DIGITAL LEXICOGRAPHIC SYSTEMS
AND TRADITIONAL PAPER DICTIONARIES
(FROM TRADITIONAL PAPER DICTIONARIES
TO DIGITAL LEXICOGRAPHIC SYSTEMS)

Abstract

Main problems of modern lexicography are under analysis. The theory of lexicographical systems is presented as well as its applications for the description of the structure of lexicographical systems for new digital Ukrainian Explanatory Dictionary and Etymological Dictionary of Ukrainian Language. The concept of virtual lexicographical laboratories is presented. The implementation of two virtual lexicographical laboratories is given (the first for Ukrainian Explanatory Dictionary and next for Etymological Dictionary of Ukrainian Language).

Keywords: digital lexicography; lexicographical systems theory; virtual labs

Recently, due to the rapid development of a new cognitive-communicative environment and strengthening of the status of the natural language in lexicography, four important problems were actualized.

1. The first problem is updating of basic lexicon. Usually, large lexicographical projects are made by teams of highly skilled linguists. And the fundamental dictionaries become standard lexicographic objects. They represent examples of linguistic thought of the corresponding historical period. Such lexicons reflect linguistic consciousness of the people of a certain era. Lexicographical structures, examples of representing semantic phenomena, examples of solving linguistic and cognitive tasks with lexicographic means we rate as invariant values of the dictionary works. It is reasonable to make a full use of the most valuable things in the lexicographic works of such type. This leads to the problem of updating the fundamental lexicographic works, bringing them into the line with new realities, modernization and modification of their content and structure. To do this, you must define the linguistic and cognitive principles of formation of the new vocabu-
lary “over” the old one, develop the technology and organize the implementation of
the lexicographic project. In this case technological problems are moved into the
foreground. And this is connected with a significant decreasing of the time, avail-
able for the modernization of the lexicon, and the necessity to represent results
of work in different “nonisomorphic” to each other forms (a traditional book form,
a laser disk, an Internet site, etc.).

The second problem relates to the need of converting the printed lexicographic
heritage into digital form, which requires the creation of a special technology. Scan-
ning is widely used to convert text from paper form to a digital form. Scanned text
is processed by optical character recognition software. As a result of this procedure
we obtain a text in digital format that can be indexed. And indexed text can be
used in search engines. But this is completely insufficient for dictionaries. The pro-
cedure of recognition of scanned text has its own peculiarities. Meta-language of
many dictionaries contains nonstandard symbols that are difficult to identify. First
of all, symbols which are not the part of natural languages alphabets. Besides, it
is necessary to recognize and save graphic markup of the dictionary entry (fonts,
paragraphs, etc.) in the digital format. Present OCR software is far from perfect
and makes a lot of mistakes in the process of their applying. Authenticity of the
recognized text is an undoubted requirement for the dictionary. Finally, work with
dictionaries that were converted into computer text formats is extremely inefficient.
Texts must be converted into formats of lexicographic databases. For this purpose,
the technology of parsing of vocabulary texts is developing. This is special task,
unknown to classical lexicography.

The third problem — is usage of a large dictionary arrays for forming a linguistic
component systems of knowledge representation (such as ontologies). This requires
to identify hidden semantic structure in the traditional vocabulary texts. On the
other hand, onto- and concept- graphic problems stimulates active and substantive
integration of lexicography and modern knowledge industry.

Finally, the fourth problem is related to the overcoming the globalization era’s
“Babylon syndrome”, i.e. a solution of the multilingualism problem. This problem
requires powerful and flexible lexicographic tools for its solving. Primarily, we use
computer tools of multilingual lexicographic representation in remote mode. This
technology should provide a creative interaction between geographically allocated
international lexicographers’ teams.

2. These problems, in our opinion, stimulate the revision of the basic assumptions
of classical lexicography. The need to formulate some new principles, regulations,
practices, methods, etc. is realized.

