

Масив іменників поділено на двадцять п'ять тематичних файлів, кожний з яких має справу з окремим первісним семантичним компонентом, наприклад: {дія, діяльність}, {тварина, фауна}, {артефакт}, {атрибут, властивість} та інші. Ці ієрархії дуже відрізняються за розмірами і не є взаємно виключаючіми альтернативами — деякі перехресні посилання іноді потрібні — але в цілому вони покривають окремі поняттєві та лексичні області. Ієрархії мають десь в середині рівень, до якого прив'язана більшість характеристичних властивостей. Їх було вирішено ввести таким самим чином, як у традиційних словниках, додавши короткі пояснювальні нотатки до синсетів, що включають багатозначні слова. Вони відмежовані дужками від решти синсета.

EuroWordNet є розширенням WordNet, призначеним для побудови багатомовної бази даних, що складається з баз даних типу WordNet на декількох європейських мовах. Кожна база даних структурована таким самим чином, як і WordNet, тобто синоніми групуються в синонімічні ряди (синсети), які в свою чергу зв'язані за допомогою базових семантичних відношень. Крім того, кожне значення під'єднується відношенням еквівалентності до синсета базового WordNet, таким чином створюючи багатомовну базу даних.

Найбільша відмінність EuroWordNet від WordNet — це багатомовність. Усі мовно-специфічні бази даних зберігаються в центральній лексичній системі баз даних. Відношення еквівалентності між синсетами різних мов та базового WordNet реалізуються в явному вигляді у міжмовному індексі (Inter-Lingual-Index, ILI). Кожний синсет в одномовній базі даних буде мати як мінімум одне відношення еквівалентності з синсетом або записом в ILI. Мовно-специфічні синсети, під'єднані до одного ILI-запису, будуть у такий спосіб еквівалентні для різних мов.

Фактична внутрішня організація синсетів засобами семантичних відношень залишається тою самою, що і в WordNet. Єдина організація, яку забезпечує ILI, являє собою дві окремі онтології, під'єднані до ILI-записів:

— онтологія найвищих понять, що є ієрархією мовно-незалежних понять, яка явно відображає відношення протиставлення (наприклад, “живий” та “неживий”) та може представляти різні семантичні поля в словнику WordNet;

— ієрархія доменних позначок, яка відноситься до концепції базових розділів або тем (наприклад, “спорт”, “водний спорт”, “зимній спорт”, “військо”, “лікарня”).

І загальні поняття, і доменні позначки можуть бути перенесені відношеннями еквівалентності ІЛІ-записів на мовно-специфічні значення та далі, через внутрішньомовні відношення на інші значення у WordNet. Обидві ієрархії дозволяють користувачеві ввести до бази даних потрібні семантичні властивості без необхідності мати доступ до внутрішньомовних відношень кожного WordNet. Крім того, доменні позначки можуть безпосередньо використовуватися для інформаційного пошуку, щоб згрупувати поняття іншим чином, оснований на розділах, а не на класифікації.

Загальна модульна структура бази даних EuroWordNet може бути узагальнена у такий спосіб. По-перше, це мовні модулі, до яких входить поняттєвий лексикон кожної мови. По-друге, це мовно-незалежний модуль, який охоплює ІЛІ, доменну онтологію та онтологію вищих понять. Така модель багатомовної бази даних має певні переваги:

- можливість використовувати базу даних для багатомовного інформаційного пошуку, переходячи від слів однієї мови до відповідних слів іншої мови засобами ІЛІ;

- бази даних WordNet різних мов можуть бути порівняні та перевірені, що робить їх більш сумісними;

- мовно-залежні особливості можна підтримувати в окремих WordNet;

- можлива розробка різних бази даних WordNet в різних місцях відносно незалежно одна від одної;

- мовно-незалежна інформація, така як доменні знання, аналітичні вищі поняття та приклади може бути збережена тільки один раз та зроблена доступною для всіх мовно-специфічних модулів засобами міжмовних відношень;

- база даних може бути пристосована до потреб користувача шляхом коригування найвищих понять, доменних позначок та прикладів без необхідності знати окремі мови або мати доступ до мовно-специфічних баз даних.

6.2.2. Л-система Словника синонімів української мови

Зараз, коли структури WordNet та EuroWordNet фактично претендують на статус стандарту побудови електронних словників тезаурусного типу, розробка аналогічної системи для української мови стає нагальною необхідністю. Основою для створення такої системи можуть стати Словник української мови та інші електронні

словники в рамках проекту “Словники України”, електронні ЛБД яких розроблені в Українському мовно-інформаційному фонді.

Для виділення синсетів доцільно використати Л-систему електронного словника синонімів (ССУМ), про який йшлося у четвертому розділі при викладі засад ІЛС “Словники України”. Але структура ЛБД синонімічної підсистеми ІЛС “Словники України” розроблялася в розрахунку тільки на перегляд словника користувачем і не передбачала можливості проведення досліджень та аналізу, тому вона є дуже спрощеною і складається всього з трьох полів: унікального номеру статті, реєстрового слова та власне тексту статті. Для відображення словникових статей на екрані комп'ютера така структура є достатньою, але для побудови на її основі системи типу WordNet — ні. Крім того, дуже багато статей словника є відсилковими, тобто в кожному синонімічному ряді виділено слово, яке вважається основним, і саме на нього подається словникова стаття. Для всіх його синонімів подається тільки відсилка на цю статтю, що ускладнює її пошук і знов-таки суперечить принципам організації WordNet. Більше того, як виявив аналіз, деякі з наявних у синсетах синонімів взагалі не представлено в реєстрі словника навіть у вигляді відсилання.

Тому внутрішню структуру словникових статей було проаналізовано та виявлено правила її побудови, які узгоджуються з лексикографічною моделлю ССУМа, але розкривають його структуру більш детально (що можна вважати непрямым використанням принципу лексикографічної редукції). Вони характеризуються певними поліграфічними ознаками, які дозволяють виділити в тексті статті її окремі елементи. Правила структурування словникової статті є такими:

- стаття може складатися з кількох блоків, що позначаються цифрами 1, 2, 3 і т. д. У такій статті заголовкове слово є полісемічним, і кожне його значення разом із синонімами являє собою окремий синсет;
- стаття може мати римську нумерацію: I, II і т. д. У такий спосіб у словнику подаються статті до омонімічних слів;
- синонім може мати фонетичні варіанти, які відрізняються написанням. Такі варіанти подаються в квадратних дужках після більш поширеної форми (наприклад: **ЗІГРІВАТИ [ЗОГРІВАТИ]**). Якщо синонім є словосполученням, що складається з кількох слів, то можливі варіанти, в яких деякі з цих слів пропусकाються;
- синонім може мати стилістичний або граматичний коментар, який подається після нього курсивом (наприклад: **ЗА-**

ГРІВАТИ *діал.*). Якщо синонім має варіанти, то коментар може подаватися до кожного варіанту окремо (відразу після нього), або до всіх в цілому (після всіх).

- після коментаря в дужках може подаватися стисле тлумачення або пояснення до синоніму (наприклад: **ГРІТИ** (на вогні й т. ін. — робити теплим, гарячим); **ПРОГРІВАТИ** (повністю));
- якщо синсет представлений дієсловами недоконаного виду, то після всіх синонімів та позначки “Док.:" перераховуються відповідні їм дієслова доконаного виду. Так само, для дієслів доконаного виду відповідні недоконані перераховуються після позначки “Недок.:". Ці синоніми також можуть мати коментарі та тлумачення;
- після всіх синонімів наводиться спільний для них блок ілюстрацій;
- після ілюстрацій може бути блок порівняння, що позначається “Пор.”, де наводяться реєстрові слова синсетів, які є семантично подібними до даного.

Було розроблено та створено ЛБД для відображення інформації про синоніми та синсети, яка має включати всі елементи згідно з перерахованими вище правилами. Зазначена ЛБД має структуру, зображену в табл. 6.1.

