Iryna Ostapova/Volodymyr Shyrokov/Yevhen Kupriianov/
Mykyta Yablochkov

ETYMOLOGICAL DICTIONARY IN DIGITAL ENVIRONMENT

Abstract The digital environment represents a qualitatively new level of service for research work with linguistic information presented in dictionary form. And first of all, this applies to index systems. By dictionary indexing we mean a set of formalized rules and procedures, on the basis of which it is possible to obtain information about certain linguistic facts recorded in the dictionary. These rules are implemented in the form of user interfaces. However, one should take into account the fact that the effectiveness of automatic construction of index schemes for a digital dictionary is possible only in a sufficiently formalized environment. This article describes the method and technology of indexing the Etymological Dictionary of the Ukrainian Language (EDUL). For the language indexing of the dictionary, a special computer instrumental system (VLL – virtual lexicographic laboratory) was developed, and adapted to the structure of the EDUL and focused on the creation of indexes in automatic mode. The digital implementation of the EDUL made it possible to access the entire corpus of the dictionary text regardless of the time of publication of the corresponding volume and opened up opportunities for various digital interpretations of etymological information.

Keywords Ukrainian language; etymology; formal model; lexicographical system; etymological data base; index

1. Introduction

Etymological Dictionary of the Ukrainian Language in 7 volumes (hereinafter – EDUL) is a fundamental lexicographic work in the field of Ukrainian etymology (ESUM 1982–2012). The first volume of this dictionary was published in 1982, and the sixth in 2012. Currently, the text of the seventh volume is being formed, which is a multilingual index to the entire array of the dictionary. The text of the dictionary includes 26,165 dictionary entries, in the development of which 277 languages were used. For each of these languages, a separate index was built with the exact localizations of each word – the index unit. The total volume of the index is about 210,000 index elements (without the Ukrainian language). The dictionary is organized according to the nest principle. The total number of index elements in the nests (Ukrainian language) is about 230,000.

For the language indexing of the dictionary, a special computer instrumental system (Virtual Lexicographic Laboratory) was developed (Shyrokov 2018), and adapted to the structure of the EDUL and focused on the creation of indexes in automatic mode. The digital implementation of the EDUL made it possible to access the entire corpus of the dictionary text regardless of the publication time of the corresponding volume and opened up opportunities for various digital interpretations of etymological information.

To build the VLL, the following steps were performed:

1) Development of a formal model of the EDUL lexicographic system;
2) Preparation of a digital version of the EDUL text and identification of elements of its metalanguage marking elements of the structure of its lexicographic system;
3) Development of a database structure corresponding to the structure of the EDUL lexicographic system, taking into account the markers of its metalanguage;

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

4) Automatic conversion of the electronic text of the EDUL into a lexicographic database with a developed structure;
5) Creation of a tool package that supports user interfaces in on-line mode.

2. Etymological Dictionary of the Ukrainian Language (digital version)

2.1 Method

The digital transformation of lexicographic works requires some general theoretical framework to describe and represent the widest possible class of lexicographic objects. In our developments, we are based on the theory of lexicographic systems.

The dictionary is considered as a special type of information system – lexicographic system. It is an abstract language-information object focused on the implementation of a comprehensive information description of the lexical and grammatical structures of a particular language or set of languages (Shyrokov 2015).

The architecture of the system corresponds to the standard three-level architecture of ANSI/X3/SPARK information systems, according to which the conceptual, internal and external data levels are distinguished in the information system (Shyrokov 2014).

The types, structures, and formats of data representation, storage, and manipulation are defined internally.

At the external level, a set of procedures that allow the user to manipulate the data presented at the internal level is implemented.

The conceptual level of representation (conceptual model) is a symbolic, semantic model in which the ideas of various specialists about the subject area are integrated in an unambiguous, finite and non-contradictory way.

