

УДК 004.912:81'374(043.3)

DOI 10.30837/bi.2024.2(101).02

Е. В. Купріянов<sup>1</sup>, І. В. Остапова<sup>2</sup>, М. М. Яблочков<sup>3</sup><sup>1</sup>НТУ «ХПІ», Харків, Україна, eugeniokuprianov@gmail.com,  
ORCID iD: 0000-0002-0801-1789<sup>2</sup>УМІФ НАНУ, Київ, Україна, irinaostapova@gmail.com, ORCID iD: 0000-0001-8221-3277<sup>3</sup>УМІФ НАНУ, Київ, Україна, gezartos@gmail.com, ORCID iD: 0000-0003-1175-1603

## ТЕХНОЛОГІЧНІ АСПЕКТИ СТВОРЕННЯ ВІРТУАЛЬНОЇ ЛЕКСИКОГРАФІЧНОЇ ЛАБОРАТОРІЇ НА БАЗІ ТЛУМАЧНОГО СЛОВНИКА (НА ПРИКЛАДІ ВЛЛ DLE 23)

Стаття присвячена висвітленню окремих важливих питань створення програмно-інструментального комплексу – віртуальної лексикографічної лабораторії ВЛЛ DLE23 для роботи з іспанським словником у цифровому середовищі. Детально викладено процес побудови концептуальної моделі, на основі якої створено базу даних та інтерфейс, вироблено принцип репрезентації тексту словника у лексикографічній базі даних та створено інтерфейс для роботи з текстом словника.

КОМП'ЮТЕРНА ЛЕКСИКОГРАФІЯ, ЛЕКСИКОГРАФІЧНА СИСТЕМА, ВІРТУАЛЬНА ЛЕКСИКОГРАФІЧНА ЛАБОРАТОРІЯ, ЦИФРОВЕ СЕРЕДОВИЩЕ, БАЗА ДАНИХ NOSQL, ІНТЕРФЕЙС ВЛЛ.

**Kupriianov Y. Technological background for creating a virtual lexicographic laboratory based on explanatory dictionary (case of VLL DLE 23).** The article considers some important issues of creating a software and tool complex – the virtual lexicographic laboratory DLE23 for working with the Spanish dictionary in the digital environment. The article describes in detail the process of building a conceptual model on the basis of which the database and interface were created, the principle of representation of the dictionary text in the lexicographic database was developed, and the interface for working with the dictionary text was created.

COMPUTER LEXICOGRAPHY, LEXICOGRAPHIC SYSTEM, VIRTUAL LEXICOGRAPHIC LABORATORY, DIGITAL ENVIRONMENT, NOSQL DATABASE, VLL INTERFACE.

### Вступ

Задоволення інформаційних потреб дослідників користувачів і сучасних лексикографів є одним із найважливіших завдань сучасної лексикографії. У цьому напрямку ведеться постійний пошук різноманітних способів використання широкого спектру можливостей цифрового середовища. Цифрові лінгвістичні технології є основою сучасного словникарства. Насамперед мова йде про цифрові системи укладання та оновлення словників (DWS, скорочено від Dictionary Writing Systems) і корпусні технології (Corpus Query Systems, або CQS). Крім того, слід зазначити, що процес укладання словників вимагає залучення ІТ-фахівців, які працюють над розвитком та підтримкою цифрових технологій у лінгвістиці. Головна проблема, з якою стикається сучасне електронне словникарство, стосується не лише розроблення ефективних форматів репрезентації лексикографічних даних та вироблення відповідних правил використання цих форматів.

Насамперед мова йде про великі (а найчастіше багатотомні) тлумачні словники національних мов, роботи над якими здійснюються в режимі нон-стоп з використанням технологій CQS та DWS, тобто без стадії завершення, як у випадку з Оксфордським словником англійської мови або Словником української мови у 20-ти томах. Незважаючи на суттєвий прогрес у створенні інформаційно-лінгвістичних ресурсів, їхні можливості пошуку, аналізу і узагальнення лінгвістичної інформації досі залишаються доволі обмеженими, особливо для професійних лінгвістів. Насамперед це

стосується вилучення лінгвістичної інформації з великого текстового масиву словника для її подальшого використання фахівцями у своїх дослідженнях.

Виходячи з вищесказаного, метою нашої розвідки є висвітлення окремих важливих технологічних аспектів, що стосуються розроблення цифрового дослідницького середовища – віртуальної лексикографічної лабораторії (ВЛЛ), починаючи формальною моделлю словника як об'єкта ВЛЛ й закінчуючи схемами інтерфейсу. Така ВЛЛ має надавати користувачам інструменти для роботи з великими текстовими масивами словників під час проведення лінгвістичних досліджень.