У таблиці поля мають такий зміст:

- **ID** — унікальний ідентифікатор синоніма;
- **ID_Set** — ідентифікатор синсета;
- **ID_Syn** — порядковий номер синоніма в синсеті;
- **Word** — синонім;
- **Comm** — стилістичний або граматичний коментар;
- **Comm2** — додатковий коментар до фонетичного варіанта синоніма;
- **Interpr** — стисле тлумачення або пояснення;
- **Sign** — ознака, що характеризує синонім;
- **ID_R** — “римський” порядковий номер статті;
- **ID_Int** — порядковий номер синсета в статті;
- **Digit** — цифровий код синоніма для сортування (в цьому коді українські літери від “а” до “я” замінюються на двоцифрові коди від “01” до “33”, всі інші символи ігноруються);
- **Nom** — унікальний ідентифікатор слова відповідно до реєстру граматичного словника.

Таблицю проіндексовано за полями: **ID** (Unique), **ID_Set**, **ID_Syn**, **Word**, **ID_R**, **ID_Int**, **Digit**, **Nom**.

Таблиця синонімів ("Synsets")

ID	ID_Set	ID_Syn	Word	Comm	Comm2	Interpr	Sign	ID_R	ID_Int	Digit	Nom
1	1	1	АБОРИГЕН< <57->	<> нини< <?-> > < >		{представник< <128255-> корінного< <65872-> населення< <84290-> країни< <66689->, місце- вості< <77175->}< <В>	0	0	0102192111040718	97	
2	1	2	АВТОХТОН< <747->	<> нини< <?-> > < >			0	0	0103231926231918	747	
3	1	3	ТУЗЕМЕЦЬ< <68786->	<> ділю< <140829->< < >			0	0	2324100117072731	198786	
4	1	4	ТУШЛЕЦЬ< <168707->	<> заст< <?->			0	0	2324021216072731	168707	
5	2	1	АБО#РТ< <99->			{передачне< <105403-> припіння< <130510-> вагітності< <11581->},	0	0	0102192123	99	
6	2	2	ВИ#ВИДЕНЬ< <15462->				0	0	0311151106071831	15462	
7	3	1	АБРИКО#СА< <122->			{плодове< <114034-> дерево< <34155->, кісто- чковий< <62811-> плід< <113958-> цього< <179933-> дерева< <34155->},	0	0	01022211115192201	122	
8	3	2	АБРИКО#Ж< <121->			{плодове< <114034-> дерево< <34155->, кісто- чковий< <62811-> плід< <113958-> цього< <179933-> дерева< <34155->},	1	0	01022111151922	121	
9	3	3	ЖЕРДЕ#ЛЯ< <42289->				0	0	09072106071833	42289	
10	3	4	МОРЕ#ЛЯ< <78191->			{с< < > дрібними< <38854-> плодами< <113958->},	0	0	171921071633	78191	

Детальніше розглянемо поле **Sign**. Залежно від того, до якого блоку належить синонім, воно може приймати значення:

- 5, якщо синонім входить до блока порівняння;
- 10, якщо синонім подається у блоці дієслів доконаного виду;
- 20, якщо синонім подається у блоці дієслів недоконаного виду;
- до поточного значення додається 1, якщо синонім є фонетичним варіантом.

Таблиця 6.2.

Таблиця ілюстрацій ("III")

ID	Illustr
1	Серед<!--151144--> наших<!--85709--> пращурів<!--128169--> були<!--11373--> пришельці<!--132191--> й<!--!--> аборигени<!--97--></I> (I<!--!-->. Цюпа<!--180984-->); <I>Слов'яни<!--154826--> не<!--85871--> тільки<!--!--> були<!--11373--> автохтонами<!--747--> на<!--79204--> цій<!--179933--> території<!--164485-->, а<!--!--> й<!--!--> сягають<!--162353--> генетичним<!--182201--> корінням<!--65874--> ще<!--379354--> в<!--!--> перше<!--108960--> тисячоліття<!--165412--> до<!--!--> нашої<!--85709--> ери<!--41141--></I> (з<!--!--> журналу<!--42956-->); <I>Розповідають<!--145076-->, що<!--377627--> під<!--!-->109945--> час<!--181287--> одного<!--95827--> із<!--!--> своїх<!--150050--> плавань<!--?--> Христофор<!--179272--> Колумб<!--64273--> був<!--!-->11373--> оголошений<!--95294--> туземцями<!--168786--> мало<!--73521--> не<!--85871--> богом<!--9369--></I> (з<!--!--> газети<!--26778-->); <I>П'ятеро<!--137528--> чорношкірих<!--182650--> тубільців<!--168707--> ретельно<!--139911--> працювали<!--128159--></I> (O<!--?-->. Донченко<!--37534-->).
2	<!--> <I>Мушиш<!--78947--> пам'ятати<!--102316-->, що<!--377627--> після<!--113180--> першого<!--108960--> аборту<!--99-->, таке<!--162647--> часто<!--181341--> трапляється<!--167293-->, ніколи<!--90746--> більше<!--8564--> важкою<!--11668--> не<!--85871--> станеш<!--158930-->...</I> (Є<!--11373-->. Гуцало<!--32429-->); <I>Спори<!--157791--> сажки<!--148445--> отруйні<!--100967--> для<!--35988--> тварин<!--163398--> і<!--!--> викликають<!--15519--> викидні<!--15462--> у<!--!--> корів<!--65959--> та<!--!--> свиней<!--149921--></I> (з<!--!--> посібника<!--125154-->).
3	<I>Цвігуть<!--179857--> в<!--!--> Херсоні<!--178000--> абрикоси<!--121--></I> (Л<!--?-->. Дмитерко<!--35994-->); <I>Марта<!--74372--> Кирилівна<!--62271--> задумалась<!--45218--> і<!--!-->, йдучи<!--59482--> по<!--114375--> доріжці<!--37878-->, зачепила<!--51285--> головою<!--30141--> гілку<!--28998--> абрикоса<!--121--></I> (I<!--!-->. Нечуй<!--90294--> Левицький<!--90294-->); <I>Єгипта<!--?-->.. каже<!--59743-->, щоб<!--377629--> не<!--85871--> забувала<!--43421--> і<!--!--> що<!--377627-->він<!--22257--> їй<!--24451--> із<!--!--> Тарасівки<!--?--> привез<!--128822--> повний<!--115933--> кіш<!--62842--> абрикос<!--122--></I> (O<!--181857--> та<!--!--> жерделі<!--42289--> і<!--!--> тане<!--163015--> день<!--34068--> в<!--!--> заметі<!--47099--> запашнім<!--47926-->,.. гойдають<!--30054--> весни<!--13856-->мої<!--76629--> пісні<!--113241--></I> (H<!--?-->. Забіла<!--43210-->); <I>В<!--!--> цвіту<!--179859--> гойдаються<!--30055--> морелі<!--78191--> Під<!--109945--> дзвін<!--34788--> працюючі<!--168472--> бджоли<!--6354--></I> (M<!--?-->. Стельмах<!--159128-->).

У цій таблиці поля мають такі значення:

- **ID** — ідентифікатор блока ілюстрацій, який співпадає з ідентифікатором синсета;
- **Illustr** — текст ілюстрацій.

Таблицю проіндексовано за полем **ID** (Unique).

Розроблена в УМІФі комп'ютерна технологія забезпечує автоматичне формування ЛБД з вищенаведеною структурою, виходячи з електронного тексту ССУМа або з бази даних ІІС “Словники України”. При цьому потреба у статтях відсилкового типу взагалі відпадає, оскільки для кожного елемента синсету тепер є можливість одержання повноцінної лексикографічної інформації. Реєстр новоствореної ЛБД визначається набором послідовностей синонімів, які входять до складу словникових статей.

При аналізі внутрішньої структури статті на першому кроці відбувається розділення її на блоки: звичайні синоніми, блок докона-ного/недоконаного виду, ілюстрації та блок порівняння. На наступному кроці аналізується кожний блок для подальшого розкриття елементів структури. Синоніми визначаються як позначені жирним шрифтом, варіанти — за квадратними дужками, коментарі — за курсивом, тлумачення — за круглими дужками. Внутрішня структура ілюстрацій не аналізується, їхній блок заноситься “як є” до поля **Illustr** таблиці III.