As a conceptual model, we use a lexicographic data model (Shyrokov 2018). Below we present it in a somewhat simplified form:

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

where D is the object (subject area) of modeling, in our case it is the Etymological dictionary of the Ukrainian language; $I_0(D) = \{x_i\}$ indicates the set of registry units of the dictionary (in the theory of lexicographic systems it is commonly called the set of *elementary information units*); $V(I_0(D))$ refers a set of descriptions (interpretations) of elementary information units, that is, texts of dictionary entries: (for dictionary $V(I_0(D)) = \{V(x_i)\}$ denotes matches the set of texts of dictionary entries, where $V(x_i)$ means a dictionary entry with a title word (head word) x_i; β denotes a set of structural elements abstracted as a result of the analysis of the dictionary text; $\sigma[\beta]$ is a structure that is generated by a β certain operator σ and represents a system of meaningful relations that reflect the semantics of the subject area under consideration; restriction $\sigma[\beta]$ on $V(x)$ generates a microstructure $\sigma(x)$ dictionary entry $V(x)$; $Red[V(I_0(D))]$ is a *recursive reduction* mechanism that makes it possible to consistently identify more and more subtle details of the structure of the lexicographic system.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

2.2 Structure of the dictionary entry

The conceptual model of the dictionary is based on the analysis of the printing version of the EDUL, that is, the typographic design, organization and structure of printed texts of dictionary entries are analyzed, which are interpreted as identifiers of the corresponding elements of lexicographic structures β and σ [β].

As a basic structural element, an *etymological class* was introduced, which is a block of linear text of a dictionary entry that describes certain genetic relationships of a registered Ukrainian word. Etymological classes are distinguished according to formal criteria: a structural unit is identified as an *etymological class* if unique sign sequences used as separators can be identified in the text of a dictionary entry. The following types of etymological classes were identified for EDUL: the *class of the head word* (*HEAD*), a *class of derivatives* (*DER*), a *class of Slavic correspondences* (*SLAV*), a *language class* (*LANG*), *bibliographic class* (*BIBL*), a *reference class* (*LINK*). Note that the phrases "etymological class" and "language class" are used only as names of structural elements in the formal model of a dictionary entry, and not as linguistic terms. Each of these classes has a unique text structure, which gave us the opportunity to build a procedure for identifying the type of each etymological class in a dictionary entry by formal features.

Let's give a brief description of each class.

The head word class contains the registry (head) word itself and its specific parameters. This class is unique and is necessarily included in the dictionary entry.

The following descriptions belong to the *language class*: a) reconstructed forms of the head word or their bases at different stages of the development of the Proto-Slavic language, presented in an anti-chronological order; b) etymologically related to the registered word, the words of other Indo-European languages, starting with the phonetic and word-formation forms closest to the Proto-Slavic; c) etymologically related to the head word, the words of Semitic-Hamitic or Ural-Altaic languages; d) the etymological relation of the word has not been established, for example, "the etymology is unclear". EDUL analysis showed that there are maximum of two such classes in a dictionary entry, but we do not limit their number in our model. The analysis of the EDUL showed that there are two such classes in a dictionary entry maximum, but we do not limit their number in our model. The dictionary entry must include at least one class of this type.

The class of the head word and the *language class* make up the minimal structure of the dictionary entry. Etymological classes of other types are optional.

The class of derivatives contains words related to the registered word of the Ukrainian language, that is, the closest etymologically meanings. There can be no more than one etymological class of this type in the text of a dictionary entry.

The class of Slavic correspondences contains the correspondences of the head word from all Slavic languages in which they are recorded. There can be no more than one structural unit of this type in a dictionary entry.

Bibliographic class is a block of text containing information about scientific works that consider the etymology of the corresponding Ukrainian word or related words of other languages. The number of structural units of this type is not limited.