Primarily the modeling object should be revised. In the digital lexicography fo-
cus has shifted from the describing of units and language relations to the modeling of
linguistic competence. It means that in the digital lexicographic systems models of
language understanding, subjected to lexicographic representation, must function.
For example, if we are dealing with inflection of inflected language, the correspond-
ing digital system must, at least, “to be able” to make word’s decomposition to the
variable and invariable part, and the variable part identify as part of the descrip-
tion of a certain paradigmatic class (which, in turn, implies a series of grammatical
identification’s operations), and then synthesize the inflectional paradigm, with all its details provided for the model. More advanced systems can equip oneself with smarter functions for the completion and development of the inflectional system, adding, modifying and editing of paradigmatic classes set, advanced morphological, phonemic, semantic and etymological analysis of the inflectional system.

Generalizing the above, we come to the idea of an explication of the system relations of the language (usually implicit) by means of digital lexicography. In cognitive terms these properties of such systems are the most interesting, as the most relevant for the intellectual tasks of natural language processing. Of course, that marked modus of digital lexicography leads to the necessity to apply to the classic problems of linguistics in order to understand how they are reformulated due to the new challenges. Let’s start with the fact that, in our opinion, the concept of the language systemacy should be reconsidered. We use term “lexicographic system” within the meaning of the theory of lexicographic systems, involving rather deep insight into the information properties of the language. The exact definition of lexicographic system will be described below, and further presentation will be based on this definition.

It is known, that vocabulary and grammar are basic methods of language description. We think that relationship between the grammatical and lexicographical descriptions should be explored further. According to the classical canons, vocabulary and grammar are being in some opposition, and at the same time being in relation of complementarity in a sense, close to Niels Bohr’s concept. The vocabulary, par excellence, is specialized in the description of the repertoire of language units, while the grammar — in the description of the rules. We believe, that as information systems, they present declarative (reflexive) and procedural (operational) description of the object domain, respectively. At the same time considering the presence of “grammar” elements in the vocabulary and “vocabulary” elements in the grammar, we acknowledge the potency to unite these two basic ways of describing the linguistic system, which is reflected in Apresyan’s concept of integral description of language, as well as in our formalism of vocabulary and grammar integration (Shyrokov, 2008, pp.25–27; 2011, pp.335–337).

Finally, the last introductory remark touches the fundamental difference between the scientific description of language and what we are seeing in other sciences. Indeed, other sciences describe their Phenomena with the help of Language that is Substance, which differs from its Phenomena. But linguistics, which Phenomenon belongs to Language, is doomed to describe Language by means of itself. In our opinion, implementation of a “meta-language” concept represents nothing more but an attempt to save the situation. Such a course of action reminds of Baron Munchausen’s efforts to pull himself out of the mire by his hair. That’s why, the explanatory mode of linguistics we, with some irony, will interpret as

---

1We believe that the language systemacy is revealed in the relations within the triad “structure – substance – subject”, to which a special work will be dedicated.


3Complementarity according to Niels Bohr.

a “principle of Munchausen”.

3. We intend to use said above as general conceptual guidelines for the further description. Our attention will be focused on the digital interpretation of the lexicography, which provides a whole range of lexicographic objects, unusual for its paper version. Among them we mark: user dictionaries (both local and online ones); resident dictionaries used in the circuits of word processing; instrumental dictionary systems: used for a) dictionary creation (formation); b) interactive texts labeling. Furthermore, we should mention: heterogeneity aspects of dictionary structures that adapt different types of information (text, audio, video, etc.); processes of automatic dictionary generation from texts etc.

For a unified description of all mentioned objects, it is advisable to unite them in a single conceptual framework. In our opinion, a theory of lexicographic systems allows to do this. This theory provides a universal definition of all the objects that we perceive as dictionaries. At the same time, it clarifies the mechanisms that stimulate language substance (and — wider — substance defined by information) to take the form of dictionary. Further we will move to the concise presentation of the theory of lexicographic systems.

The theory is based on a phenomenological principle that common to the structure and behavior of any system in which there are information processes. Its essence is as follows.