Значення ідентифікаційних полів таблиці **Synsets** визначаються послідовно в процесі аналізу. При цьому **ID_Set** нарощується при переході до наступної статті або блоку з наступним номером за нумерацією в межах статті. Домінанта синонімічного ряду завжди отримує номер 1, який записується в **ID_Syn**, але тепер вона вже не має “головного” статусу і є рівноправною серед інших слів цього ж синсета. Зв'язок з таблицею ілюстрацій здійснюється через поле **ID_Set** — його значення записується в поле **ID** таблиці III для відповідного даному синсету блоку ілюстрацій. Оскільки кожна стаття аналізується послідовно, то порядок слідування синонімів у синсеті завжди такий: звичайні синоніми, синоніми доконаного або недоконаного виду (якщо є), порівняльні синоніми (якщо є).

Під час роботи алгоритму виникали деякі ускладнення. Наприклад, синонім до “будь-де”, представлений в словнику у вигляді “ДЕ Б [ТО (ТАМ)] НЕ БУЛО [НЕ БУДЕ]”, як видно з цього запису, може мати декілька варіантів в залежності від наявності або відсутності “ТО (ТАМ)” та варіації “НЕ БУЛО” / “НЕ БУДЕ”. Таких складних випадків небагато, але вони не можуть бути проаналізовані автоматично, тому їх довелося виявляти та розглядати окремо.

В результаті було отримано ЛБД, в таблиці **Synsets** якої кожний запис представляє синонім з його властивостями: належність до певного синсету та його місце в цьому синсеті згідно з вищенаведеними ознаками. Зауважимо, що синоніми доконаного/недоконаного виду та порівняльні, представлені у словниковій статті, фактично належать до інших синсетів, але наводяться у тому самому синсеті з відповідними ознаками для збереження зв'язків.

У наведених вище фрагментах таблиць можна побачити, що в усіх текстових полях за кожним словом слідує HTML-коментар — число, яке є унікальним ідентифікатором слова в реєстрі граматичного словника. Ці коментарі були введені до ЛБД під час індексації при підготовці електронного варіанту словника синонімів. Найбільший інтерес в даному випадку представляють ідентифікатори, що стоять безпосередньо після синонімів у полі **Word**. Для тих синонімів, що представлені окремими словами, такі ідентифікатори було виділено, перетворено у цифрову форму та збережено в полі **Nom**. Таким чином, значення цього поля може використовуватися для інтеграції отриманої ЛБД з іншими електронними словниками, створеними в УМІФі.

Для цієї ЛБД було написано клієнтську програму, яка буде та виводить на екран словникові статті згідно з поліграфічними правилами, що мали місце у вихідному словнику синонімів. Але, на відміну від колишнього словника, в новому варіанті відпадає потреба у відсилкових статтях, тому що для кожного синоніма легко може бути отриманий синсет, до якого він належить, та побудовано статтю для цього синсету. Якщо слово є полісемічним та входить до кількох синсетів, то їх можна вивести послідовно один за одним. Отже, така база даних зберігає всю інформацію, яка була у вихідному словнику синонімів, але вміщує цю інформацію у більш структурованому та придатному для подальшої обробки вигляді і доступ до неї значно спрощується.

Програма перегляду нового словника синонімів також дає можливість представити на екрані внутрішню структуру синсету, відобразити окремі його елементи та зв'язки між ними.

Алгоритм формування словникових статей із синсетів для виводу їх на зовнішній пристрій, розроблений для даної програми, було використано також для побудови статей словника синонімів ІЛС "Словники України", починаючи з версії 1.2. Оскільки при цьому були збережені всі реєстрові ідентифікатори слів, то отримана на виході база даних залишилася проіндексованою, тобто кожне слово нового словника синонімів так само є точкою входу для системи,

як це було у попередній версії ЛДС “Словники України”. Новий словник є більш прозорим та зручним у використанні; крім того, його структура гарантує, що всі синоніми, наявні в словникових статтях, обов’язково присутні в реєстрі.

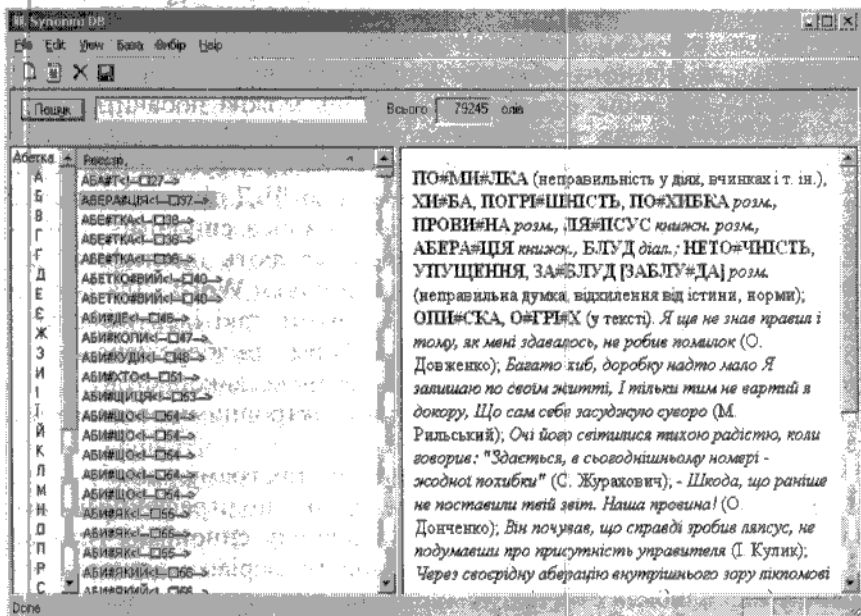


Рис. 6.1. Програма перегляду словника синонімів

6.2.3. Інтеграція синонімічної ЛБД з ЛБД СУМа

Наступним кроком стало створення зв’язку між синонімічною ЛБД та ЛБД СУМа так, щоб вони мали семантичний зв’язок через словникові дефініції (формули тлумачення). Для цього необхідно, по-перше, введення додаткових полів до структури обох БД. По-друге, такий зв’язок не може бути створений автоматично, тому що СУМ та Словник синонімів уклалися практично незалежно, хоча при укладанні останнього СУМ, безперечно, використовувався. Тому виникає необхідність створити інтерфейс редагування зв’язків між словниковими статтями.

Нагадаємо, що всі формули тлумачення в ЛБД СУМа незалежно від того, відносяться вони до власне тлумачень або їх відтінків, фразеологізмів, еквівалентів слів тощо представлені в таблиці

formula. Звідси випливає, що саме з цією таблицею необхідно зв'язати ЛБД синсетів.

Але спочатку над таблицею Synsets було виконано ряд перетворень. Принцип побудови WordNet передбачає, що кожному синсету має бути приписане тлумачення, спільне для всіх його елементів. Проте у побудованій нами структурі БД є тільки поле тлумачення для окремих синонімів. При цьому структурою словника синонімів, на основі якого була побудована ЛБД, визначається, що тлумачення до домінанти в ряді насправді відноситься до всього синонімічного ряду. Зрозуміло, що такий підхід є незручним, тому було вирішено ввести до таблиці *Illustr*, кожний запис якої відповідає цілому синсету, поле **Interpr**, до якого й перенести значення тлумачення з однойменного поля для перших в рядах синонімів таблиці Synsets. Після такого перетворення домінанти синсетів втрачають статус "головних" і можуть мати свої додаткові тлумачення, як і всі інші.

Наступні перетворення призначені для того, щоб виділити синсети коректніше, ніж це було зроблено при першому формуванні ЛБД:

1. Синоніми, що належать до блоків доконаного або недоконаного виду, виносяться в окремі синонімічні ряди.

2. Синоніми, що належать до блоків порівняння, виключаються із синсетів (фактично вони взагалі не є синонімами і були включені до ЛБД тільки для сумісності з вихідним Словником синонімів).

3. Синоніми-фонетичні варіанти пов'язуються з основними варіантами написання через додатково введене до таблиці ідентифікаційне поле **ID_Phon**. Наявність такого зв'язку дозволяє контролювати, щоб значення тлумачення фонетичного варіанту завжди співпадало з тлумаченням основного варіанту. При цьому додаткові коментарі з поля *Comm2* поєднуються з коментарями *Comm* до основних варіантів; таким чином, поле *Comm2* стає надлишковим.