The *class of links* includes those text blocks that describe links with other dictionary entries.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Let's illustrate what has been said by the example of a small, but quite representative from the point of view of the structure of a dictionary entry with the head word **вуж** (*grass snake*). The text is presented in an authentic printed form:

(1) **вуж,** *вужа́к, вужа́ка, вуженя́, вужи́ха* Я, [*вужо́вник*] «змійовик» (мін.) Ж, [*гужа́к*] ЛЧерк, *уж, ужа́ка,* [*вужа́чий*] Я, *вужи́ний,* [*вужо́вий*], [*вужува́тий*] Ж; — р. бр. *уж,* п. wąż, ч. слц. užovka, вл. нл. wuž, слн. vóž; — псл. *ǫžь; — споріднене з лит. angìs «змія», прус. angis, лат. anguis, двн. unc «тс.», сірл. escung «вуж, вугор» (букв. «водяна змія»); ie. *ang^{u}(h)i-, з яким пов'язується також *вуго́р.* — Фасмер IV 150—151; Holub—Kop. 405; Machek ESJČ 673; Топоров I 86—87. — Пор. **вуго́р**[1], **вуго́р**[2].

The distribution by class can be represented as follows:

HEAD: **вуж**

DER: вужа́к, вужа́ка, вуженя́, вужи́ха Я, [*вужо́вник*] «змійовик» (мін.) Ж, [*гужа́к*] ЛЧерк, *уж, ужа́ка,* [*вужа́чий*] Я, *вужи́ний,* [*вужо́вий*], [*вужува́тий*] Ж

SLAV: р. бр. *уж*, п. wąż, ч. слц. užovka, вл. нл. wuž, слн. vóž

LANG: псл. *ǫžь

LANG: споріднене з лит. angìs «змія», прус, angis, лат. anguis, двн. unc «тс.», сірл. escung «вуж, вугор» (букв. «водяна змія»); ie. *ang^{u}(h)i-, з яким пов'язується також *вуго́р*

BIBL: Фасмер IV 150—151

BIBL: Holub—Kop. 405

BIBL: Machek ESJČ 673

LINK: **вуго́р**[1]

LINK: **вуго́р**[2]

The linear sequence of the text blocks of the article can be schematically represented as follows (in curly brackets are sequences of characters that play the role of separators):

HEAD{,}DER{;—}*SLAV*{;—}*LANG*{;—}*LANG*{;—}
BIBL{;}*BIBL*{;}*BIBL*{;— Пор.}*LINK*{,}*LINK*

The selection of linear text blocks corresponds to the traditional division of the text of a dictionary entry into zones.

In the text of each etymological class, the connections of the registered word with certain words of other languages are established. We will call all these words, including head words, *etymons* (we are aware of the controversy of this term and use it for our model). When analyzing the texts of etymological classes, eight parameters were identified by which etymons are described: a *marker of linguistic affiliation* (P_L), a *remark to the marker of linguistic affiliation* (P_{RL}), a *symbolic representation of etymon* (P_A), *belonging to the dialect vocabulary* (P_D), a *homonymy marker* (P_O), *interpretation* (P_S), *remark* (P_R), *bibliography* (P_B). We have listed parameters in the order in which they usually appear in the text of the corresponding etymological class. Two parameters are required: P_L (*marker of language affiliation*) and P_A (the *symbolic representation of the etymon*). These two parameters ensure the uniqueness of each etymon of the dictionary entry: etymons with the same sign form may have different linguistic affiliation (for example, for a dictionary entry **уж** р. бр. *уж*; ч. слц. užovka; вл. нл. wuž), or etymons with the same language affiliation may have different sign forms. The

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

other parameters are optional. A formal procedure is defined for each parameter, which allows you to isolate the corresponding parameter from the text for each etymological class.

The set of parameters $\{P_L, P_{RL}, P_A, P_D, P_O, P_S, P_R, P_B\}$ we will call the *etymon structure* and denote as $ETYM(e_i)$, where *ei* is the corresponding etymon; index *i* is the ordinal number of the appearance of this etymon in the text of the etymological class. We believe that the order of the parameters in the etymon structure is not essential. The ordinal number of the etymon is significant and is recorded in the database.