Виходячи з вищесказаного, метою нашої розвідки є висвітлення окремих важливих технологічних аспектів, що стосуються розроблення цифрового дослідницького середовища – віртуальної лексикографічної лабораторії (ВЛЛ), починаючи формальною моделлю словника як об'єкта ВЛЛ й закінчуючи схемами інтерфейсу. Така ВЛЛ має надавати користувачам інструменти для роботи з великими текстовими масивами словників під час проведення лінгвістичних досліджень.

Створювана ВЛЛ DLE 23 призначена для роботи зі словником іспанської мови «Diccionario de la lengua española. 23ª edición» (далі за текстом DLE 23) у цифровому середовищі. Словник DLE 23 був виданий Королівською академією Іспанії у жовтні 2014 р. Наразі він досі залишається найбільш повним та репрезентативним тлумачним словником іспанської мови. Його онлайн версія доступна на сайті академії [4].

## 1. Реалізація проекту ВЛЛ DLE 23

Проект ВЛЛ DLE 23 передбачає створення двох версій. Перша буде «скороченою» і слугуватиме для тестування деяких технологічних рішень, а друга – повністю функціональною і матиме розширений інтерфейс за результатами функціонування. Незважаючи на «скорочену» версію, ВЛЛ вже має більше можливостей для роботи зі словником у цифровому середовищі, ніж його оригінальна онлайн-версія. Поточна версія ВЛЛ DLE 23 доступна у корпоративному режимі (через логін і пароль) за адресою: <https://svc2.ulif.org.ua/Dics/ResIntSpanish>. Проект реалізується у три етапи, про які детально висвітлено нижче.

Другий етап полягав у створенні бази даних і виборі типу бази даних для ВЛЛ DLE 23. Як свідчить наш досвід, реляційні бази даних виявилися непридатними для створення ефективних цифрових лексикографічних систем, оскільки дані зберігаються в неявному вигляді як сукупність кількох таблиць і зв'язків між ними. Робота з окремими таблицями як з єдиним об'єктом вимагає побудови потужної програмної інфраструктури. Крім цього, еволюційний потенціал такого цифрового об'єкта обмежений непрозорістю бази даних.

Оскільки словникові статті репрезентують елементи лексикографічної системи з чітко визначеною структурою, цілком логічним є репрезентація їх у вигляді класів об'єктно-орієнтованих мов програмування, разом із подальшим опрацюванням, редагуванням і зберіганням в імпліцитній формі. Таку можливість надають бази даних NoSQL (бази даних документного типу). Для баз даних цього типу основним елементом, що підлягає зберігання та опрацюванню, є документ (об'єкт) із чітко описаною структурою.

Основна перевага баз даних NoSQL для нашого проекту – це їхня здатність зберігати лексикографічні об'єкти в явному вигляді без зміни їхньої внутрішньої структури, що відкриває прямий доступ до кожного елемента лексикографічного об'єкта і значно спрощує можливість його редагування та подальшої модифікації (розширення).

Заключним етапом було створення веб-додатка для роботи з базою даних ВЛЛ DLE 23. Структура бази даних повинна відповідати концептуальній моделі.

## 2. Лексикографічна модель даних

Побудова концептуальної моделі має в основі теорію лексикографічних систем (Л-систем) провідного фахівця з комп'ютерної лінгвістики та лексикографії, академіка В. А. Широкова. Основні положення теорії та її реалізації детально висвітлені у працях ученого в [1, 2, 3, 5].

Відповідно до теорії Л-систем, тлумачний словник (так само як і будь-який інший словник) розглядається як Л-система. У свою чергу, Л-система визначається як інформаційна система, в межах якої реалізується один або кілька лексикографічних

ефектів. Будь-яка Л-система визначається такими компонентами:

- $D$  – фрагмент дійсності, що являє собою об'єкт лексикографічного опису;
- $S$  – суб'єкт, який здійснює лексикографічний опис  $D$  (у нашому випадку асоціюється з авторами словника);
- $Q$  – лексикографічний ефект, що спостерігається суб'єктом  $S$  у  $D$  і трансформується в набір елементарних інформаційних одиниць  $I^Q(D)$  (у нашому випадку ми інтерпретуємо його як набір мовних одиниць, що складають список заголовних слів словника);
- $V(I^Q(D))$  – множина описів  $I^Q(D)$ ;  $S: I^Q(D) \rightarrow V(I^Q(D))$ .