4. Приймається за правило, що крапка з комою в синонімічному ряді розділяє різні синсети; при цьому домінанта ряду належить до кожного з цих синсетів (тобто є полісемічною). Так, наприклад, зображена на рис. 3.2 стаття при застосуванні цього правила розпадається на три з синонімічними рядами відповідно:

- ПОМІЛКА, ХИБА, ПОГРІШНІСТЬ, ПОХИБКА, ПРОВІНА, ЛЯПСУС, АБЕРАЦІЯ, БЛУД;
- ПОМІЛКА, НЕТОЧНІСТЬ, УПУЩЕННЯ, ЗАБЛУД, [ЗАБЛУДА];
- ПОМІЛКА, ОПІСКА, ОГРІХ.

Блок ілюстрацій також необхідно було б розбити на три частини, але цю операцію майже неможливо виконати автоматично, тому на цьому етапі весь блок копіюється до кожного з трьох новостворених синсетів.

Крім того:

- в таблицю ілюстрацій було добавлено поле **Part**, що визначає частину мови синсета;
- в таблицю Synsets добавлені поля **Hyperonym** та **ID_Hyp**, що визначають гіперонім до синоніма та його ідентифікатор. На даному етапі ці поля не заповнюються, але вони введені для того, щоб уможливити подальше створення в ЛБД синсетів ієрархічних зв'язків, які мають місце в структурі WordNet.

Далі було введено поля для зв'язку таблиць:

- **IntSum** в таблиці Synsets ЛБД синонімів для зберігання ідентифікатора формули тлумачення, з якою пов'язаний синонім;
- **Synset** в таблиці formula ЛБД СУМа для зберігання ідентифікатора синсета, з яким пов'язана формула тлумачення.

Отримані в результаті зв'язки можна представити схемою, зображеною на рис. 6.2.

Слід зауважити, що реалізація таких зв'язків саме і означає інтеграцію ЛБД СУМа та синсетів до єдиної ЛБД. Допоміжним кроком до такої інтеграції стало копіювання таблиць Synsets та III до ЛБД СУМа.

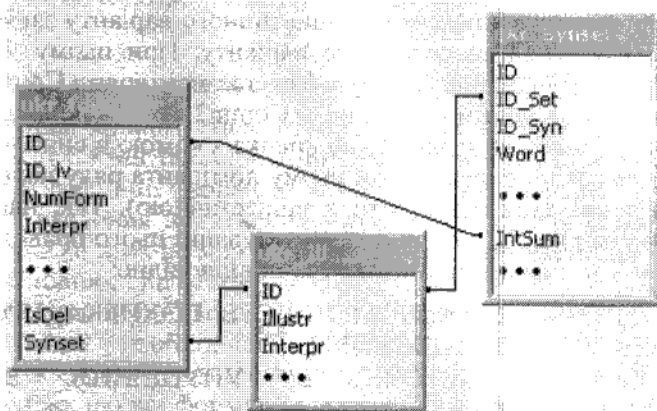


Рис. 6.2. Зв'язки між таблицями ЛБД СУМа та ЛБД синсетів

- Для того, щоб зв'язати формулу тлумачення з відповідним синонімічним рядом, необхідно виконати наступні дії:
- знайти всі синсети, до яких входить реєстрове слово СУМа з вибраною формулою тлумачення;
 - вибрати з цих синсетів той, значення якого відображає ця формула тлумачення.

Для реалізації таких дій в програмі редагування СУМа було розроблено діалогове вікно, приклад якого наведений на рис. 3.4.

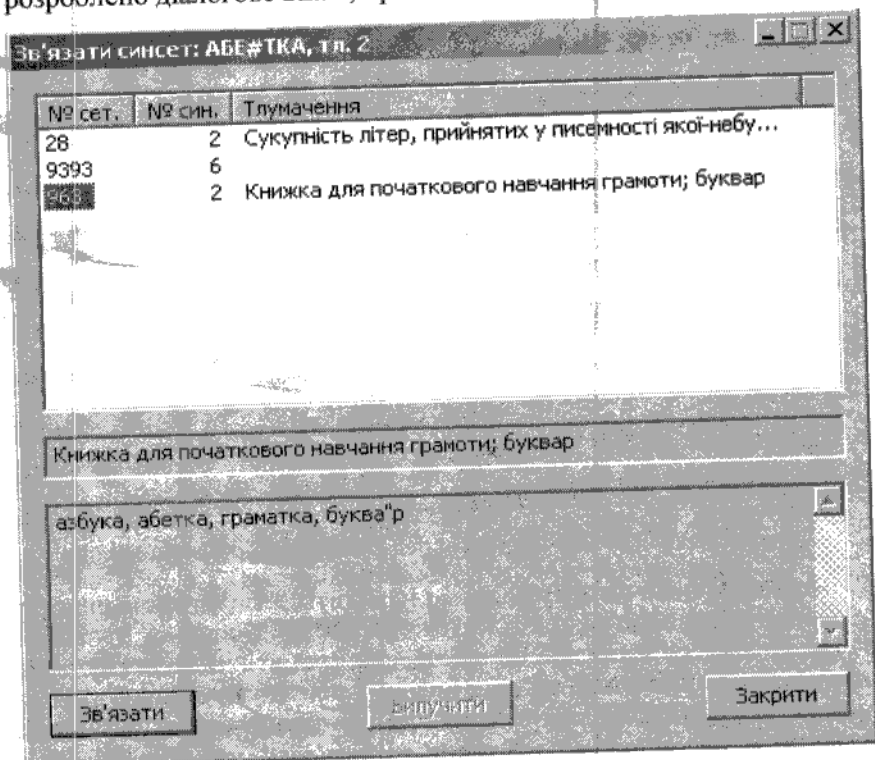


Рис. 6.3. Встановлення зв'язку між формулою тлумачення та синсетом

Також було реалізовано зворотній зв'язок, який дозволяє отримати формулу тлумачення з СУМа для конкретного синоніма. Але спочатку програму перегляду синонімів було розширено та доповнено функціями редагування.

Виявилось, що поліграфічна форма представлення синонімічної статті, досить зручна і компактна при використанні її в паперовому

виданні, в той же час є неоптимальною при відображенні статті для технологічного редагування, тому що не дає можливості чітко розрізнити структурні елементи. Щоб уникнути цього недоліку, було розроблено спеціальний формат візуального представлення синонімічних статей. Приклад зовнішнього вигляду статті у такому форматі наведено на рис. 6.4.

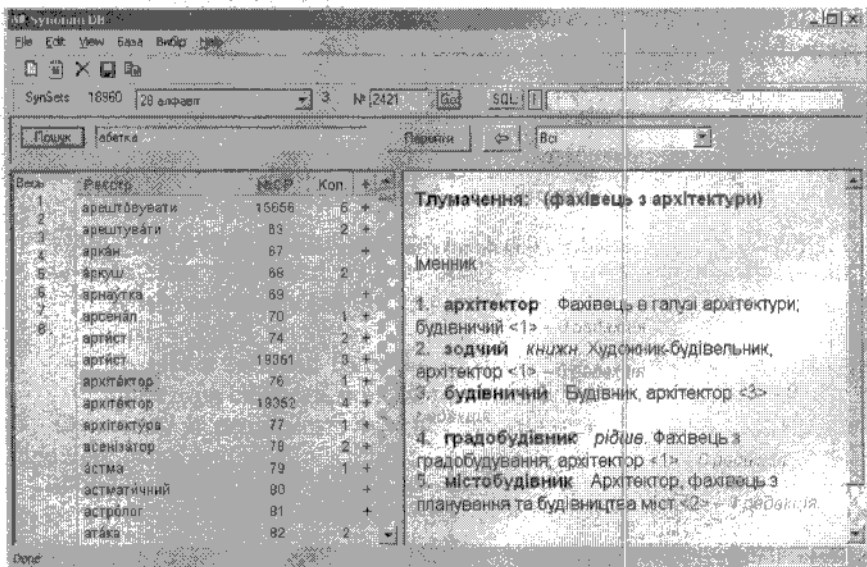


Рис. 6.4. Головне вікно програми редагування синонімічних рядів

Як можна побачити, тут окремо наводиться тлумачення синсета та частина мови, до якої він належить. Кожний синонім дається з нового рядку, де чітко представлені його граматичні та стилістичні параметри, тлумачення, зв'язок з формулою тлумачення СУМА та номер останньої редакції синоніма. При цьому виділене жирним шрифтом поле синоніма є активним і дозволяє створити зв'язок з формулою тлумачення СУМА через діалогове вікно, зображене на рис. 6.5. Текст вибраної формули тлумачення заноситься до відповідного поля таблиці Synsets або III, залежно від того, необхідно його встановити для всього синсета або тільки для вибраного синоніма.