Not all parameters are relevant for each etymological class. The text that we identify as an etymological class uses its own subset of parameters; not every etymon is required to be described by a complete set of parameters. However, in order to achieve structural uniformity, one type of etymon structure is constructed for each class; if a certain parameter is not involved or cannot be distinguished by formal features, then an empty line of text corresponds to its meaning. The etymon structure is constructed only if it was possible to isolate P_A. Formally, we believe that at least one etymological structure corresponds to each etymological class. If a language class does not have any etymon (or it was not possible to identify it by a formal procedure), then we consider it a degenerate etymological class and an empty etymon structure corresponds to it. An example of a dictionary entry with such a language class:

(2) **[андріяк]** «опій»; — походження неясне.

Here is an example of etymon structures (only basic parameters) for the etymological classes *SLAV* and *LANG*.

(1) р. бр. уж, п. wąż, ч. слц. užovka, вл. нл. wuž, слн. vóž

(2) $ETYM\,(e_1)$ = {*PL* = <р.>, *PA* = <уж>}

(3) $ETYM\,(e_2)$ = {*PL* = <бр.>, *PA* = <уж>}

(4) $ETYM\,(e_3)$ = {*PL* = <п.>, *PA* = <wąż>}

(5) $ETYM\,(e_4)$ = {*PL* = <ч.>, *PA* = <užovka>}

(6) $ETYM\,(e_5)$ = {*PL* = <слц.>, *PA* = <wuž>}

(7) $ETYM\,(e_6)$ = {*PL* = <вл.>, *PA* = <wuž>}

(8) $ETYM\,(e_7)$ = {*PL* = <слц.>, *PA* = <vóž>}

(9) $ETYM\,(e_8)$ = {*PL* = <вл.>, *PA* = <wuž>}

(10) $ETYM\,(e_9)$ = {*PL* = <нл.>, *PA* = <wuž>}

(11) $ETYM\,(e_{10})$ = {*PL*= <нл.>, *PA* = <wuž>}

(12) псл. *ǫžь

(13) $ETYM\,(e_1)$ = {PL = <псл.>, PA = <*ǫžь>}

2.3 Multilingual index

In a printed dictionary, the text is organized in such a way that the head word has a font emphasis and is the first word of the dictionary entry, which ensures its structural significance by the capabilities of the printed text. In the proposed model, the structure-forming parameters of the ethymon structure are mandatory: entry into the dictionary is possible for any language and for any word in the alphabet of this language.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Each index element ind_el_k (k = 1, 2, … K; K is the number of elements in the corresponding index) has the following structure:

$ind\_el_k \equiv \{e_k, lang(e_k), loc(e_k)\}$, where e_k — etymon, $lang(e_k)$ – marker of language affiliation, $loc(e_k)$ — localization of the etymon in the dictionary text.

For the printed index of the EDUL, the following form of localization of the etymon is proposed: volume number, page number (on which the etymon is printed), the head word of the corresponding dictionary entry. The numbers of volumes and pages are a tribute to tradition and binding to the printed version of the dictionary, individual volumes of which were published with a significant time interval. Before conversion to the database, the text of each article was linked to its printed original, the volume and page numbers were recorded in the corresponding database fields. Redundancy of localization parameters can be explained by the desire to combine the approaches of the digital and printed versions: the head word in the printed version is analogous to the ID of the dictionary entry; the page number is useful when the dictionary entry occupies several pages of text (this is typical for verbs). In the digital version of EDUL, the etymon is localized up to the ordinal number of the word in the etymological class string.

The task of constructing language indexes is assigned to the instrumental system. A separate index is formed for each of the 277 languages recorded in the dictionary. The index is edited in two stages: 1) first, these word structures are edited in the database at the level of a dictionary entry; 2) a file of index elements is formed on request to the database, alphabetized by head words for a given language and edited in .doc format. After editing, the file is ready for re-conversion to the database. For the publishing system, files are processed additionally by two programs: 1) the signs that are used for etymons are inventoried; 2) an alphabet is formed and index elements are ordered according to this alphabet.

