualization of linguistic technologies. Ukrainian Lingua-Information Fund, Kyiv (2018).
13. Vincze, O., Alonso Ramos, M.: Testing an electronic collocation dictionary interface: Diccionario de Colocaciones del Español. In: Kosem, I., Kallas, J., Gantar, P., Krek, S., Langemets, M., Tuulik, M. (eds.): Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference, 17-19 October 2013, pp. 328–337, Tallinn, Estonia (2013).