В програмі редагування реалізовано ряд функцій, зокрема:

- додавання, вилучення, копіювання та редагування синсета;
- пошук синсета чи групи синсетів за словом, яке до них входить, або за ідентифікатором;

- фільтрація реєстру: режим “всі синоніми” або “тільки синсети”, вибір технологічного тому, а також можливість встановлення фільтра за частиною мови або довільним SQL-запитом;
- вибір режиму відображення словникової статті;
- відображення стану редагування синсета.

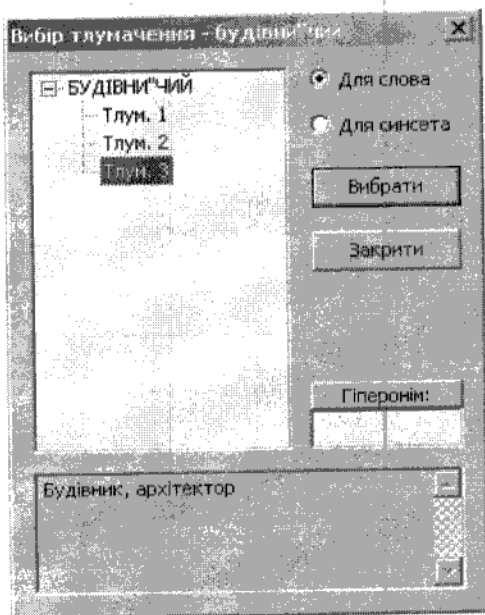


Рис. 6.5. Поєднання синоніма з формулою тлумачення.

Також вікно перегляду структури синсета було перетворене у вікно редагування та значно розширене. Представлений ним інтерфейс дозволяє виконувати такі дії:

- додавання, видалення та редагування синонімів, що входять у синсет;
- переставлення синонімів в будь-якій послідовності;
- редагування тлумачення синсета або вибраного синоніма, граматичних та стилістичних параметрів;
- редагування значення синоніма у однині, якщо його основна форма представлена множиною;
- редагування гіпероніма до синоніма;
- встановлення ознаки фонетичного варіанту;
- редагування блоку ілюстрацій до статті.

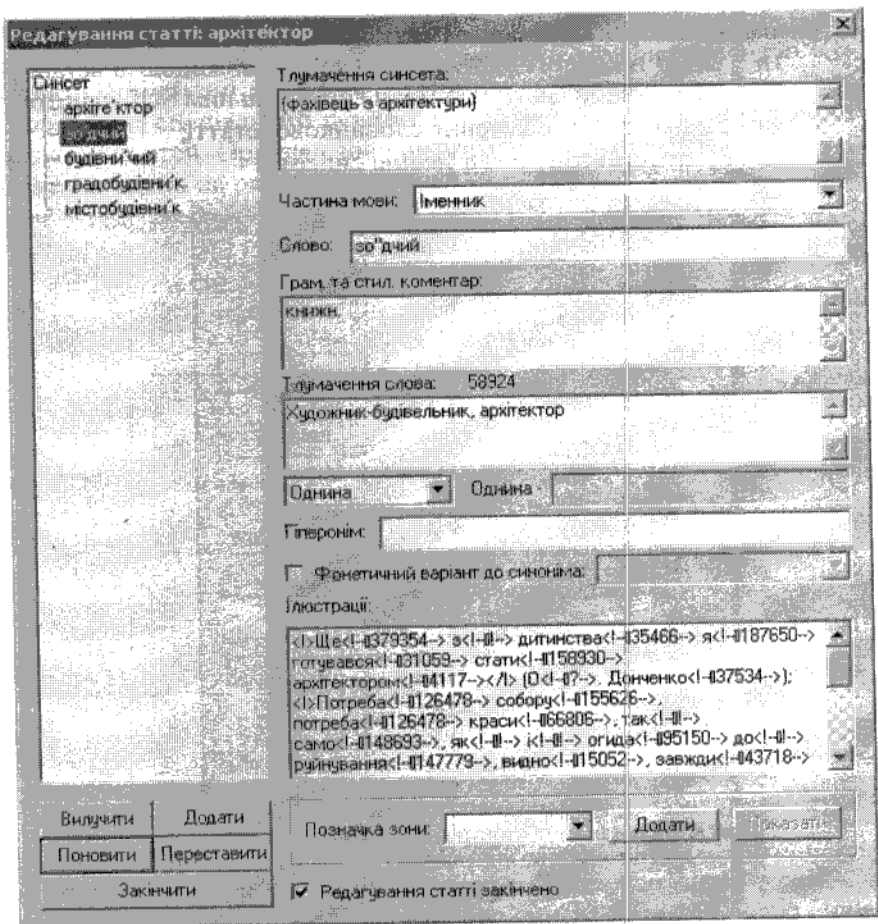


Рис. 6.6. Редагування словникової статті

Отримана інтегрована ЛБД СУМа та синонімічних рядів має стати основою для побудови української системи WordNet. Завдяки наявності поля частини мови для кожного синсета стає можливим структурувати базу даних за частинами мови так, як цього вимагає структура WordNet. Далі ЛБД може бути розширена шляхом введення до неї ряду зв'язків між словами та синонімічними рядами, кінцевим результатом чого повинна стати ЛБД, придатна до безпосередньої інтеграції у WordNet та EuroWordNet.

6.3. Системотехніка двомовної лексикографії

Завдяки наявності комп'ютерних реалізацій Л-систем, викладених у попередніх розділах, постає можливість формування спеціального інструментального комплексу для підтримки процесів фундаментальної двомовної лексикографії. Тут ми схематично наведемо приклад такого інструментального комплексу, який формується в Українському мовно-інформаційному фонді НАН України в рамках спільного з Інститутом російської мови Російської академії наук проекту створення великого комп'ютерного російсько-українського словника (Большой Русско-Украинский Словарь — БРУС).

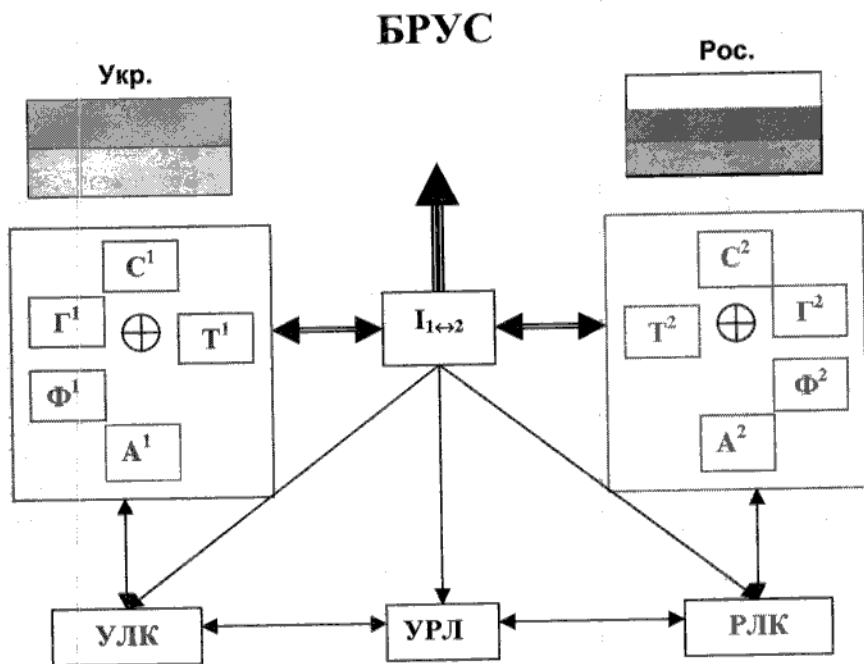


Рис. 6.7. Схема інструментального комплексу БРУС

На схемі рис. 6.7. використано такі позначення:

- T^1, T^2 — тлумачні словники у мовах 1, 2 відповідно;
- C^1, C^2 — словники синонімів у мовах 1, 2 відповідно;
- Γ^1, Γ^2 — граматичні словники у мовах 1, 2 відповідно;
- Φ^1, Φ^2 — фразеологічні словники у мовах 1, 2 відповідно;
- A^1, A^2 — словники антонімів у мовах 1, 2 відповідно;

⊕ — операція інтеграції Л-систем T, C, Г, Ф, А;

$I_{1 \leftrightarrow 2}$ — Л-система — інтерфейс між 1 і 2;

УЛК, РЛК, УРЛК — український, російський і паралельний українсько-російський лінгвістичні корпуси. Концепція, лінгвістичні та системотехнічні засади побудови та використання лінгвістичних корпусів докладно викладено у нашій монографії⁶, де наведено також і приклади застосувань інструментарію лінгвістичного корпусу в конкретних мовознавчих дослідженнях.