There is a fragment of the indexes for the languages that are used in the dictionary entry (the names of the languages are given in the order of their appearance in the text of the dictionary entry):

Ukrainian language
вуж 1, 437 **вуж**
вужа́к 1, 437 **вуж**
вужа́ка 1, 437 **вуж**
вуженя́ 1, 437 **вуж**
вужи́ха 1, 437 **вуж**
[вужо́вник] 1, 437 **вуж**
[гужа́к] 1, 437 **вуж**
уж 1, 437 **вуж**
ужа́ка 1, 437 **вуж**
[вужа́чий] 1, 437 **вуж**
вужи́ний 1, 437 **вуж**
[вужо́вий] 1, 437 **вуж**
[вужува́тий] 1, 437 **вуж**
…

Russian language
уж 1, 437 **вуж**
Belarusian language
уж 1, 437 **вуж**
Polish language
1, 437 **вуж**
Czech language
užovka 1, 437 **вуж**
Slovak language
užovka 1, 437 **вуж**
Upper Lusatian language
wuž 1, 437 **вуж**
Lower Lusatian language
wuž 1, 437 **вуж**
Slovenian language
vóž 1, 437 **вуж**

Proto-Slavic language
*ǫžь 1, 437 **вуж**
Lithuanian language
angìs 1, 437 **вуж**
Prussian language
angis 1, 437 **вуж**
Latin language
anguis 1, 437 **вуж**
Old High German language
unc 1, 437 **вуж**
Middle Irish language
escung 1, 437 **вуж**
Indo-European language
*ang^{u}(h)i- 1, 437 **вуж**
…

The * sign is used for reconstructed words; we include it in the alphabet.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

2.4 Representation of the dictionary text in the structure of the lexicographic database

The current digital version of the dictionary uses a relational database of data. The entire corpus of dictionary entries is organized in 6 tabs. The tables **uketym_etym_classes**, **uketym_biblography**, **uketym_links** contain the texts of etymological classes from which the texts of dictionary entries are formed. In the **uketym_etymons** table, the parameters of all the explicated etymon structures are shown. The **uketym_languages_all** table contains information about the dictionary languages. The **ukretym_heads table** organizes the text of the dictionary entry. In the **uketym_etym_classes table**, the texts of etymological classes of dictionary entries (*HEAD*, *DER*, *SLAV*, *LANG*) are saved. All inter-article dictionary links are organized in **uketym_links**, all bibliographic links are in the **uketym_bibliography table**. The **uketym_language_all** table contains information about languages used in the dictionary. Based on this table, all user language registries are formed. The selection of an index array for a given language registry is performed from the **uketym_etymons table**.

The parsing of the dictionary text and the conversion of the text to the database was performed by a single program without the formation of intermediate files (in the future we abandoned this approach).

To perform parsing, the texts of dictionary entries of all volumes were converted into HTML format. The volumes of the dictionary were prepared for printing by various publishing technologies. The first three volumes are in monotype, pre-computer technology. Therefore, the printed texts were scanned and recognized using the ABBYY FineReader program, and then the texts were proofread. The last three volumes were prepared by means of a computer publishing system (MS Word was used). The sign system of the entire Dictionary text has been unified according to UNICODE 3.0 encoding. This made it possible to carry out an inventory of the alphabet symbols to represent the etymons of each language. To link between the printed and digital versions of the dictionary, each dictionary entry was marked (manually) as follows: volume number, page number at the beginning of the text of the dictionary entry, page number at the end of the text of the dictionary entry.

2.5 User interfaces

Basic VLL (virtual lexicographic laboratory) functions:

1) traditional entry by the head word and dictionary entry text visualization according to its structure (see fig. 1);
2) editing of any structural element of a dictionary entry;
3) building a dictionary entry of a given structure;
4) automatic indexing for each EDUL language (or a specified set of languages) (see fig. 2, fig. 3).