З огляду на вищесказане, нижче наведено твердження буде вірним для будь-якої заголовкової слова  $x$ :

$$I(D) = \{x\}; \forall x \in I^Q(D) S: x \rightarrow V(x) \quad (1)$$

де  $V(x)$  у словнику – це текст словника, що описує заголовне слово  $x$ . Отже,  $V(I^Q(D))$  – це сукупність усіх словникових статей. На множині описів  $V(I^Q(D))$  і, зокрема, на кожному  $V(x)$  можна визначити дві структури:  $\beta$  та  $\sigma[\beta]$ . Вони є носіями мовних фактів і закономірностей у Л-системі. Водночас  $\beta$  – це сукупність «дуже простих» структурних елементів словника (слів, аббревіатур, позначок, приміток, рисунків, елементів граматичного та лексичного опису тощо). Це можна сформулювати наступним чином. Для кожного  $x \in I^Q(D)$  визначається набір структурних елементів  $\beta(x)$ , які складають  $V(x)$  за наступними принципами:

1.  $x \in \beta(x)$ ;
2. Будь-який фрагмент словникової статті  $V(x)$  можна побудувати з елементів  $\beta(x)$ ;
3. Принцип формування елементів  $\beta(x)$  має бути спільним для всіх  $V(x)$ , тобто для всіх  $x \in I^Q(D)$ .

Необхідно вказати на важливість сформульованих принципів формування  $\beta$ -структур у лексикографії. Правило 2 фактично є вимогою щодо універсальності метамови словника: будь-який мовний факт, зафіксований у певному словнику, має бути відображений у його метамові. Правило 3 передбачає, що всі однотипні мовні факти та явища повинні мати уніфіковану репрезентацію в лексикографічному описі. Наведені правила створюють об'єктивні передумови для формалізованого визначення процесу лінгвістичного пізнання за допомогою Л-системи.

Продемонструємо на прикладах  $\sigma[\beta]$  лексикографічний опис заголовного слова *agua*. Текст словникового елемента подано у форматі, що зберігає шрифтову розмітку, використану в онлайн-версії словника (рис. 1). На основі аналізу текстів онлайн-версій статей DLE 23 ми визначили такі параметри лівої частини  $L_0$ : *RR* (форми лем), *DUPL* (регіональний варіант), *ETYM* (етимологія), *MORPHO* (флексія), *ORTHO* (орфографія) і *UNCRT* (невизначений параметр). Кожен параметр представлений у нашій моделі у вигляді текстового рядка. Права частина  $P_0$  складається з елементів опису лексичного значення. Полісемія

заголовного слова визначається кількістю цих описів. Кожен опис може включати кілька структурних елементів, а саме: *MNGN* (номер визначення), *REM* (набір позначок), *DEF* (визначення), *ED* (енциклопедична довідка), *COM* (коментар) та *IL* (ілюстрація).

agua

Del lat. *agua*.

1. f. Líquido transparente, incoloro, inodoro e insípido en estado puro, cuyas moléculas están formadas por dos átomos de hidrógeno y uno de oxígeno, y que constituye el componente más abundante de la superficie terrestre y el mayoritario de todos los organismos vivos. (Fórm. H<sub>2</sub>O).
2. f. Líquido que se obtiene por infusión, disolución o emulsión de flores, plantas o frutos, empleado como refresco o en medicina y perfumería. *Agua de azahar, de cebada, de limón.*
3. f. *lluvia* (l acción de llover). U. t. en pl. con el mismo significado que en sing.
4. f. *lágrimas* (l gotas de la glándula lagrimal). *Se le llenaron los ojos de agua.* U. t. en pl. con el mismo significado que en sing.
5. f. Vertiente de un tejado. *Una cubierta a dos aguas.*
6. f. *Mar. marea* (l movimiento periódico de las *aguas* del mar).
7. f. *Mar.* Rotura, grieta o agujero por donde entra en la embarcación el *agua* en que ella flota. *Abrirse, descubrirse un agua.*
8. f. pl. Visos u ondulaciones que tienen algunas telas, plumas, piedras, maderas, etc.
9. f. pl. Visos o destellos de las piedras preciosas.
10. f. pl. Manantial de *aguas* minero-medicinales.
11. f. pl. *agua minero-medicinal.* *El balneario es famoso por sus aguas.*
12. f. pl. Zona marítima próxima a la costa de un lugar. *Naufrió en aguas de Cartagena.*
13. f. pl. *Mar.* Comientes del mar. *Las aguas tiran o van haciatal parte.*
14. f. pl. *Mar.* Estela o camino que ha seguido un buque. *Buscar, ganar, seguir las aguas de un buque.*
15. interj. *jerg.* U. para avisar de la presencia de cualquier tipo de autoridad.
16. interj. *Mar. hombre al agua.*