РУС	ТОЛКОВАНИЕ	ТЛУМАЧЕННЯ	УКР
X (РУС_РЕВСТР_ЕД)	C(X)	C(Y)	Y (УКР_РЕЕСТР_ОД)
G(X)			G(Y)
РЕМАРКИ			РЕМАРКИ
СИНСЕТ S(X)			СИНСЕТ S (Y)
ГИПЕРОНИМ H(X)			ГПЕРОНИМ H (Y)
ГИПОНИМ h (X)			ГПОНИМ h (Y)
АНТОНИМ: A(X)			АНТОНИМ A (Y)
УПРАВЛЕНИЕ			КЕРУВАННЯ
СЛОВСОЧЕТАНИЯ			СЛОВСПОЛУЧЕННЯ
СЛОВООБРАЗОВАТЕЛЬНОЕ ГНЕЗДО			СЛОВТВІРНЕ ГНІЗДО
АВТОР		ДАТА	

Рис 6.8. Формат для укладання словникових статей БРУС

Репертуар окремих словників, звичайно, є відкритою множиною і буде розширятися і доповнюватися по мірі потреби. Для укладачів БРУС буде надано формат (Рис. 6.8.), який вони заповнюватимуть, звертаючись в інтерактивному режимі до Л-систем, наведених на Рис. 6.7. При цьому вибір лексикографічних структур зі згаданих Л-систем інструментального комплексу відбувається автоматично.

⁶ В.А.Широков та ін. Корпусна лінгвістика. — К.: Довіра, 2005, 502 с.

У такий спосіб буде створено комп'ютерний масив структурованої лексикографічної інформації, з якого словники, необхідні для конкретних застосувань в системах опрацювання природної мови, формуватимуться вже цілком автоматично за допомогою спеціально розроблених процедур.

6.4. Системи лінгвістичного діалогу

6.4.1. Всеукраїнський лінгвістичний діалог в контексті теорії Л-систем

Об'єднавчі процеси, характерні для сучасного періоду розвитку світової цивілізації, активізували дві суперечливі тенденції в мовній сфері: з одного боку, зумовлені парадигмою інформаційного суспільства явища уніфікації мовно-інформаційних процесів з домінуванням вузького кола так званих світових мов, а з другого — природне прагнення кожного народу до самоідентифікації, захисту, збереження й розвитку власного національно-культурного начала. Саме мовна сфера, що покликана обслуговувати всі аспекти суспільної практики у самій державі й водночас становити основу для забезпечення необхідних комунікативних функцій на міжнародному рівні, подекуди стає ареною драматичних колізій між відзначеними тенденціями.

Останнім часом об'єктивно зростають передумови для обмеження сфери застосування рідної мови на користь інших мов, які мають більше поширення та міжнародний статус. Тим часом досвід переконує, що в тих державах, де недооцінюються мовні проблеми, завжди існує й відповідна внутрішня та зовнішня напруженість. Тому кожна країна, яка не бажає перебувати у колі світових аутсайдерів, опікується розвитком своєї мови (мов) і вживає необхідних заходів щодо підтримки її (їх) належного статусу. Відносно новий аспект мовного статусу, що надзвичайно актуалізувався останніми роками, а саме — технологічний, пов'язаний із застосуванням механізмів природної мови в інформаційних технологіях⁶. Системи природномовного пошуку, машинного перекладу, безклавіатурного введення мовної інформації до комп'ютерів, автоматичного редагування, реферування та індексування, природномовні інтерфейси до інформаційно-комп'ютерних систем різного призна-

⁶ Широков В.А. Гуманітарна традиція і технологічний статус мови // Мовознавство.— 2001.— № 3.— С.120 — 132.

чення тощо все наполегливіше набувають ознак обов'язковості у програмному інструментарії комп'ютерів.

Усвідомлення мови як специфічного природного феномену спонукає при проведенні досліджень дотримуватися певного методу, спрямованого на якомога глибшу об'єктивізацію досліджуваних мовних явищ та їхню формалізацію. Прагматичний компонент, що спонукає до створення конкретних лінгвістичних продуктів, визначає необхідність розробки й застосування якомога ефективнішого інформаційно-технологічного інструментарію лінгвістичного дослідження.

Розроблення та створення комп'ютерних систем зазначеного класу потребує відповідної лінгвістичної інформації, яка не лише зосереджена в колі опису власне фактів мови (це було й раніше), а й вимагає наявності ефективних, формалізованих моделей мовної та лінгвістичної компетенції людини, що ставить перед мовознавцями суттєво нові завдання.

Зазначимо, що наведені вище напрями природномовної інформатики, які бурхливо розвиваються у світі впродовж останніх двох десятиліть, досі майже не були представлені в україномовному варіанті. Основною причиною такого стану є, на нашу думку, недостатня розвиненість національної словникової бази, що становить одну з ознак і загальнокультурного, і технологічного статусу мови. Незважаючи на природність даного твердження та світовий досвід науково-технічні здобутки української мови з цього напрямку поки що слід оцінювати як досить скромні. Численні приклади з розвитку фахової лексикографії підтверджують, що за репертуаром і кількісними та якісними показниками українська лексикографія поки що відстає, наприклад, від англійської, німецької, французької, російської й не повністю відповідає сучасним потребам. А через відсутність багатьох типів словників досі неможливе і створення комп'ютерних лексикографічних систем, що програмує відставання в галузі вітчизняної лінгвістичної технології взагалі.

Подолати це відставання покликана програма створення національної словникової бази, ініційована Указом Президента України від 7 серпня 1999 р. № 967, яка передбачає формування цілісної та універсальної системи української лексикографії, заснованої на сучасних концептуальних лінгвістичних та технологічних засадах⁷. Ядром відзначеної системи виступає серія академічних словників

⁷ Широков В.А., Манаків В.В. Організація ресурсів національної словникової бази // Мовознавство.— 2001.— № 5.— С. 3—13.

нового покоління — “Словники України”, близько 50 випусків якої вже побачило світ. Завершується й робота над укладанням нового 20-томного тлумачного Словника української мови⁸. Але не дарма побутує вислів, що “велика книга — це велике зло”⁹. Стосовно багатотомного словника це означає не лише те, що в ньому подекуди буває важко знайти необхідну інформацію, а й що такі видання зазвичай застарівають ще до їхнього виходу. І хоч новий тлумачний словник буде завершено у надзвичайно короткі строки, очевидно, що й він не встигне за швидкоплинним життям. Крім цього, ми не бачимо рації у “демонтажі” науково-організаційного, а тим паче технологічного апарату укладання словника після завершення роботи над ним — навпаки, вважаємо за доцільне продумати комплекс заходів щодо його збереження та розвитку.

На нашу думку, існує унікальний шанс реалізації варіанта саморозвинної лексикографічної системи тлумачного словника української мови, здатної до еволюції, “відстеження” в реальному часі розвитку самої мови і збагачення новими мовними фактами.

