

## ИНДЕКСИРОВАНИЕ ЭТИМОЛОГИЧЕСКИХ ЛЕКСИКОГРАФИЧЕСКИХ СИСТЕМ

И. В. Остапова (irinaostapova@gmail.com, iros@zeos.net)

Украинский языково-информационный фонд НАН Украины, Киев, Украина

**Ключевые слова:** этимологический словарь, индекс лексикографической системы, лексикографические базы данных, виртуальная лексикографическая лаборатория

### INDEXING THE ETYMOLOGICAL LEXICOGRAPHIC SYSTEMS

I. V. Ostapova I. V. (irinaostapova@gmail.com, iros@zeos.net)

Ukrainian Lingua-Information Fund,

National Academy of Sciences of Ukraine, Kyiv, Ukraine

The main problems and directions for development of the etymological lexicographic systems in the digital environment were studied. The formal conceptual model of a lexicographic system of the fundamental academic Etymological Dictionary of the Ukrainian Language (EDUL) was developed. The lexicographic structure of the EDUL individual elements was developed and described. The EDUL metalanguage was studied and described. The formal model and technology of the EDUL parsing was developed. That made it possible to automatically convert the EDUL text into the lexicographic database, which corresponds to the conceptual model of the lexicographic system. The conceptual foundations of an instrumental complex to form the etymological dictionaries were developed. This was the bases for creating the Virtual Lexicographic Laboratory «Etymological Dictionary of the Ukrainian Language», which was implemented with a modern approach on the real lexicographic array of the EDUL. That allowed forming the database of the EDUL multilingual index (about 250 languages) in the automatic mode. This index is a basis of the seventh and final volume of the EDUL. The possibility of applying the developed models to other etymological dictionaries was studied. The conceptual foundations of integration of the etymological lexicographic systems were developed.

**Keywords:** etymological dictionary, index of the lexicographic system, lexicographic databases, virtual lexicographic laboratory

Одним из основных инструментов для эффективной работы со словарём является индекс. Индекс в его латентной форме присутствует в любом словаре: это заголовочные слова словарных статей, графически выделенные из массива текста и организованные в алфавитном порядке. При возрастании сложности структуры словарной статьи создание эффективной системы индексирования становится методологической и технологической проблемой, причём равно актуальной как для печатных словарей, так и для цифровых.

Исторически сложилось, что индексы разрабатывались, как правило, для двух типов словарей: фразеологических и этимологических. В первом случае

необходимость индекса обусловлена структурной сложностью реестровой единицы; с помощью индекса организуется доступ к словарной статье по любой компоненте фразеологического словосочетания. Для этимологических словарей индекс является инструментом для установления генетической (этимологической) связи по параметру языковой принадлежности слова.

В идеале индекс должен быть построен по любому лексикографическому параметру словаря. Для реализации этого требования элементы индекса должны быть структурообразующими параметрами лексикографической системы, которая индексируется. Очевидно, развитые технологии индексирования могут быть построены только для цифровых лексикографических систем.

Как и все фундаментальные лексикографические работы, этимологические словари постепенно переводятся в цифровой формат. В интернете наиболее доступен «Этимологический словарь русского языка» М. Фасмера, первая его цифровая версия была издана на CD-ROM ещё в 2004 году. Однако все цифровые репрезентации подобных лексикографических систем являются не более чем разновидностями машиночитаемых текстов, отсутствуют возможности масштабирования и динамического индексирования.

В данной работе рассматриваются метод и технология индексирования Этимологического словаря украинского языка [1] (ЭСУЯ, в печатной версии – 6 томов), которые предлагаются как универсальные для любой этимологической лексикографической системы. Разработана формальная модель словарной статьи, в которой структурные элементы разделяются на два класса: лингвистические (описательные) и структурообразующие, последние вычленяются по формальной процедуре из лингвистических. На основе этой модели построена схема компьютерной базы данных (в рассматриваемой реализации — реляционной) и технология парсинга текста словаря, что позволило в автоматическом режиме конвертировать текст. Для поддержки цифровой версии ЭСУЯ разработан инструментальный комплекс, в который входит подсистема индексирования словаря по любому сформированному пользователем регистру.

В качестве базового структурного элемента словарной статьи определён *этимологический класс* (обозначается *ECL*), представляющий собой блок линейного текста, в котором описываются определённые генетические связи заголовочного

слова. Для ЭСУЯ этимологические классы подразделяются следующим образом: класс реестрового слова (*HEAD*), класс дериватов (*DERIVAT*), класс славянских соответствий (*SLAVIA*), языковой класс (*LANG*), библиографический класс (*BIBL*) и классы ссылок (*REF* и *COMP* в зависимости от типа ссылки). Каждый из этих классов обладает уникальными формальными признаками, по которым идентифицируется в линейном тексте словарной статьи, и имеет индивидуальное содержание и структуру. Классы ссылок и библиографический рассматриваются как служебные. В свою очередь классы могут распределяться на подклассы. Более сложной структурной иерархии не было выявлено. Минимальная структура словарной статьи состоит из двух структурных элементов: класса заголовочного слова и одного языкового класса. Все остальные элементы структуры словарной статьи являются факультативными [2].