The common feature of all of the processes of information exchange is conversion of information from one form to another. All modern scientific theories quite definitely confirm fundamentally discrete (“quantum”) nature of information processes of interaction and exchange. This implies fundamental discretization of processes of subjective perception and interpretation (description) of reality. This discretization has at least one common feature for all known processes: in the evolution of system of any nature a certain subsystem of relatively stable discrete entities (“subsystem of order”) is induced in its structure. These entities will be called basic information units of the system. All the phenomena of the system are represented by combinations of basic information units organized in a certain way.

This subsystem has properties, which in some way related to the properties of the lexical system of natural language: it “generates” in its structure something like thesaurus and grammar, with such properties of these structures as system of signs, meaning, polymorphism, and etc.; it has both “expression domain” and “content domain”.

Sets of basic information units have a property of “substantiality”. Characteristics of these sets are relatively stable.

Described set of phenomena is the content of lexicographic effect. It can be said that during studying any subject domain specialists actually examine lexicographic effects which occur in them. Thus lexicographic effect has not only a phenomenological component, but also a methodological. It has a certain “operationality potential” that during the process of simulation of various systems stimulates to establish and determine the appropriate set of basic information units; considers, specifies and represents their properties. In this review the concept of the lexicographic effect acts as a method of data abstraction.
Lexicographic effect is expressed in the representation of a continual universum by means of discrete sets. The possibility of its formalization via “Skolem paradox”, in the theory of models known as Löwenheim-Skolem theorem (Pauza, 2001), seems attractive. Theorem states that under a certain restrictions, there is a certain isomorphism between uncountable and countable sets; in some ways, the potential infinity can have finite interpretation, per se a finite model of infinity is constructed. Löwenheim-Skolem theorem states that every consistent theory of the first order, which has an uncountable model, also has a countable model. This means that there is a representation of the infinite description of the object through a finite one, which contains all the information about the infinite object. Thus Löwenheim-Skolem theorem plays the role of a formal correlate of lexicographic effect (Golovko & Shyrokov, 2006, pp.201–202).

In what follows, the set of basic information units of system D, induced due to the development of lexicographic effect Q, denote by IQ(D) or simply I(D), if the reference to the type of lexicographic effect is insignificant.

Structure and architecture of Lexicographic Systems
Lexicographic systems (the parallel reducter therm L-systems is also used) we define as the basic constructive of information architecture, where one or more lexicographic effects are actualized. Lexicographic systems represent rather common type of formalized structures such as data models, formal grammars, formal systems, canonical calculus in finite alphabets etc.

We should mention that in science and technology special cases (or implementations) of lexicographic systems have been operated for a very long time, and they are: a large number of different information systems, databases and knowledge-base, which include all traditional dictionaries and computer dictionary systems. According to stated above we consider lexicographic system as a special information (semiotic and semantic) structure in which a certain lexicographic effect (or a set of lexicographic effects) is registered.

To construct a useful scheme of modelling of mentioned phenomena, we should define a set of information form-factors that specify structural elements of L-systems, which leads to the construction of a lexicographic data model. In accordance with the information interpretation of the perception processes let’s determine the results of the reception of class of basic information units (BIU) I^Q(D) by the subject S in regard to lexicographic effect Q as a certain set V(I^Q(D)) — a set of descriptions of units, which belong to the class I^Q(D); we consider that this set is the result of the process:

\[ S : I^Q(D) \rightarrow V(I^Q(D)) \]  

(1)

that is why, each element \( x \in I^Q(D) \) is explicitly (within a model) determined by its description \( V(x) \) as an element of the set \( V(I^Q(D)) \): \( V(x) \in V(I^Q(D)); Sx = V(x) \).