Основна ідея полягає у створенні лексикографічної комп'ютерної системи, що функціонуватиме в середовищі Інтернету, забезпечуючи співпрацю лінгвістів різних мовознавчих установ України через уніфікований інтерфейс і на єдиних наукових та організаційних принципах. Реалізація цієї ідеї є наслідком теорії лексикографічних систем, зокрема так званих віртуальних лексикографічних систем (ВЛС), що є основою для системотехнічної розробки віртуальної лексикографічної лабораторії — лінгвістично-технологічної структури, спроможної об'єднати всі зацікавлені фахові словникарські колективи України. Викладемо основні положення системотехніки ВЛС.

6.4.2. Віртуальні лексикографічні системи

Поняття віртуальної лексикографічної системи, введене в теорії лексикографічних систем, визначається у наступний спосіб.

У процесі функціонування агрегованих лексикографічних систем в реальних соціотехнічних середовищах може відбуватися процес віртуалізації лексикографічних систем. Це буває, коли

⁸ Концепція та технологія укладання нового Словника української мови викладені у працях: Русанівський В.М., Широков В.А. Інформаційно-лінгвістичні основи сучасної тлумачної лексикографії // Мовознавство. — 2002. — № 6. — С. 7 — 48; Широков В.А., Рабулець О.Г., Костишин О.М., Шевченко І.В., Якименко К.М. Технологічні основи сучасної тлумачної лексикографії // Там же. — С.49 — 86.

⁹ The Oxford English Dictionary. — Oxford, 1970. — V. 1: A — B. — P. VII.

“предметна галузь” (X) — носій суперсистеми лексикографічних систем — має розподілені системні характеристики і параметризована певною структурованою множиною системних (мережевих) адрес, на кшталт, наприклад, таких, що прийняті в Інтернеті. Тоді кожен елемент $x \in X$ постає функцією від певного кортежу адрес:

$$x = f(a_1 a_2 \dots a_n), \quad (6.1)$$

тобто лексикографічна система ELX/L перетворюється на віртуальний об'єкт, розподілений у фізичному просторі, що зображається точкою (a_1, a_2, \dots, a_n) у просторі мережевих адрес. При цьому узгодження відповідних моделей даних на концептуальному та внутрішньому рівнях навіть не вимагається (хоча для зручності таке узгодження було б, звичайно, вельми корисним). Необхідне лише узгодження на рівні зовнішніх моделей, зокрема на рівні мережевих протоколів, що забезпечує мінімальну цілісність віртуальної лексикографічної системи й можливість її ідентифікації як єдиного об'єкта.

Таку ідею лексикографічної системи можна реалізувати, коли визначимо $(a_1 a_2 \dots a_n)$ як множину мережевих координат українського сегменту Інтернету, а ELX/L спорядимо структурою елементарної Л-системи, клас елементарних інформаційних одиниць якої ототожнимо з класом слів — наприклад, української мови.

Практична реалізація цієї ідеї передбачає виконання наступних етапів.

По-перше, визначення й локалізація множини $(a_1 a_2 \dots a_n \dots)$ мережевих координат українського сегменту Інтернету.

Зрозуміло, що ця множина є відкритою й динамічною, але необхідно визначити її системні характеристики. Як відомо, в мережі Інтернет існує два зв'язаних між собою методи адресації. Перший з них — цифровий метод, який було прийнято ще на самому початку розвитку TCP/IP мережевих технологій при створенні прототипу Інтернету — ARPANET. Для цифрового методу характерне представлення адреси певного мережевого вузла чотирма трійками чисел, наприклад, 194.45.132.231. Така нотація мережевої адреси зумовлена тим, що при розробці системи ARPANET було визначено максимальне загальне число вузлів мережі — 2^{32} . Тоді цифрові значення вузлів варіюють від 0.0.0.0 до 255.255.255.255.

Проте такий метод адресації є незручним для людського сприйняття і групою розробників Стенфордського університету (США) було запропоновано мнемонічний метод адресації, який надалі на-

був широкого розповсюдження й отримав назву служба доменних імен — Domain Name Service (DNS)¹⁰.

DNS — це розподілена база даних, яка дає можливість локального керування її окремими сегментами, а також зробити дані кожного сегмента бази доступними всій мережі. Цю властивість системи DNS реалізовано завдяки використанню клієнт-серверної технології. Надійність та продуктивність системи DNS забезпечується механізмами реплікації та кешування. Серверна частина клієнт-серверного механізму DNS представлена спеціалізованими програмами — *DNS-серверами* (*name servers* — серверами імен). DNS-сервери володіють інформацією про сегменти бази даних та забезпечують до неї доступ клієнтам — *пошуковим аналізаторам* (*resolvers*). Звичайний DNS-клієнт набір бібліотечних функцій, що реалізують запити й передають їх через мережу до сервера імен.

Цифровий та мнемонічний способи адресації пов'язані між собою, але цей зв'язок неоднозначний. Для прикладу розглянемо типовий фрагмент бази даних DNS-сервера.

Name	Type		Data
	NS		monster.umif.net.
	MX	[10]	thor.umif.net.
monster	IN	A	192.168.1.100
win	IN	A	192.168.1.11
nataly	IN	A	192.168.1.23
thor	IN	A	192.168.1.24
enigma	IN	A	192.168.1.27
www	CNAME		enigma.umif.net
www2	CNAME		enigma.umif.net
ftp	CNAME		enigma.umif.net

Як видно з таблиці, у базі даних сервера імен наявні три типи записів, а саме:

— спеціалізовані записи типу NS (*name server*) — сервери імен, MX (*mail exchanger*) — поштові сервери;

— записи, що відповідають адресам вузлів — IN A, які забезпечують трансляцію цифрової адреси у мнемонічну й навпаки;

— записи типу CNAME (т.н. *Canonical Names*), які забезпечують псевдонімізацію, що дозволяє надати одній цифровій адресі декілька символічних імен.

¹⁰ DNS и BIND. Пол Альбиц и Крикет Ли, O'Reilly, 2002.

З фрагменту бази даних сервера імен видно, що на комп'ютері, який має адресу 192.168.1.27, функціонує два сервіси http та один ftp-сервіс, що демонструє відзначену неоднозначність.

Таким чином, множиною $(a_1 a_2 \dots a_n)$ мережевих координат у розробленій моделі варто визначити саме множину мнемонічних адрес українського сегменту Інтернету.

Для ідентифікації україномовних систем на множині $(a_1 a_2 \dots a_n)$ в Українському мовно-інформаційному фонді розроблено спеціальну програмну процедуру — ідентифікатор української мови, яка дозволяє автоматично визначати тексти, написані українською мовою, і отже — україномовні сайти в Інтернеті. Ця програма, поперше, аналізує метадані вибраного ресурсу, а потім — і саме його текстове наповнення. Після цього тексти з виявлених таким способом україномовних сайтів автоматично індексуються за допомогою спеціального модуля індексування віддалених ресурсів “SPIDER_ua”. При індексації текст має розглядатися як послідовність слів, але з точки зору операційної системи він є послідовністю символів. Тому виникає необхідність у побудові алгоритму виділення слів у тексті. Такий алгоритм було розроблено та реалізовано в окремому модулі — “analyze”. Його основна функція — отримання наступного слова в тексті — саме і дає можливість працювати з текстом, як з послідовністю слів. Модуль “analyze” визначає не тільки саме слово, а й код, яким визначено його тип. Цей код показує, чи належить слово до мови, з якою працює процедура індексації (звичайно це українська мова), тобто чи складається воно з символів українського алфавіту та, можливо, дефіса і апострофа, або воно є іншомовним; чи воно є числом, пунктуаційним знаком, скороченням, кінцем речення, ініціалом, послідовністю символів певного діапазону, тощо. Далі при індексації обробляються тільки слова вибраної мови, всі інші ігноруються.

Структура текстів Інтернет-ресурсів, як правило, є нелінійною — крім безпосередньо тексту, вона містить метадані відповідно до стандартів HTML 4.01, XML 1.1 або XHTML 1.0¹¹. Крім того, й власне текст має внутрішню структуру — частини, розділи, параграфи і т. ін. Формальні ознаки, за якими автоматично визначаються ці структурні елементи, є стандартними.

У результаті виконання відзначених процедур в автоматичному режимі динамічно формується лексикографічна система $ELSx[U]$, класом ЕЮ якої є “УКРАЇНСЬКІ СЛОВА”, а інтерпретаційною частиною — їхні локалізації на відповідних сайтах.