Приведём пример словарной статьи ЭСУЯ, достаточно полно демонстрирующей структуры текста этимологического описания (полиграфическая разметка повторяет разметку оригинального текста, цветом выделены разделители классов и подклассов).

- (1) **[баяти]** «розповідати; ворожити», **[байкати]** «розповідати (писати) байки; балакати», **[бай]** «казкар» Пі, **байка**, **байкар**, **байкарство**, **байківниця** «збірник казок і легенд» Ж, **[байко]** «казкар, брехун» Я, **баєчник** «тс.», **[байла]** «ворожбит» Ж, **[байло]** «тс.» Ж, **[байчар]** «балакун, базіка, пліткар», **[баюн]** «казкар, брехун» Я, **[баян]** «співець» Я, **[байкий]** «говіркий» Ж, **баєчний**, **байбай** (приспів колискової пісні), **[баюбаю]**, **[баюлі]** «тс.» Я, **[забаяч]** «заклинач, ворожбит, знахар» Ж; — р. **баять** «говорити», бр. **[баиць]** «говорити, базікати, плести нісенітницю», др. **баяти** «розповідати байки; чарувати», п. **bajać** «розповідати казки; базікати, вигадувати», ч. **bájiť** «розповідати казки», ст. **báti** «тс.», ст. **baju** «тс.», слц. **báj** «міф», **bájka** «байка», вл. **bać** «розповідати казки, говорити нісенітницю», заст. **bać** «розповідати казки, говорити нісенітницю», нл. **bajaś** «базікати, розповідати казки», болг. **бая** «замовляю, ворожу», м. **бае**, схв. **бајати** «тс.», слн. **bájati** «базікати, заклинати, ворожити, пророкувати», **[bajúlje]** «якісь дитячі пісні», стел. **баяти** «розповідати байки, замовляти, заворожувати, заклинати»; — псл. **bajati**; — споріднене з дінд. **bhánati** «говорить», **sabhā** «збори», лат. **fārī** «говорити», гр. **φημί** (дор. **φῶμι**) «говорю», вірм. **ban** «слово, мова», пн.-фриз. **bālen** «говорити», дангл. **bōian** «хвалитися»; з іс. **\*bhā-**; невірно пов'язується (напр., *Sławski* I 25; *Шанский ЭСРЯ* I 2, 65) з лит. **bóti** «звертати увагу», яке походить від п. **dbać** чи бр. **дбаць** і не має приписуваного йому іноді при перекладі значення «питати». — *Фасмер—Трубачев* I 140; *ЭССЯ* 1, 138—139; *Sł. prasł.* I 182; *Sadn.—Aitz.* VWb. I 115; *Bern.* I 39; *Büga* I 584; *Fraenkel* 53; *Maurohofer* II 469—470; *Holthausen PBrB* 48, 460; *Pokorny* 105—106. — Пор. **базікати**, **балакати**, **баніт**, **бари**, **барити**, **басні**.

Распишем все структурообразующие элементы для этой статьи (текстовые блоки будем подавать в угловых скобках и, по возможности, в сокращённом виде; двойным индексом выделяются подклассы):

$ECL_1(\mathbf{баяти}) \equiv HEAD(\mathbf{баяти}) \equiv \langle [баяти] \text{ «розповідати; ворожити»} \rangle$ ;  
 $ECL_2(\mathbf{баяти}) \equiv DERIVAT(\mathbf{баяти}) \equiv \langle [байкати] \text{ «розповідати (писати) байки; балакати»}, \dots, [забаяч] \text{ «заклинач, ворожбит, знахар» Ж} \rangle$ ;

$ECL_3(\text{баяти}) \equiv SLAVIA(\text{баяти}) \equiv \langle \text{р. баять «говорити», ..., стсл. багати «розповідати байки, замовляти, заворожувати, заклинати»} \rangle$ ;

$ECL_4(\text{баяти}) \equiv LANG_1(\text{баяти}) \equiv \langle \text{псл. bajati} \rangle$ ;

$ECL_{5_1}(\text{баяти}) \equiv LANG_{2_1}(\text{баяти}) \equiv \langle \text{споріднене з дінд. bhánati «говорить», sabhā «збори», лат. fāgī «говорити», гр. φημί (дор. φαμί) «говорю», вірм. ban «слово, мова», пн.-фриз. bālen «говорити», дангл. bōian «хвалитися»} \rangle$ ;