Рис. 1. Словникова стаття DLE 23 для слова *agua*

Текстовий рядок цієї словникової статті в DLE 23 можна розбити на менші фрагменти, кожен з яких містить мітку певного типу: *REM-GR* (граматика); *REM-US* (вживання); *REM-ST* (стилістика); *REM-DOM* (домен); *REM-REG* (географічний регіон). Як правило, лексичне значення у вхідному тексті описується структурним елементом *DEF*. Додаткові коментарі (*COM*) узгоджуються з визначенням. До кожної дефініції та коментаря можуть додаватися власні ілюстрації (*IL*). Структура тлумачення може включати декілька *DEF*, *COM* та *IL*. Поділ тексту на структурні елементи для заголовкового слова *agua* (вода) показано в таблиці 1. Для прикладу взято описи лексичних значень 1, 2, 4, 7 і 15.

Таблиця 1

Значення елементів правої частини словникової статті *agua* (вода)

Елемент	Значення
MNGN	1
REM-GR	f.
DEF	Líquido transparente, incoloro, inodoro e insípido en estado puro, cuyas moléculas están formadas por dos átomos de hidrógeno y uno de oxígeno, y que constituye el componente más abundante de la superficie terrestre y el mayoritario de todos los organismos vivos
ED	(Fórm. H <sub>2</sub> O)
MNGN	2
REM-GR	f.
DEF	Líquido que se obtiene por infusión, disolución o emulsión de flores, plantas o frutos, empleado como refresco o en medicina y perfumería

Елемент	Значення
IL	Agua de azahar, de cebada, de limón
MNGN	4
REM-GR	f.
DEF	lágrimas (   gotas de la glándula lagrimal)
IL	Se le llenaron los ojos de agua
COM	U. t. en pl. con el mismo significado que en sing.
MNGN	7
REM-GR	f.
REM-DOM	Mar.
DEF	Rotura, grieta o agujero por donde entra en la embarcación el agua en que ella flota
IL	Abrirse, descubrirse un agua
MNGN	15
REM-GR	f.
REM-US	jerg.
DEF	U. para avisar de la presencia de cualquier tipo de autoridad

### 3. Репрезентація тексту словника в структурі лексикографічної бази даних

Для формування лексикографічної бази даних словника було відібрано 89528 словникових статей з онлайн-версії іспанського словника, доступної на сайті Королівської іспанської академії (<https://dle.rae.es/>). Тексти статей були отримані в оригінальній розмітці (HTML 5.0) зі збереженням структури змісту кожної статті та неявно представлених маркерів елементів статті. Враховуючи, що дослідницьку систему планується створювати в автономному режимі на різних етапах проектування, база даних розроблена з можливістю легкої модифікації її структури в майбутньому. Як мову програмування та технологічну платформу було обрано, відповідно, C# та .Net. База даних, яку було використано для словника, належить до документного типу, перевагами якої є: легкість у використанні, підтримка транзакційних механізмів, можливість паралельного доступу до БД, безкоштовна для дослідницьких потреб.

Додатковою перевагою баз даних документного типу для нашої роботи є можливість зберігати лексикографічні об'єкти в явному вигляді, не порушуючи їхньої внутрішньої структури, що забезпечує прямий доступ до кожного елемента лексикографічного об'єкта і, крім того, значно спрощує можливість їхнього редагування та модифікації (розширення). Для нашого проекту було обрано базу даних документного типу LiteDB (<http://www.litedb.org/>). Це відносно проста, умовно-безкоштовна версія бази даних MongoDB. Завдяки тому, що LiteDB побудована у вигляді одного бібліотечного файлу (dll) та одного конфігураційного файлу (xml), а не цілого програмного пакету, ця база даних також має перевагу в тому, що її легко встановлювати та підключати.