So, it is logical to assume that \( V(I^Q(D)) \) is represented as a union:

\[ V(I^Q(D)) = \bigcup_{x \in I^Q(D)} V(x). \]  

(2)

At this, each \( V(x) \) is represented as a word (text) in a certain finite alphabet \( A = \{a_1, a_2, ..., a_n\} \), i.e. a finite sequence of symbols (finite chains) from \( A \). Then
words in the alphabet $A$ we call $A$-words. For example, if we consider the Ukrainian language dictionary (ULD), its alphabet $A$ includes the following elements: an ordinary alphabet of the Ukrainian language (capital and lowercase letters); punctuation; Arabic numerals; Roman numerals; space and paragraphs symbols; special symbols ($//, \Delta, \blacklozenge, \bullet, ...$); font types, and others. Thus, the description $V(x)$ of any basic information unit $x$, $x \in I^Q(D)$, is represented by the $A$-word:

$$V(x) = v_1(x)v_2(x)...v_k(x), v_i(x) \in A, i = 1, 2, ..., k(x), k(x) \geq 1,$$

wherein each symbol $v_i(x)$ is taken from the alphabet $A$.

Let us define the concept of an elementary lexicographic data model and elementary lexicographic system.

Elementary lexicographic data model is formally defined as the seven elements:

$\{D; S; Q; I^Q(D), V(I^Q(D)), \beta, \sigma[\beta]\}$,

where $D$ is a fragment of reality that have to be lexicographically represented; $S$ — subject, that carrying out analysis of $D$; $Q$ — lexicographic effect, observed by the subject $S$ in the $D$, as a result of which a complex of basic information units (entities) $I^Q(D)$ is forming. In the dictionary this complex corresponds to the register; $V(I^Q(D))$ — is amount of descriptions $I^Q(D)$, which, as it was noted above, is formally determined by the set of words (texts), presented by linear string of symbols from a finite alphabet $A$:

$$S : I^Q(D) \to V(I^Q(D)); \forall x \in I^Q(D)S(x) = V(x).$$

Elements $\beta$ and $\sigma[\beta]$ determine the structure of L-system $V(I^Q(D))$ and they are specified for each $x \in I^Q(D)$. Therefore, the expressions $\beta(x)$ and $\sigma[\beta(x)]$ define the structure of the description $V(x)$; and the structure of the dictionary entry with the registry unit $x$ is identical to them. Elements $\beta$ are “elementary” structural elements of the dictionary (words, abbreviations, labels, remarks, numeration...) and, as a rule, are given explicitly. Formally it implements as follows. For each $h \in I^Q(D)$ a set $\beta(x)$ of “subwords” are formed from $V(x)$ with following properties:

1. $x \in \beta(x)$
2. The whole $V(x) \in \beta(x)$
3. A rule of forming the elements $\beta(x)$ is the same for every $V(x)$, i.e. for all $x \in I^Q(D)$.

Let’s assume that $\beta$ consists of elements $\beta_1, \beta_2, ..., \beta_q$. Then construct a table:

<table>
<thead>
<tr>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\ldots$</th>
<th>$\beta_q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1(x_1)$</td>
<td>$\beta_2(x_1)$</td>
<td>$\ldots$</td>
<td>$\beta_q(x_1)$</td>
</tr>
<tr>
<td>$\beta_1(x_2)$</td>
<td>$\beta_2(x_2)$</td>
<td>$\ldots$</td>
<td>$\beta_q(x_2)$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$\beta_1(x_M)$</td>
<td>$\beta_2(x_M)$</td>
<td>$\ldots$</td>
<td>$\beta_q(x_M)$</td>
</tr>
</tbody>
</table>
Elements $\beta_1, \beta_2, \ldots, \beta_q$ can be interpreted as the attribute names of some of the relational model, and elements $\beta_i(x_j), i = 1, 2, \ldots, q; j = 1, 2, \ldots, M,$ — as domains of corresponding attributes. Such representation allows to interpret the lexicographic model in terms of the relational model. More aggregated structures $\sigma[\beta]$ are constructed from the structures $\beta$. In the language of the relational model $\sigma[\beta]$ can be interpreted as relations of the model (or Codd’s algebra). In particular, left and right sides of the dictionary entries, along with special relations that define the semantics of the dictionary, defined from the elements of $\sigma[\beta]$. Many (if not most) of the structures $\sigma[\beta]$ in dictionaries are usually defined implicitly.