$ECL_{5_2}(\text{баяти}) \equiv LANG_{2_2}(\text{баяти}) \equiv \langle \text{з іє. *bhā-} \rangle$ ;

$ECL_{5_3}(\text{баяти}) \equiv LANG_{2_3}(\text{баяти}) \equiv \langle \text{невірно пов'язується (напр., Sławski I 25; Шанский ЭСРЯ I 2, 65) з лит. bóti «звертати увагу», яке походить від п. dbač чи бр. дбаць і не має приписуваного йому іноді при перекладі значення «питати»} \rangle$ ;

$ECL_{6_1}(\text{баяти}) \equiv BIBL_{1_1}(\text{баяти}) \equiv \langle \text{Фасмер—Трубачев I 140} \rangle$ ;

...

$ECL_{6_{10}}(\text{баяти}) \equiv BIBL_{1_{10}}(\text{баяти}) \equiv \langle \text{Pokorny 105—106} \rangle$ ;

$ECL_{7_1} \equiv COMP_{1_1} \equiv \langle \text{базікати} \rangle$

...

$ECL_{7_6} \equiv COMP_{1_6} \equiv \langle \text{басні} \rangle$

Таким образом, словарная статья  $V(\text{баяти})$  в своём составе имеет семь этимологических классов: класс реестрового слова, класс дериватов, класс славянских соответствий, два языковых класса, библиографический класс и класс ссылок. Второй языковой класс имеет в своём составе 3 подкласса, библиографический – 10 подклассов, а класс ссылок – 6. Как линейную последовательность текстовых блоков эту словарную статью можно представить в таком виде:

$$V(\text{баяти}) \equiv HEAD\{, \} DERIVAT\{; —\} SLAVIA\{; —\} \\ LANG_1\{; —\} LANG_{2_1}\{; —\} BIBL\{; —\} COMP.$$

В фигурных скобках подаются последовательности символов, которые выполняют роль разделителей структурных элементов. Для разделения на подклассы используется только один тип разделителя  $\{; \}$ .

Каждому этимологическому классу, за исключением служебных, ставится в соответствие *этимон-структура*, представляющая собой совокупность параметров слов (этимонов), с которыми была установлена генетическая связь. Для ЭСУЯ было определено восемь параметров: маркер языковой принадлежности (обозначается  $PAR_L$ ), ремарка к маркеру языковой принадлежности ( $PAR_{RL}$ ), знаковое представление этимона в определённом алфавите ( $PAR_A$ ), принадлежность к диалектной лексике ( $PAR_{DER}$ ), признак омонимии ( $PAR_{OM}$ ), толкование ( $PAR_S$ ), ремарка ( $PAR_{REM}$ ), библиография ( $PAR_{BIBL}$ ). Все параметры вычленяются из текста по определённым формальным признакам. Обязательными являются параметры языковой

принадлежности ( $PAR_L$ ) и знаковое представление ( $PAR_A$ ). Эти два параметра обеспечивают уникальность каждого этимона в тексте словарной статьи. Для каждой статьи строится как минимум одна этимон-структура (для класса заголовочного слова).

Приведём примеры этимон-структур языковых классов (из факультативных параметров включаем только толкования):

(2)

$LANG_1(\text{баяти}) \equiv \langle \text{псл. bajati} \rangle;$

№	$PAR_L$	$PAR_A$	$PAR_S$
1	псл.	bajati	

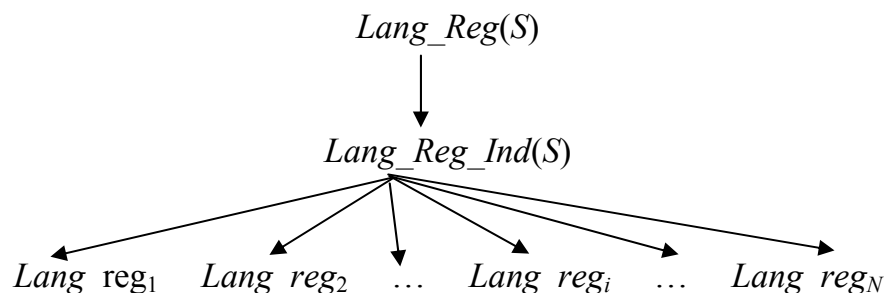
(3)

$LANG_2(\text{баяти}) \equiv \langle \text{споріднене з дінд. bhánati «говорить», sabhā «збори», лат. fāī «говорити», гр. φημί (дор. φαμί) «говорю», вїрм. ban «слово, мова», пн.-фриз. bālen «говорити», дангл. bōian «хвалитися»} \rangle$