Для опрацювання HTML-файлів словникових статей було використано бібліотеку для розбору високоструктурованих текстових рядків – Html Agility Pack (<https://html-agility-pack.net>). На першому етапі ми визначили початкові параметри текстів словникових статей для їхнього індексування. Було виділено лише ті параметри, які були явно представлені в структурі словникових статей. Більш глибокий аналіз словникових статей було відкладено до наступної ітерації ственду. Ці параметри можна поділити на такі підтипи: 1) параметри, що характеризуються наявністю чи відсутністю певного текстового рядка або підрядка в HTML-коді статті; 2) параметри, визначені шляхом аналізу HTML-коду статті.

Аналіз різних типів словникових статей дав можливість виокремити такі явні характеристики: ознака омонімії, опис типу заголовного слова, опис структури заголовного слова, тип заголовного слова, кількість словосполучень «іменник + прикметник», кількість словосполучень іншого типу та первинний опис структури словникової статті (зона дефініції) [].

Ці характеристики обрано, по-перше, тому, що їх легко ідентифікувати в тексті словникової статті, а по-друге, тому, що вони являють собою достатній набір індексних параметрів для формування вибірки з лексикографічної бази даних та аналізу структур і текстів словникових статей. Було розроблено три типи класів для представлення у внутрішній моделі програмного додатку, що розробляється:

- Клас типу «Article», що містить текст статті у форматі HTML.

- Клас типу «R\_Member», що містить лише заголовне слово словникової статті, а також його параметри, які фактично є елементами індексу словника.

- Клас типу «R\_Unit», що містить посилання на класи типу «R\_Member» і, таким чином, представляє заголовне слово разом з його родовими формами.

Оскільки додаток використовує документну базу даних, то дані в ній зберігаються відповідно до їх репрезентації у внутрішній моделі. Клас типу «Article» має таку структуру:

- Внутрішній ідентифікатор бази даних: словника цілочисельна змінна `_ID`.

- Ідентифікатор словникової статті, до якої належить даний клас: текстова змінна `Inner_ID`: текст запису у форматі HTML: текстова змінна `Text`.

Клас типу «R\_Member» складають такі структурні елементи:

- Внутрішній ідентифікатор бази даних словника: цілочисельна змінна `_ID`.

- Ідентифікатор словникової статті, до якого належить цей клас: текстова змінна `Inner_ID`.

- Ідентифікатор «повної» словникової статті у випадку, якщо оригінальний запис містить лише посилання: `Orig_Inner_ID`.

- Форма роду заголовного слова: текстова змінна `Member`.

Наведемо приклад класів `R_Member` та `Article` для заголовного слова *abacalero*, словникова стаття якого подана на рис. 2.

## abacalero, ra

1. *adj. Filip.* Pertenciente o relativo al abacá o al abacal.

2. *m. y f. Filip.* Persona que cultiva abacá o comercia con él.

Рис. 2. Словникова стаття *abacalero, ra*

Клас `R_Member` для аналізованого заголовкового слова матиме такий вигляд:

```
{
  "_id": 135,
  "Inner_ID": "00HQuS9",
  "Orig_Inner_ID": null,
  "Member": "abacalero",
  "Omonim": 0,
  "Compon": 11,
  "Struc": 21,
  "Type": 30,
  "Mean_Count": 2,
  "Phras1_Count": 0,
  "Phras2_Count": 0,
  "Phras_Link_Count": 0
}
{
  "_id": 136,
  "Inner_ID": "00HQuS9",
  "Orig_Inner_ID": null,
  "Member": "abacalera",
  "Omonim": 0,
  "Compon": 12,
  "Struc": 21,
  "Type": 30,
  "Mean_Count": 2,
  "Phras1_Count": 0,
  "Phras2_Count": 0,
  "Phras_Link_Count": 0
}
```