There is a non-optional element in the $\sigma[\beta]$ structure, it is a basic framework relation in each dictionary, namely, the relation “form – content”. Due to this, in the structure of each dictionary entry and dictionary $V(I^Q(D))$ there are two parts: “left” $\Lambda(I^Q(D))$ and “right” $P(I^Q(D))$, and they are responsible for the formal and substantial properties of lexicographically represented units, respectively. Thus, we come to the canonical format of the presentation of lexicographic model, which is represented by the following commutative diagram:

\[
\begin{array}{c}
V(I^Q(D)) \\
\downarrow F \quad \downarrow C \\
\Lambda(I^Q(D)) \quad \downarrow H \quad P(I^Q(D))
\end{array}
\]

(7)

where commutativity condition is defined by the relation $H \cdot F = C$. Macrostructures are induced on $\Lambda(I^Q(D))$ and $P(I^Q(D))$:

$F \sigma[\beta] = \Lambda[\beta]$ and $C \sigma[\beta] = \rho[\beta]$

(8)

and corresponding microstructures:

$\Lambda[\beta]|_{V(x)} \equiv \Lambda(x); \rho[\beta]|_{V(x)} \equiv \rho(x)$

(9)

As restrictions $\Lambda[\beta]$ and $\rho[\beta]$ on $V(x)$.

Considering not an optional decomposition (7), it is logical to assume that mapping $F$, $C$ and $H$ are the basic elements of the structure of elementary lexicographic data model, which takes the following form:

$\{D; S; Q; I^Q(D), V(I^Q(D)), \beta, \sigma[\beta], F, C, H\}$

(10)

and $FV(I^Q(D)) = \Lambda(I^Q(D)); CV(I^Q(D)) = P(I^Q(D)); H \cdot F = C$.

Lexicographic model is transformed into a lexicographic system if it is implemented in a particular information architecture $\Sigma$. As such, we use the architecture of ANSI/X3/SPARK, which consists of three levels of data description: conceptual ($CM$), inner ($INM$) and external ($EXM$)). Thus, the elements of architecture ARCH_LS of lexicographic system are the following: $\{CM, EXM, INM; \Phi, \Psi, \Xi\}$, where symbol $CM$ denotes the conceptual model of lexicographic system $LS$; symbol $EXM = \{exM\}$ denotes a set its external model, that corresponds to this
conceptual model $CM$, and $INM = \{\text{inM}\}$ is a corresponding set of its internal models. A set of mapping $CM$ in $EXM$, is denoted through $\Phi = \{\varphi\}$.

$$\varphi : CM \rightarrow \text{exM}, \text{ where } \text{exM} \in EXM;$$

(11)

$\Psi = \{\psi\}$ is a set of mapping $CM$ in $EXM$ respectively:

$$\Psi : CM \rightarrow \text{inM}, \text{ where } \text{inM} \in INM;$$

(12)

$\Xi = \{\xi\}$ is a set of mappings $INM$ in $EXM$:

$$\xi(\text{inM}) = \text{exM}.$$  

(13)

In this case we use the following interpretation of the elements of architecture. The conceptual model has following properties:

1. **Semiotics.** The conceptual model is implemented in an environment of some sign system. Its structure involves the joint consideration of the subject area, the thinking domain and the signs domain.

2. **Semantics.** Objects and relations between them, that are essential for an adequate description of knowledge of subject area, is displayed in the categories of model.

3. **The uniqueness.** The conceptual model describes the subject area in accordance with the principle of unambiguous naming, according to which each sign of a model has one sense, one meaning. Homonymy and polysemy in determining the elements of the model is assumed to be removed. Usage of qualifications and contextual mechanisms are not allowed.

4. **The consistency.** For each state of the conceptual model the boundaries between its categories are absolute: during classification or description of the objects of subject domain a uniquely defined set of disjoint states corresponds to each one of them.

5. **Integratedness.** Different specialists’ concepts about the subject area are combined in a conceptual model.