№	$PAR_L$	$PAR_A$	$PAR_S$
1	дінд.	bhánati	«говорить»
2	дінд.	sabhā	«збори»
3	лат.	fāī	«говорити»
4	гр.	φημί	«говорю»
5	вїрм.	ban	«слово, мова»
6	пн.-фриз.	bālen	«говорити»
7	дангл.	bōian	«хвалитися»

В печатном словаре текст организуется таким способом, что заголовочное слово имеет шрифтовое выделение и идёт первым словом ряда, чем обеспечивается его структурная значимость возможностями печатного текста. В предлагаемой модели структурообразующими являются обязательные параметры этимон-структуры: вход в словарь возможен по любому языку и по любому слову в алфавите этого языка

Обобщённую схему многоязыкового индексирования этимологического словаря можно представить в таком виде:



$Lang\_Reg(S)$  — регистр всех языков, наявних в этимологических описаниях словаря (к языкам также относим наречия и диалекты).

$Lang\_Reg\_Ind(S)$  — языковой регистр, выбранный для индексирования словаря. Она может совпадать с  $Lang\_Reg(S)$ , или может быть его подмножеством:  $Lang\_Reg\_Ind(S) \subseteq Lang\_Reg(S)$ .

$Lang\_reg_i$  — некоторое подмножество индексного регистра, для него выполняется требование  $Lang\_reg_i \cap Lang\_reg_j \equiv \emptyset$ , если  $j \neq i$ , то есть перечень языков для каждого из этих субрегистров уникальн. Как правило, в состав субрегистров входят родственные языки. Каждый субрегистр, в свою очередь, может подразделяться на более специализированные регистры ( $Lang\_reg_{ij}$ ). Количество уровней иерархии и языковой состав регистров зависят от задач, которые ставят перед собой авторы словаря.

С каждым регистром соотносится определённое множество индексных элементов, которое мы будем называть индекс-множеством, или индексом, этого языкового регистра.

Каждый индексный элемент  $ind\_el_k$  ( $k = 1, 2, \dots, K$ ;  $K$  — количество элементов в соответствующем индексе) имеет такую структуру:

$$ind\_el_k \equiv \{e_k, lang(e_k), loc(e_k)\}, \text{ где}$$

$e_k$  — ЭТИМОН,

$lang(e_k)$  — маркер языковой принадлежности,

$loc(e_k)$  — локализация этимона в тексте словаря.

Для печатного индекса ЭСУЯ предложена такая форма локализации этимона: заголовочное слово соответствующей словарной статьи, номер тома и номер страницы, например *bajati* **баяти** I, 356 или *bajati* *нсл.* **баяти** I, 356:

$$(4) ind\_el \equiv \langle bajati \text{ нсл. } \mathbf{баяти} \text{ I, } 356 \rangle,$$

$$e \equiv \langle bajati \rangle,$$

$$lang \equiv \text{нсл.}$$

$$loc \equiv \langle \mathbf{баяти} \text{ I, } 356 \rangle;$$

Если статья размещается на нескольких страницах, то указывается первая и последняя. Номера томов и страниц — дань традиции и привязка к печатной версии словаря, отдельные тома которого выходили со значительным временным интервалом. Перед конверсией в базу данных текст каждой статьи был привязан к своему печатному оригиналу, номера томов и страниц записаны в соответствующие

поля базы данных. В цифровой версии ЭСУЯ этимон может быть локализован с точностью до порядкового номера слова в строке этимологического класса.

Так как для каждого языка словаря планируется отдельный регистр, то был предложен такой формат печатного индекса (названия регистров на украинском языке):

**АБХАЗЬКА**

...

**АВЕСТІЙСЬКА**

...

varəsa **во́лос** I, 420–421

varəz **бульва́р** I, 293

varəz- **верста́т** I, 357

**АВСТРАЛІЙСЬКІ...**

**ГРЕЦЬКА**

...

κύμβαλον **кімва́л** II 446

...

**...ПРАСЛОВ'ЯНСЬКА..**

...

bajati [**баяти**] I, 356

...

**...ШУМЕРСЬКА**

...