Клас `Article`, у якому репрезентовано інформаційні елементи словникової статті для аналізованого заголовкового слова, виглядає таким чином:

```
{
  "_id": 118,
  "Inner_ID": "00HQuS9",
  "Text": "<article id=\\"00HQuS9\\">\n<header title=\\"Definicion de abacalero, abacalera\\" class=\\"f\\">abacalero, ra</header>\n<p class=\\"j\\" id=\\"009O119\\"><span class=\\"n_acep\\">1. </span><abbr class=\\"d\\" title=\\"adjetivo\\">adj.</abbr> <abbr class=\\"c\\" title=\\"Filipinas\\">Filip.</abbr> <mark data-id=\\"SkQu9TG|SkSTU0O\\">Pertenciente</mark> <mark data-id=\\"QlqTEX0|Qlr66uc|Qltkqueu\\">o</mark> <mark data-id=\\"VpYCENC\\">relativo</mark> <mark data-id=\\"IPTvt8b\\">al</mark> <mark data-id=\\"00GxCz2\\">abaca</mark> <mark data-id=\\"QlqTEX0|Qlr66uc|Qltkqueu\\">o</mark> <mark data-id=\\"IPTvt8b\\">al</mark> <mark data-id=\\"00H2PMC\\">abacal</mark>.</p>\n<p class=\\"j1\\" id=\\"009oafT\\"><span class=\\"n_acep\\">2. </span><abbr class=\\"d\\" title=\\"nombre masculino y femenino\\">m. y f.</abbr> <abbr class=\\"c\\" title=\\"Filipinas\\">Filip.</abbr> <mark data-id=\\"SjUIL8Z|SjwafWr\\">Persona</mark> <mark data-id=\\"UkbUarn\\">que</mark> <mark data-id=\\"Beoktf0\\">cultiva</mark> <mark data-id=\\"00GxCz2\\">abaca</mark> <mark data-id=\\"QlqTEX0|Qlr66uc|Qltkqueu\\">o</mark> <mark data-id=\\"9vXHQNND\\">comercia</mark> <mark data-id=\\"A5cH5M4\\">con</mark> <mark data-id=\\"ESSilCu\\">el</mark>.</p>\n</article></p>
```

## 4. Інтерфейс ВЛІ DLE 23

Для можливості функціонування DLE 23 у цифровому середовищі та, відповідно, побудови інструментарію для проведення лінгвістичних досліджень

на масиві словника, було здійснено декомпозицію його текстової структури на інформаційні елементи за процедурою, описаною в розділі 3. У табл. 2 подано інформаційні елементи, які містить концептуальну модель Л-системи вони доступні для користувача через інтерфейс ВЛЛ DLE 23. Зірочкою позначено елементи, які вже доступні в поточній версії лабораторії. Усі інші будуть реалізовані у фінальній версії.

Таблиця 2

Інформаційні елементи Л-системи DLE 23

No	Елемент	Маркер	Шрифт	Колір
1	Лема			
1.1.a	Слово*	Немає	Жирн. / жирн. курсив	Синій
1.1.b	Словосп. Ім. + прикм.	Немає	Жирний	Темно-коричн.
1.1.c	Словосп. інш. типу	Немає	Жирний	Світло-коричн.
1.2	Омонім*	Верхній індекс	Жирний	Синій
2	Реєстровий ряд*	Немає	Жирний	Синій
2.1	Форма чол. р.	Словоф.	Жирний	Синій
2.2	Форма ж. р.	Закінч-я	Жирний	Синій
3	Дублети*	“Тв.”	Прямий	Темно-синій
4.2	Етимон	Немає	Курсив для лат. ет-нів та прямий для інших	Зелен.
4.3	Додатк. інф.	Немає	Прямий	Зелен.
5	Морфологія	Немає	Прямий	Темно-синій
6	Орфографія	“Escr.”	Прямий	Зелен.
6.1	Орфографічна особливість	Немає	Прямий	Темно-синій
6.2	Номер тлумачення	“en aser.”	Прямий	Темно-синій
7.	Блок тлумачень	№ перш. тл. “1”	Жирний	Чорний
7.1	Блок ремарок	Немає	Прямий або курсивний	Темно-синій
7.2	Дефініція	Немає	Прямий	Чорний
7.2.2	Ілюстративні приклади	Немає	Курсивний	Фіолетовий
7.2.3	Додаткові коментарі	Немає	Прямий	Темно-синій
7.2.4	Енциклопедичний елемент	“(“ та “)”	Прямий	Чорний

На рис. 3 показано головне вікно ВЛЛ DLE 23, досягне на сайті <https://services.ulif.org.ua:44359/>, (корпоративний доступ за логіном та паролем). Його складають: верхній рядок меню (1), панель пошуку разом із словниковим реєстром (2), поле перегляду словникових статей (3) та поле перегляду HTML-тексту словникових статей (4). Верхня панель меню включає два інструменти: «Вибір» та «Статистика».

Перший подає групу параметрів для формування вибірки словникових статей, що містять лінгвістичні ознаки головного слова (тип, структура реєстрового слова, омонімія, кількість лексичних значень тощо). Другий генерує статистику для конкретної вибірки словникових статей або всього словника. Інтерфейс цифрової версії DLE 23 забезпечує три входи: за реєстровими одиницями, за структурним профілем словникової статті та повнотекстовий пошук. Зробимо короткий аналіз можливостей кожного входу.