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

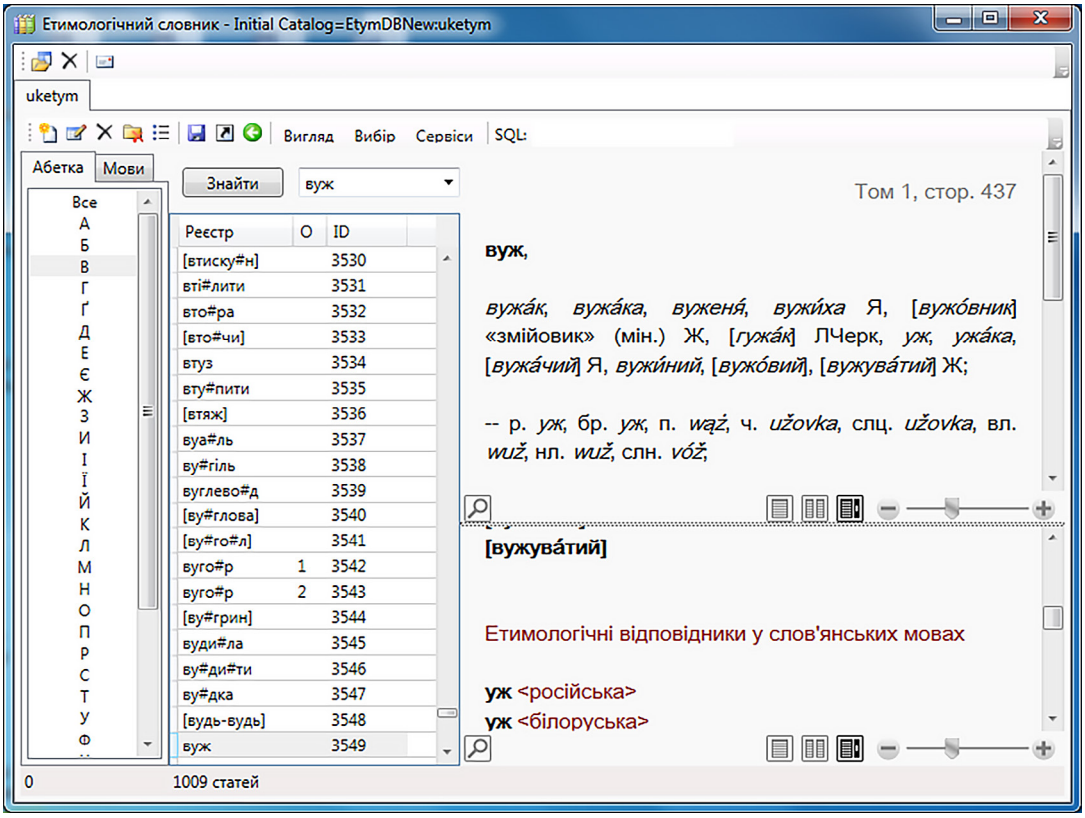

Fig. 1: Main window (article "**уж**")