**...ЯКУТСЬКА**

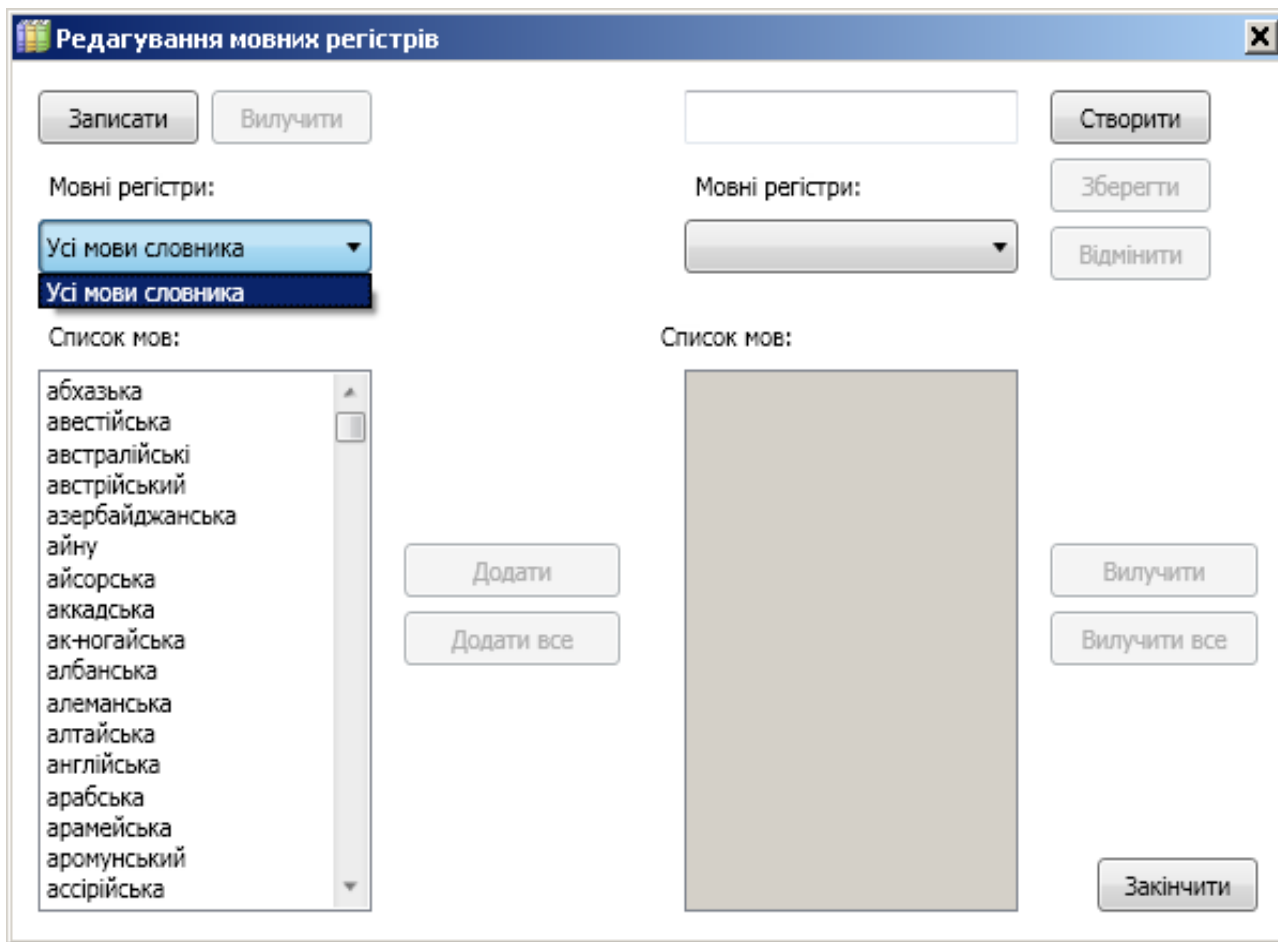
...

**ЯПОНСЬКА**

Задача построения языкового индекса для словаря возложена на инструментальную систему.

При разработке инструментального комплекса языкового индексирования мы стремились минимально ограничивать пользователя системы. Разработанный инструментарий позволяет формировать языковой регистр с любым составом языков, зафиксированных в словаре, то есть объединять в одном регистре произвольно выбранные языки, не учитывая их родственность. Можно создать любое количество таких регистров. Регистры должны различаться только именем (названием), которое присваивается пользователем. Можно задать перечень структурных элементов (этимологических классов), которые будут индексироваться.

На рис. 1 показано окно интерфейса, в котором формируются регистры. На рисунке зафиксировано начальное состояние системы: ещё ни один пользовательский регистр не сформирован.



*Рис. 1.* Окно формирования языкового регистра

Левая панель предназначена для выбора уже созданных регистров как неизменных шаблонов, правая – для формирования новых регистров и их редактирования. Регистр «Усі мови словника» («Все языки словаря») является системным и содержит названия всех языков, зафиксированных в словаре. Любой сформированный регистр, в том числе и системный, могут быть использованы для индексирования словаря.