6. **Typed property.** All of conceptually specified properties of the objects of a model must have an interpretation certain types of data.

7. **Algorithm property.** All objects of the conceptual model and connections, mappings and operations with them, have a finite description, i.e. an interpretation in terms of ultimate sophistication algorithms of designed objects.

So the conceptual model (conceptual level of representation) of subject area is a sign, a semantic model in which all concepts of different experts about the subject area are integrated in unambiguous, finite and consistent form.

Types, structures and formats of presentation, storage and manipulation of data, algorithmic database and operational and software environment in which the conceptual model “submerges” during its implementation are determined in the internal model (internal level of presentation).

External model (external level of representation) reflects the views of final users and application programmers on L-system. It implements a set of tools that enable
a user to perform an allowed contacts and manipulation of the data, presented in the internal level.

Several internal and external models can correspond to the one conceptual model, so mappings \( \Phi : CM \to EXM \) and \( \Psi : CM \to INM \), generally speaking, are not injective. But mappings \( \varphi : CM \to exM \) and \( \psi : CM \to inM \) are injective (but not bijective, as usual). Let’s define a set \( \Xi \) of mappings \( INM \) in \( EXM : \) for \( \forall inM \in INM \) and \( \forall exM \in EXM \) \( \exists \xi \in \Xi \) it is:

\[
\xi(inM) = exM. \tag{14}
\]

\[
\xi \circ \Psi = \varphi. \tag{15}
\]

is commutative \( \xi \circ \Psi = \varphi \). The commutativity condition is essential, as it ensures consistency between all levels of the system architecture.

Our experience shows that any dictionary, paper or computer, are satisfied the described formalization. Let us illustrate it on the example of explanatory dictionary of the Ukrainian language.

To determine the structure of the Ukrainian language dictionary, first of all it is necessary to make a basic diagram (7) for it, i.e. to determine objects \( \Lambda(IQ(D)) \), \( P(IQ(D)) \), \( F \), \( C \) and \( H \). This means that it is necessary to determine the structure of objects \( \Lambda(x) \) and \( P(x) \) for all \( h \in W \), where symbol \( W \) denotes class of words of the Ukrainian language.

If it is glossary, there are two types of basic information units: the units of the lexical level and phrases, which have an idiomatic status in the language. Thus this dictionary integrates lexicographic effects of two types. Therefore it is natural to introduce the structure of the dictionary entry \( V(x) \) in the form of a unified description of the structural units of both types:

\[
V(x) \equiv V_{Lex}^{x} \cup \left[ \bigcup_{i} V_{Fras}^{x i} \right],
\]

where \( V_{Lex}^{x} \) — a description of the semantics (grammatical and lexical) of lexeme \( x \); \( V_{Fras}^{x i} \) — a description of the phrase \( i \)-th type; \( n(x) \) — amount of phrases in the dictionary entry \( V(x) \) with the register’s word \( x \).

Each lexicographic complex — both \( V_{Lex}^{x} \) and \( V_{Fras}^{x i} \) — is associated with the basic structure:
If \( V^{Lex}(x) \) than \( \Lambda_0 \) is an entry word with the corresponding grammatical characteristics. For \( V^{Fras}_i(x) \Lambda_0 \) is the phrase in the register dictionary form with certain grammatical characteristics (optionally). The formal structure of the right-hand side is the same for lexemes and for phrases.

Analysis of the right sides of \( P(x) \) provides the following structural elements: the definition of lexical word meanings, definitions of shades of meaning, illustrations for each meaning and each shade of meaning. We denote the definition of i-th meaning of register’s unit \( x \) by \( S_i \), \( S(Shade)_{ij} \) — the definition of j-th shade of i-th meaning, \( I_{ijk} \) — illustration where \( i \) identifies the meaning and \( j \) — connotation, with which the illustration with order nummer \( k \) correlates. Right side of the structure can be represented as a graph:

The structure of illustration divide into substructures \( I(Text) \) and \( I(Passport) \) — the text of illustration and bibliographic description of its source.