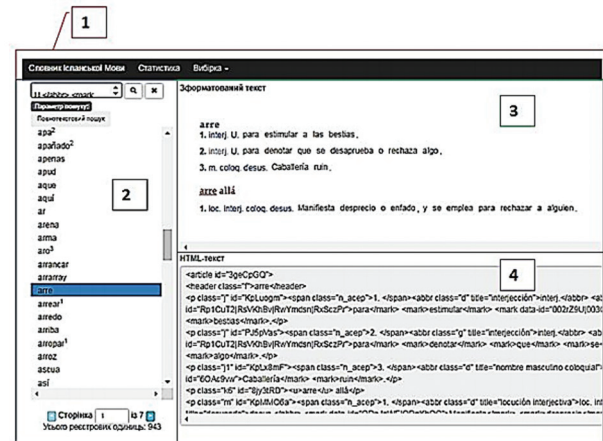


Рис. 3. Вигляд основного інтерфейсу ВЛЛ DLE 23

Інтерфейс цифрової версії DLE 23 забезпечує три входи: за реєстровими одиницями, за структурним профілем словникової статті й повнотекстовий пошук. Зробимо короткий аналіз можливостей кожного входу.

Вхід за реєстровими одиницями дає можливість доступу до словникових статей через реєстр словника. Реєстрову одиницю можна обрати двома способами: безпосередньо в реєстровому вікні або задати послідовність символів, яка має точно збігтися з реєстровою одиницею. Із цієї послідовності символів має починатися чи закінчуватися реєстрова одиниця, вона також може бути її складником.

Вхід за структурним профілем словникової статті в актуальній версії ВЛЛ DLE 23 (рис. 4) призначено для створення вибірки словникових статей словника, що задовольняють параметрам елементів словникової статті, репрезентованих у концептуальній моделі DLE 23. Режим активується натисканням меню «Вибір», після чого з'являється діалогове вікно. Діалогове вікно має дві вкладки «Параметри реєстру» та «Параметри тлумачної частини».

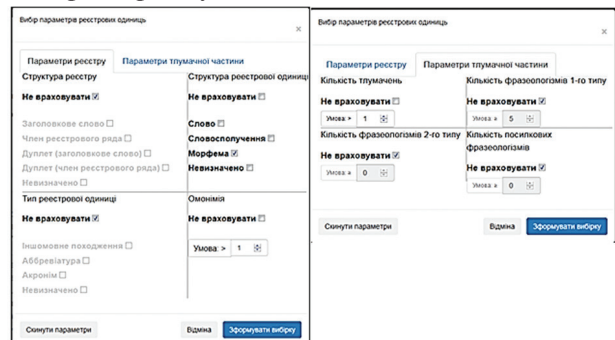


Рис. 4. Вікна для створення вибірки словникових статей словника

У першій користувач може вибрати параметри відбору реєстрових слів:

- структура реєстрового ряду: заголовкове слово, компоненти реєстрового ряду, дуплети;
- структура реєстрової одиниці: слово, словосполучення, морфема;
- тип реєстрової одиниці (аббревіатура, акронім, слово іншомовного походження);
- омонімія; полісемія; словосполучення (різних типів та посилкові фразеологізми).

Інструмент «Вибірка» дає можливість вибирати статті, заголовкові слова яких мають спільні граматичні, лексичні та інші особливості, відображені в тексті словникових статей.

Друга вкладка призначена для вибору записів, які відповідають параметрам дефініцій / колокацій:

- кількість дефініцій: числове значення та додаткові опції (>, ≥, =, ≤, <);
- кількість колокацій типу «іменник + прикметник»: числове значення та додаткові опції (>, ≥, =, ≤, <);
- кількість колокацій інших типів: числове значення та додаткові опції (>, ≥, =, ≤, <);
- кількість відсилкових колокацій: числове значення та додаткові опції (>, ≥, =, ≤, <).

Словникові статті можна обирати одночасно за заголовком та параметрами дефініція / словосполучення. На рисунку 6.5 наведено вибірку словникових статей, що містять омонімічні багатозначні морфеми. Вибірка відповідає таким параметрам:

1. Структура реєстрової одиниці: морфема;
2. Омонімія: ≥1;
3. Кількість лексичних значень (дефініцій): > 1.