Пользователь может использовать несколько стратегий формирования нового регистра, например:

1) выбрать на левой панели регистр «Усі мови словника», выбрать из списка языков требуемые позиции и переписать их в именованный список правой панели;

2) выбрать на левой панели регистр с подходящим языковым составом, переписать его на правую панель, а потом отредактировать, исключив ненужные позиции;

3) объединить несколько регистров-шаблонов в один регистр; для этого последовательно выбираются регистры на левой панели и список языков переписывается на правую панель;



4) выбрать на правой панели уже существующий регистр и отредактировать его (не изменяя имени), добавляя недостающие позиции из списка левой панели.

Регистр «СЛОВ'ЯНСЬКІ МОВИ» («славянские языки») на рис. 2. сформирован следующим образом: на основе системного регистра последовательно были сформированы «східно-слов'янські мови» («восточно-славянские языки»), «західно-слов'янські мови» (западно-славянские языки), «південно-слов'янські мови» («южно-славянские языки»), а затем языковые списки были объединены.

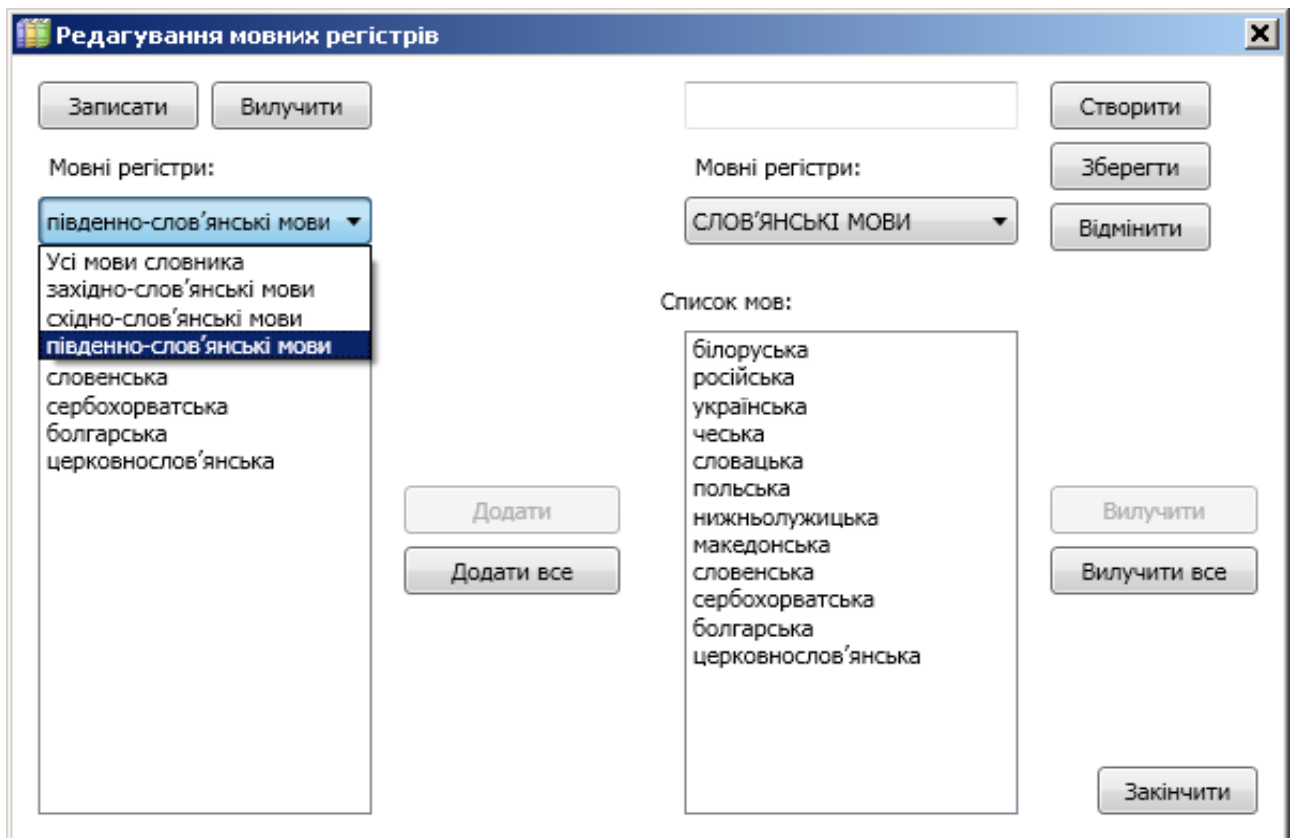


Рис. 2. Формирование регистра «СЛОВ'ЯНСЬКІ МОВИ»

Возможность использование разработанного инструментария для других этимологических лексикографических систем в первую очередь обусловлено возможностью использования формальной модели. С этой целью исследуем тексты словаря «Этимологический словарь славянских языков : Праславянский лексический фонд» [4]. Эта лексикографическая работа выбрана по следующим соображениям. Словарь выходит выпусками небольшого объема в печатном формате в течение длительного времени. Содержит лингвистическую информацию, представляющую постоянный интерес для исследователей. Технология, разработанная для Этимологического словаря украинского языка, позволяет индексировать словарь в процессе его создания, не ожидая завершения работы.