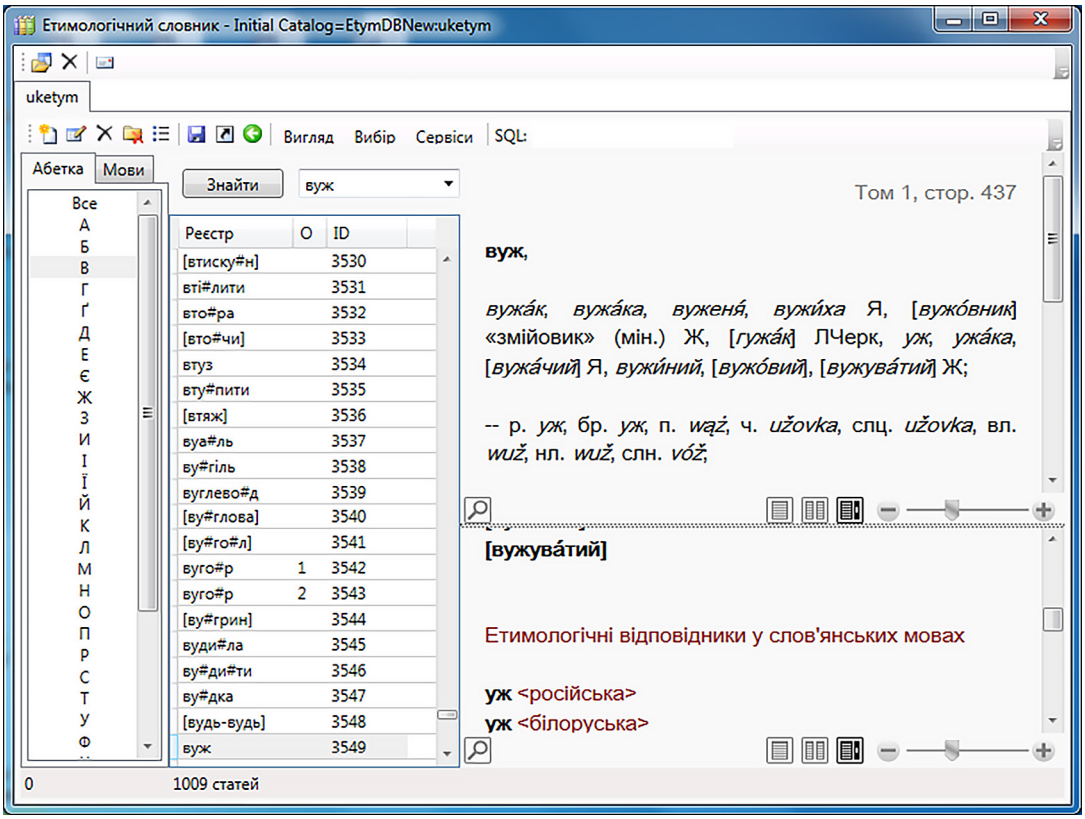

Fig. 2: Language index for a given language list (head words)