Let us generalize the structure of the dictionary entry for the lexicographic system of Etymological dictionary of the Ukrainian language (EDUL).
Let us mark a set of entry words by $I_0(E)$, $V(I_0(E))$ — is a L-system of EDUL:

$$V(I_0(E)) = \cup_{x \in I_0(E)} V(x).$$

The structure of the dictionary can be represented by the following scheme (similar to the dictionary entry scheme):

$$V(I_0(E))$$

$\Lambda_0(I_0(E))$ ---- $P_0(I_0(E))$

We denote the set of all dictionary languages by $L(E)$, $L_i$ — one of the languages, $NL(E)$ — the amount of languages recorded in the dictionary:

$$L(E) = \{L_i, i = 1, 2, ..., N_L(E)\}.$$  

We denote the set of all etymons described in dictionary entries by $Z(E)$ and the set of etymons of a certain language by $Z_{L_i}(E)$:

$$Z(E) = \cup_{L_i} Z_{L_i}(E).$$

For the Ukrainian language we can construct this sort of chain of inclusions ($Z_U(E)$ — the set of etymologized words, marked as Ukrainian words):

$$Z(E) \supset Z_U(E) \supset I_0(E),$$

i.e. we consider a set of entry words as a subset of etymons.

The structure of the left part of the dictionary $\Lambda_0(E)$ includes the following elements:

$$\Lambda_0(E) = L, L_i, Z(E), Z_{L_i}(E), I_0(E), ETYM(E),$$

where the set of dictionary’s etymon-structures (etymon structure is a set of all parameters of etymon; sign representation of etymon and marker of its linguistic identity play a formative role) denotes by $ETYM(E)$. It is obvious, that $\Lambda_0(I_0(E)) = \cup \Lambda_0(x) = ETYM(E)$.

Let’s construct the following chains of mappings:

$$I_0(E) \rightarrow ETYM(E) \rightarrow V(I_0(E));$$
$$Z(E) \rightarrow ETYM(E) \rightarrow V(I_0(E));$$
$$Z_{L_i}(E) \rightarrow ETYM Li(E) \rightarrow V_{L_i}(I_0(E));$$
$$L \rightarrow ETYM(E) \rightarrow V(I_0(E));$$
$$L_i \rightarrow ETYM Li(E) \rightarrow V_{L_i}(I_0(E)).$$

These chains are simulating inputs to the dictionary system; the first one corresponds to the traditional entrance by entry word. In the context of our model this is one of the input options by etymons of a certain language group.
Virtual Labs
In the digital environment computer database organizes text of a dictionary. All selected structural elements of the dictionary entry are displayed in the structure of the database. It provides direct access to each structural element of the lexicographic system and provides the possibility to construct different schemes of indexes. As a rule lexicographic database of a dictionary displays only basic structures. Further widening 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.

The technology of virtual lexicographic laboratory is designed to support and develop digital lexicographic systems. The main functions of the laboratory are:

— users’ authorization and identification;
— adding and deleting of new users;
— control of user’s access rights (viewing of dictionary entries, editing, access to the interfaces, etc.);
— adding new dictionary entries to the lexicographic database;
— removing dictionary entries from the database;
— editing the dictionary entries (adding, deleting structural elements within the boundaries of a given structure the dictionary entry, text editing);
— dynamic reproduction of dictionary entries in print or in any given imaging format;
— data analysis (lexicographic statistics, history of lexicographic representation of each dictionary entry with the authorization of the changes to the database, planning and recording amounts of performed work, marking the stages of lexicographic processing, and marking problematic dictionary entries, etc.);
— sampling from the databases by the different parameters (part of speech, style and industry remarks, quasi semantics formula, etc.);
— creating new types of sorting, SQL-requests and forming subsystems with specified characteristics.