Словарная статья с заголовочным словом **бајџка** достаточно полно характеризует структуру этимологических описаний:

(5)

**\*бајџка:** макед. *bajka* ж. р. ‘сказка’ (И-С), сербохорв. *bajka* ж. р. ‘басня, сказка, fabula’ (с XVII в., RJA I, 151), ‘сказка, выдумка’, ‘ворожба, заговоры’ (РСА I, 250), словен. *bajka* ж. р. ‘сказка’ (Plet. I, 10), чеш. *bajka* ж. р. ‘басня’, ‘сказка, выдумка’, словц. *bajka* ‘басня, выдумка’ (SSJ I, 66), диал. *bajka* ж. р. ‘неправда’ (Buffa. Dlhá Lúka 130), в.-луж. *bajka* ж. р. ‘сказка, басня’ (Pfuhl 5), н.-луж. *bajka* ‘басня, сказка’ (MuKa Sł. I, 11), польск. *bajka* ж. р. ‘сказка, басня’ (Dorosz. I, 301), русск. диал. *ба́йка* ж. р. ‘говор’, ‘сказка’ (Опыт 5), ст.-укр. *байка* ж. р. ‘выдумка’ (Зизаний, Тимченко I, 50), укр. *ба́йка* ж. р. ‘басня’, ‘безделица, шутка, пустяки’ (Гринченко I, 21), блр. *ба́йка* ж. р. ‘басня, небылица’ (Блр.-русс. 115).

Широко распространенное в слав. языках, но, возм., возникшее как параллельное и относительно новое образование с суфф. *-ька* от основы гл. *\*bajati* (см.).

Следуя логике модели, вычленяем четыре этимологических класса:

$EL_1 \equiv \langle *бајџка \rangle$

$ECL_2 \equiv \langle \text{макед. } bajka \text{ ж. р. ‘сказка’ (И-С), …, блр. } байка \text{ ж. р. ‘басня, небылица’ (Блр.-русс. 115)} \rangle$

$ECL_3 \equiv \langle \text{Широко … от основы гл. } *bajati \text{ (см.)} \rangle$

$ECL_4 \equiv \langle *bajati \text{ (см.)} \rangle$

Можно предложить такое распределение этимологических классов по типам: класс реестрового слова ( $ECL_1 \equiv HEAD$ ), славянский класс ( $ECL_2 \equiv SLAVIA$ ), языковой класс ( $ECL_3 \equiv LANG$ ), класс ссылок ( $ECL_4 \equiv REF$ ).

Как линейную последовательность текстовых блоков эту словарную статью можно представить в таком виде:

$HEAD\{\ : \}SLAVIA\{\text{маркер абзаца}\}LANG\{(см.)\}REF.$

Последний класс «встроен» в языковой класс; его можно переформировать таким образом  $\langle \text{См. } *bajati \rangle$ , как формат класса ссылок ЭСУЯ. Такой встроенный класс может быть и не один, как, например, в статье с заголовочным словом **бајидло**:

(6)

**\*бајидло:** словен. *bajilo* ср. р. ‘колдовство. заговор’ (Plet. I, 10).

Производное с суфф. *-dlo* от основы гл. *\*bajiti* (см.). Ср. еще *\*bajadlo* (см.) и ряд слов с основой *\*ba(d)l-* (см. ниже).

Распределение текста по этимологическим классам в этом случае следующее:

$HEAD \equiv \langle *бајидло \rangle$

$SLAVIA \equiv \langle \text{словен. } bajilo \text{ ср. р. ‘колдовство. заговор’ (Plet. I, 10)} \rangle$

$LANG \equiv \langle \text{Производное с суфф. } -dlo \text{ от основы гл. } *bajiti \text{ (см.)}. \text{ Ср. еще } *bajadlo \text{ (см.) и ряд слов с основой } *ba(d)l- \text{ (см. ниже)} \rangle$

$REF_1 \equiv \langle *bajiti \text{ (см.)} \rangle \equiv \langle \text{см. } *bajiti \rangle$

$REF_2 \equiv \langle *bajadlo \text{ (см.)} \rangle \equiv \langle \text{см. } *bajadlo \rangle$