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

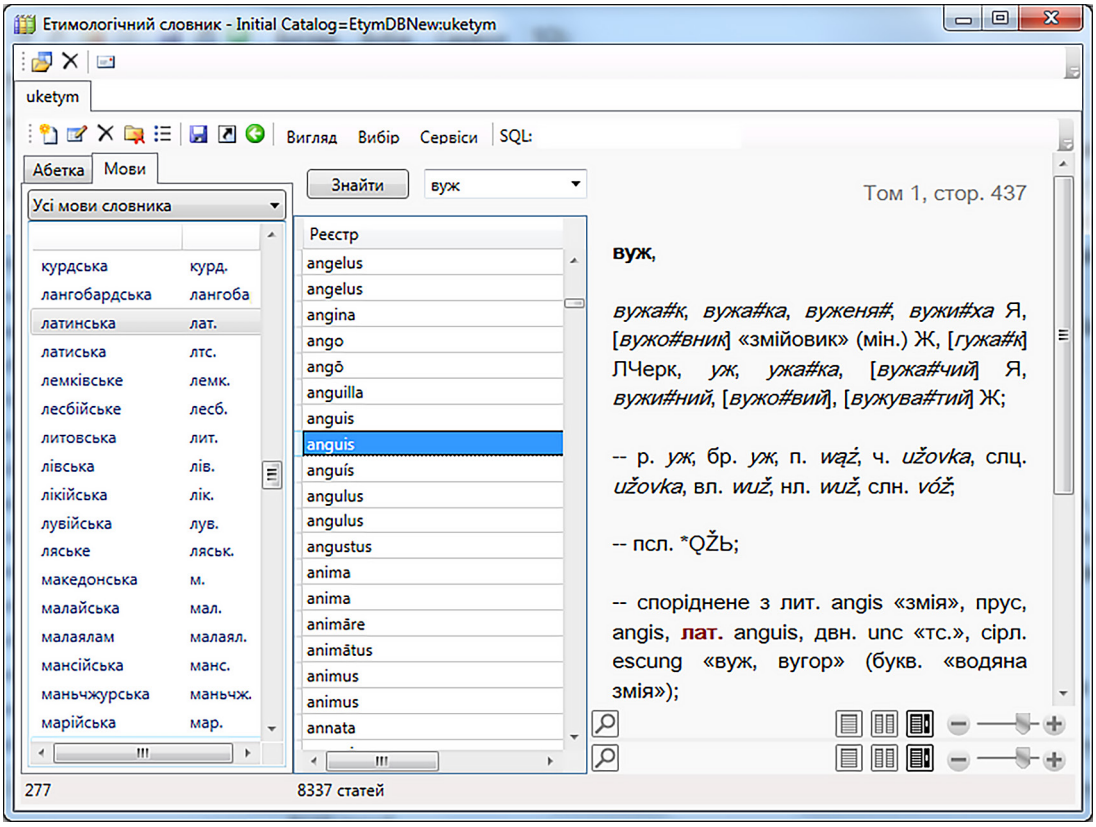

Fig. 3: Language index for a given language list (etymons)

3. Conclusion

In modern conditions, the sociologization of any lexicographic product is largely determined by its adaptation to functioning in the digital environment. This is true not only in relation to the mass consumer, but also in relation to professional communities. Such fundamental lexicographic projects as Dictionaries of national languages are switching (and some have already switched) to non-stop functioning mode: dictionary entries are updated and the registry is updated on a single lexicographic database with on-line user access.

Like most large dictionaries, etymological dictionaries also have their own representation in the digital environment. However, as a rule, they are presented in the format of machine-readable texts, which significantly limit their scalability and potential for research work.

EDUL has been created for half a century. The printed version of the index is understood more as the completion of a printed project, a tribute to tradition. Given more than modest print runs of the last three volumes (and the first three have already become a bibliographic rarity), the integrity of the dictionary and its sociologization can only be ensured by its digital version. To a large extent, the interest in the dictionary (in our understanding) will be provided by the development of the user interface, focused primarily on research tasks.

The new dictionary of the Ukrainian language (like its predecessor) does not have an etymological reference in its structure (SUM 2010–2021). The new edition of the dictionary is formed in the Digital Writing System mode and each new volume is delivered to users in the format of a printed book and a website on the Internet. Therefore, the integration of these two dictionaries is a useful technological task.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

References

ESUM (1982–2012): Etymologichnyĭ slovnyk ukraïskoĭ movy. V 7 t. (Vol. 1–6). Kyïv.

Shyrokov, V./Ostapova, I./Yakymenko, K. (2014): Digital lexicographical systems and traditional paper dictionaries (from traditional paper dictionaries to digital lexicographical systems). In: Cognitive Studies/Études cognitives 15, pp. 193–210.

Shyrokov, V./Ostapova, I. (2015): Indexing the etymological lexicographic systems. In: Cognitive Studies/Études cognitives 14, pp. 1–11.

Shyrokov, V (ed.) (2018): Computer linguistic studies: proceedings of the Ukrainian Lingua Information Fund NAS of Ukraine. Vol. V: Virtualization of linguistic technologies. https://movoznavstvo.org.ua/files/Ling_inf_studio_TOM_5_umif_B5.pdf (last access: 24-03-2022).

SUM (2010–2022): Slovnyk ukraïskoĭ movy v 20 t. Vol. 1–12. Ukrainian Lingua Information Fund NAS of Ukraine. Accesses at: https://services.ulif.org.ua/expl/, sum20ua.com (last access: 24-03-2022).

Trap-Jensen, L. (2018): Lexicography between NLP and linguistics: aspect of theory and practice. In: Čibej, J./ Gorjanc, V./ Kosem, I./Krek, S. (eds.): Lexicography in lobal Contexts. Proceedings of the 18th EURALEX International Congress 2018, 17–21 July 2018, Ljubljana. Ljubljana, pp. 25–38.

Contact information

Iryna Ostapova
Ukrainian Lingua-Information Fund of National Academy of Sciences of Ukraine
irinaostapova@gmail.com

Volodymyr Shyrokov
Ukrainian Lingua-Information Fund of National Academy of Sciences of Ukraine
Vshirokov48@gmail.com

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

Mykyta Yablochkov
Ukrainian Lingua-Information Fund of National Academy of Sciences of Ukraine
gezartos@gmail.com

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.