Figures 1–4 demonstrate windows of user’s interfaces.
Figure 1: Log in to the virtual lexicographic laboratory for a Dictionary of the Ukrainian language

Figure 2: Main user screen of the Ukrainian language dictionary
Figure 3. Selection parameters for the Ukrainian language dictionary

Figure 4: Edit screen for dictionary entry of the Ukrainian language dictionary
Other Examples
On the structures of the explanatory and etymological dictionaries were shown examples of construction of the corresponding L-systems and their tool implementations as virtual lexicographic laboratories. It should be mentioned that each of them is “integrated” already, i.e. it combines different types of linguistic units, relationships etc. For example, in the L-system “DUL” words and some types of phrases are combined. Instrumental VLL “DUL” offers the opportunity to identify other relationships between linguistic units (Shyrokov, 2012, pp. 487–510). A similar observation can be make about the VLL “EDUL” The index formed by the tool system of VLL “ESUM” gives us the ability to integrate almost any set of languages. Through the usage of instrumental mechanisms of VLL in described L-systems lexicographic effects of different nature is appearing. Furthermore, each of the mentioned L-systems contains bibliographic subsystems, and that provides a way to integrate different linguistic corps.

Of course, it is a very attractive opportunity to integrate into one object as much aspects of language as possible. A theory of Lexicographic Environments and Lexicographic Agents, that is a rather far-reaching generalization of the theory of lexicographic systems, gives the necessary formal apparatus for this (Shyrokov, 2011).

The practical implementation of one of the methods of L-systems integration, based on theory that was mentioned the above, is demonstrated by the Integrated lexicographic system “Dictionaries of Ukraine”, in which functions of a complete inflectional paradigm, phonemic transcription, phraseology, synonyms and antonyms are combined for the Ukrainian language; online version of this system is presented on the website of Ukrainian linguistic portal http://ulif.org.ua.

It is characteristic that the methods of integration, which based on the formalism of lexicographic medium and agents, are suitable for creating digital dictionary systems with various purposes: user, residential and instrumental.

All listed concepts and language-data objects, in our opinion, are representing quite a universal device for building an integrated L-systems of almost unlimited complexity. Generalization of the idea of L-systems integration naturally leads to the idea of building a digital lexicographic space, which, in stated language-information paradigm, has not only an adequate system parameterization, but also well-defined contours of technological implementation.
<table>
<thead>
<tr>
<th>Language</th>
<th>Entry</th>
<th>ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ukrainian</td>
<td>duznie</td>
<td>1023</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>duka</td>
<td>1023</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>duń</td>
<td>982258</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dwojgo</td>
<td>1003</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dworzanił</td>
<td>1003</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dydnęć się</td>
<td>1013</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dyktować</td>
<td>1014</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dynać</td>
<td>1014</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dyndać</td>
<td>1013</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dz-</td>
<td>1013</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dzban</td>
<td>1013</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>[dzandzeli]</td>
<td>1013</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dzerdza</td>
<td>1013</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dzerdza</td>
<td>1013</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dzied</td>
<td>1016</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dziedzia</td>
<td>1013</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dziez</td>
<td>1009</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dzieckować</td>
<td>1017</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dzień</td>
<td>1017</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dziob</td>
<td>1013</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>[dzopa]</td>
<td>1013</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dziab</td>
<td>1013</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dziubać</td>
<td>1014</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dziubać</td>
<td>1013</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dziubać</td>
<td>1008</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dziudinga</td>
<td>1023</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>dziunka</td>
<td>1013</td>
</tr>
</tbody>
</table>

**Figure 5:** Login to the etymological dictionary with a selected language
References


---

**Acknowledgment**

This work was supported by a core funding for statutory activities from the National Academy of Sciences of Ukraine.

The authors declare that they have no competing interests.

The authors’ contribution was as follows: concept of the study: first author; data analyses: first and second author; wrote the manuscript: first and second author; working with databases: first and second author.

This is an Open Access article distributed under the terms of the Creative Commons Attribution 3.0 PL License (http://creativecommons.org/licenses/by/3.0/pl/), which permits redistribution, commercial and non-commercial, provided that the article is properly cited.

© The Authors 2015

Publisher: Institute of Slavic Studies, PAS, University of Silesia & The Slavic Foundation