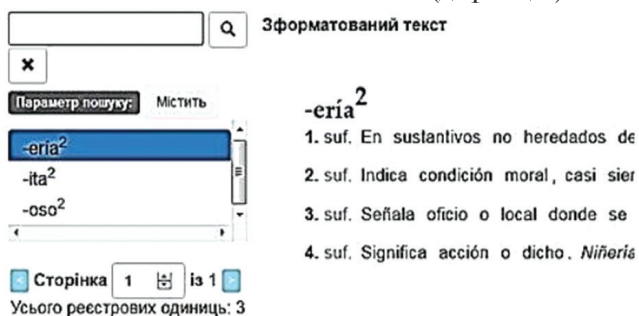


Рис. 5. Вибірка статей з омонімічними багатозначними морфемами

Як приклад роботи у режимі повнотекстового пошуку сформуємо вибірку реєстрових слів, які є гіпонімами до слова *embarcación* (корабель, судно). Для досягнення цієї мети необхідно буде створити вибірку зі статей, у яких леми тлумачаться через слово *embarcación*. У цьому випадку потрібні обидва інструменти «Вибірка» та «Повнотекстовий пошук». Наведемо послідовність кроків:

1. У діалоговому вікні «Вибірка» поставити галочку «Слово» (вкладка «Параметри реєстру») та натиснути кнопку «Сформувати вибірку»;
2. Вибрати «Повнотекстовий пошук» у вікні «Параметри пошуку»;
3. Ввести слово «Embarcación» у пошуковому полі, а потім натиснути кнопку пошуку.

У результаті отримуємо вибірку зі 109 статей, заголовкові слова яких є гіпонімами слова *embarcación*: *aljibe* (цистерна), *almadía* (дощатий човен), *barca* (човен), *barcón* (військовий корабель), *barquía* (гребний човен) тощо. Крім цього, у дефініціях можна виокремити компоненти лексичного значення, якими леми відрізняються (диференційні семи): 1) призначення; 2) конструкція; 3) такі характеристики, як форма та розмір; 4) географічний регіон; та 5) проміжок часу. Слід зазначити, що деякі дефініції можуть мати декілька компонентів. Отримані результати можна використовувати для складання глосаріїв до лекцій, вікторин-словників та вправ для запам'ятовування нових слів тощо.

## Висновки

Розроблена віртуальна лексикографічна лабораторія (ВЛЛ) є практичною реалізацією моделі Л-системи іспанської мови. Призначенням ВЛЛ є аналіз тексту словника з метою:

- інвентаризації реєстрових слів, що відповідають установленим параметрам (питоме слово, іноземне слово; морфема, аббревіатура, слово, словосполучення тощо);
- дослідження в тексті словникової статті лінгвістичних особливостей реєстрових слів. Це дає змогу виявити закономірності в іспанській мові, які подані в словнику в імпліцитній формі;
- статистичні дослідження, які демонструють частоти досліджуваних мовних явищ (наприклад, співвідношення питомої та запозиченої лексики тощо).

На ґрунті цих досліджень користувач може зробити певні висновки щодо лексико-семантичних, етимологічних, граматичних та побутових особливостей іспанських мовних одиниць. У майбутньому поточна версія VLL DLE 23 буде забезпечена розширеним набором інструментів для роботи окремо з кожним елементом словникової статті, визначаючи не лише його наявність або відсутність, але й конкретний зміст.

## Література

- [1] Language. Information. System / V. Shirokov; Ukrainian Lingua-Information Fund of the NAS of Ukraine. Kyiv: Akadempriodyka, – 2021. – 160 p.
- [2] Лінгвістично-інформаційні студії: Праці Українського мовно-інформаційного фонду НАН України: у 5 т. – Т. 3: Тлумачна лексикографія. – Кн. 2: Системна семантика тлумачних словників / В. А. Широков та ін. Київ: Український мовно-інформаційний фонд НАН України, – 2018. – 250 с.
- [3] Лінгвістично-інформаційні студії: праці Українського мовно-інформаційного фонду НАН України: у 5 т. / В. А. Широков та ін. – Т. 1: Наукова парадигма та основні мовно-інформаційні структури. Київ: Український мовно-інформаційний фонд НАН України, – 2018. 271 с.
- [4] Diccionario de la lengua española. URL: <https://dle.rae.es/el>
- [5] Купріянов Е. В. Лексикографічна система іспанської мови: феноменологія інтегрального опису: [монографія] / Е. В. Купріянов; Укр. мовно-інформ. фонд НАНУ. – Київ : УМІФ НАНУ, 2018. – 250 с.

Надійшла до редколегії 06.11.2024