$REF_3 \equiv \langle *ba(d)l- \text{ (см. ниже)} \rangle \equiv \langle \text{см. } *ba(d)l- \rangle$

Построим этимон-структуру для языкового класса словарной статьи **\*bajьka**:

(7)

№	$PAR_L$	$PAR_A$	$PAR_S$
1	макед.	<i>bajka</i>	‘сказка’
2	сербохорв.	<i>bajka</i>	‘басня, сказка, fabula’
3	словен.	<i>bajka</i>	‘сказка’
4	чеш.	<i>bajka</i>	‘басня’
5	слвц.	<i>bajka</i>	‘басня, выдумка’
6	в.-луж.	<i>bajka</i>	‘сказка, басня’
7	н.-луж.	<i>bajka</i>	‘басня, сказка’
8	польск.	<i>bajka</i>	‘сказка, басня’
9	русск.	<i>байка</i>	‘говор’
10	ст.-укр.	<i>байка</i>	‘выдумка’
11	укр.	<i>байка</i>	‘басня’
12	блр.	<i>байка</i>	‘басня, небылица’

При вычленении параметров мы руководствовались алгоритмом, разработанным для ЭСУЯ.

Построим этимон-структуру для класса реестрового слова статьи с заголовочным словом **\*alovьjь HEAD**  $\equiv \langle *alovьjь, *alovaja / *jalovьjь, *jalovaja \rangle$ :

(8)

№	$PAR_L$	$PAR_A$
1	праслав.	<b>*alovьjь</b>
2	праслав.	<b>*alovaja</b>
3	праслав.	<b>*jalovьjь</b>
4	праслав.	<b>*jalovaja</b>

В этимологических словарях для реестровых единиц языковой маркер присваивается по умолчанию, в данном случае – язык праславянский (аналогично для ЭСУЯ – украинский).

На основе вышеизложенного можно сделать вывод, что для этой лексикографической работы («Этимологический словарь славянских языков...») вполне релевантна модель, разработанная для ЭСУЯ.

На базе инструментального комплекса построена виртуальная лексикографическая лаборатория – ВЛЛ «ЕСУМ» [3], которая обеспечивает профессиональное взаимодействие лексикографов в Интернет-среде и предоставляет возможность использования всего функционала системы в режиме on-line.

Применение разработанной технологии для других этимологических лексикографических систем (в первую очередь славянских языков) открывает перспективу формирования лексикографической этимологической среды. При этом подходе словарные системы рассматриваются как относительно независимые объекты, а полнота лексикографического этимологического описания языков создаётся за счёт их интеграции. Как интегрирующая структура может быть использован языковой индекс.

### ЛИТЕРАТУРА

1. Етимологічний словник української мови : В 7 т. К. : Наукова думка, 1982–2011. – Т. 1–6.

2. *Остапова И. В.* Лексикографическая структура этимологического словаря и его представление в цифровой среде // Компьютерная лингвистика и интеллектуальные технологии : По материалам ежегодной Международ. конф. «Диалог 2009» (Бекасово, 27–31 мая 2009 г.). Вып. 8(15). М.: РГГУ, 2009. С. 359–365.

3. *Остапова И. В., Широков В. А.* Виртуальная лексикографическая лаборатория для толковых словарей // Компьютерная лингвистика и интеллектуальные технологии : По материалам ежегодной Международной конференции «Диалог» (Бекасово, 26–30 мая 2010 г.). Вып. 9(16). М.: РГГУ, 2010. С. 363–367.

4. Этимологический словарь славянских языков. М. : Наука, 1974. Вып. 1. 214 с.

### REFERENCES

Etymologichnyj slovnyk ukrajins'koji movy, I–VI, Kyiv 1982–2012

Ètimologicheskij slovar' slavanskih jazykov. Praslavjanskij leksicheskij fond. pod redakciej O. N. Trubachova, 1–31, Moskva 1974–2005

Ostapova I. V. Etymological dictionary: lexicographic structure and representation in digital environment [Leksikograficheskaja structura ètimologicheskogo slovarja i ego predstavlenie v tsifrovoj srede]. *Komp'juternaia Lingvistika i Intellektual'nye Tehnologii: Trudy Mezhdunarodnoj Konferentsii "Dialog 2009"* [Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog 2009"]. Bekasovo, 2009, pp. 359–364

Ostapova I. V., Shyrovkov V. A. A virtual lexicographic laboratory for explanatory dictionaries [Virtualnaja leksikograficheskaja laboratorija dlja tolkovykh slovarej]. *Komp'uternaia Lingvistika i Intellektual'nye Tehnologii: Trudy Mezhdunarodnoj Konferentsii "Dialog 2010"* [Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog 2010"]. Bekasovo, 2010, pp. 363–367