¹¹ <http://www.w3.org/MarkUp/>

6.4.3. Віртуальна лексикографічна лабораторія

Створена у такий спосіб система виконуватиме функції електронного архіву української мови в Інтернеті й окрім розв'язання лексикографічних проблем, допомагатиме у фіксації історії розвитку україномовного сегменту Інтернету. У випадку коли реалізовано підтримку декількох мов, система дозволить забезпечити моніторинг розвитку їх мережевих сегментів. Однак при цьому вимоги, які накладаються на технічну базу (продуктивність системи обробки інформації, обсяги пам'яті і под.) стають значними, а у певних випадках — нездійсненними.

Розумним методом розв'язання відзначеної проблеми є розпаралелення системи на декілька підсистем, кожна з яких опрацьовує обмежену підмножину мовного сегмента мережі, причому зовсім необов'язково, щоби фізично ці підсистеми були локалізовані в одному місці. Зв'язавши їх в одну лексикографічну мережу й забезпечивши маршрутизацію зовнішніх мовних запитів до відповідної підсистеми, можна одержати інформацію з будь-якої групи мов, реалізованої в лексикографічній мережі.

Але така віртуальна лексикографічна система може бути використана і в іншому режимі, а саме — як віртуальна лексикографічна лабораторія, яка надає засоби для виконання спільних лексикографічних проектів фахівцями різних, територіально розподілених інституцій і навіть різних країн.

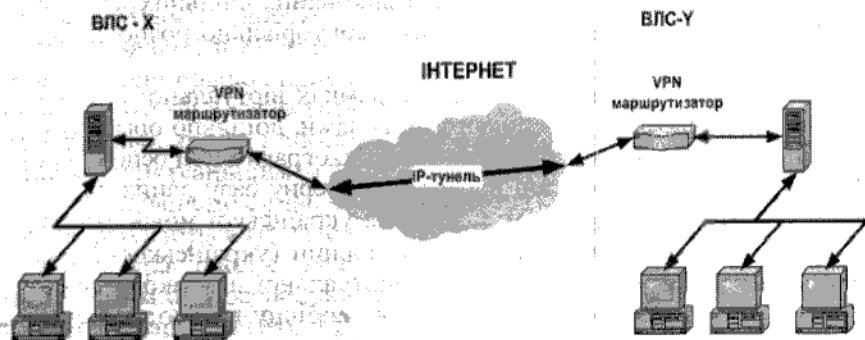
Спільну лексикографічну роботу в рамках віртуальної лексикографічної лабораторії, за нашими уявленнями, доцільно організувати у такий спосіб. Як інформаційно-лексикографічну її основу пропонуємо використати описані вище комп'ютерну базу даних та технологію розробки тлумачного Словника української мови, зануривши її у спеціально розроблений двомовний (українсько↔X) інтерфейс, реалізований у середовищі віртуальної лексикографічної лабораторії, який дозволяє здійснювати двобічну взаємодію лексикографів у режимі реального часу. Для забезпечення їхньої роботи доцільно створити двомовний лінгвістичний корпус у складі цифрової бібліотеки літературних текстів на двох мовах та автоматизованої двомовної лексичної картотеки згідно з вищеописаною системотехнікою.

Однак об'єднання віртуальних лексикографічних систем в корпоративну мережу пов'язано зі значними матеріальними витратами, а саме — створенням комунікаційної інфраструктури з високою пропускну здатністю (комунікаційні канали класу не нижче

ЕЗ/ТЗ), вартість якої у наш час є досить велика, особливо у випадку локалізації лексикографічних підсистем у різних національних сегментах комунікаційного середовища. Тим часом, вже існуюча комунікаційна інфраструктура глобальної мережі Інтернет цілком спроможна забезпечити і необхідну пропускну здатність, і нормальну роботу лексикографічної мережі у межах її запланованих функцій, що може бути досягнуто шляхом застосування системотехніки так званої віртуальної приватної мережі — virtual private network (VPN).

Зазначена технологія забезпечує внутрішні інформаційні потоки лексикографічної мережі, використовуючи комунікаційну інфраструктуру Інтернет: синхронізацію необхідних даних, маршрутизацію внутрішніх запитів, діалогову взаємодію і т.ін. без суттєвого ризику для інформаційно-лексикографічних ресурсів створеної VPN. Схему організації такої віртуальної лабораторії представлено на малюнку.

Схема X-Y віртуальної лексикографічної лабораторії



Як бачимо, окремі віртуальні лексикографічні системи (ВЛС X — ВЛС Y) підключено до глобальної мережі Інтернет через VPN-маршрутизатори, які забезпечують емуляцію локальної мережі (через створення захищених IP-тунелів) у середовищі Інтернет. Іншими словами, окремі вузли віртуальної лексикографічної мережі взаємодіють між собою так, якби вони були вузлами єдиної локальної мережі, що значно спрощує налаштування та підтримку внутрішньомережевих сервісів. Тим часом, захист даних, що передаються між

вузлами лексикографічної мережі забезпечується засобами VPN — спеціально настроювати внутрішні сервіси для захисту даних необхідності немає.

У ролі VPN-маршрутизаторів, які показано на малюнку, можуть виступати промислові маршрутизатори, NAT-перетворювачі, а також комп'ютери, що виконують функції брандмауерів. При цьому, VPN-маршрутизатори, окрім забезпечення захищених IP-тунелів між вузлами віртуальної мережі, надають вузлам також і доступ до всіх звичайних ресурсів Інтернет.

Незважаючи на те, що на схемі показано VPN об'єднання лише двох мереж, ця технологія забезпечує створення віртуальної мережі з будь-яким числом вузлів. Таким чином, вузли віртуальної лінгвістичної мережі окрім обміну внутрішньомережевою інформацією можуть водночас аналізувати й накопичувати лінгвістичні дані мовних сегментів Інтернету, які вони обслуговують.

В Українському мовно-інформаційному фонді завершується розроблення онлайн-варіанта фундаментальної академічної лексикографічної системи "Словник української мови" та опрацюється методика й порядок взаємодії між центральним і периферійними її осередками. Через деякий час ця лексикографічна система може бути інстальована на комп'ютерах практично всіх мовознавчих кафедр університетів України, які побажають взяти участь у цьому проекті. Таким чином реалізується можливість поділу праці та організації скоординованого лексикографічного процесу. Фактично це започаткує практичну роботу Української лінгвістичної мережі, що дасть можливість функціонування постійно діючого всеукраїнського лінгвістичного діалогу й безперервного вдосконалення та розвитку "Словника української мови"¹².

Метафора всеукраїнського лінгвістичного діалогу конкретизується через надання їй певних науково-організаційних рис. Для всіх учасників діалогу має бути чітко визначене коло наукових обов'язків; разом з тим вони отримають можливість користуватися й певними правами. Обов'язки впливатимуть з перспективного та поточного плану робіт, де буде відображено ролі й завдання кожного учасника. Права полягатимуть передусім у пріоритетах щодо користування ресурсами Української лінгвістичної мережі та в наявності авторського права на частину виконаної роботи. Нагадаємо, що цій роботі постановою Кабінету Міністрів України "Про за-

¹² Широков В.А. Всеукраїнський лінгвістичний діалог у контексті теорії лексикографічних систем. //Мовознавство., 2003, № 6. сс. 3—7.

твердження Державної програми розвитку та функціонування української мови на 2004—2010 роки” надано статус загальнодержавного проекту, а це забезпечить і відповідний науковий статус його учасникам.

Спектр завдань з розвитку Словника уявляється досить широким. На нашу думку, необхідними його аспектами мають бути:

- розширення та мовно-системна гармонізація реєстру;
- уточнення й удосконалення граматичних характеристик реєстрових рядів та системи граматичної семантики;
- удосконалення, уточнення та модифікація лексичної семантики;
- поглиблення фразеологічно-ідіоматичної підсистеми;
- розширення ілюстративної бази;
- розширення та уточнення діалектної підсистеми;
- удосконалення метамови Словника;
- модернізація та вдосконалення структури Словника як лексикографічної системи;
- введення нових функцій та візуалізація мовно-системних відношень, відсутніх у наявному варіанті словника або заданих імпліцитно;
- інтеграція інших словників та лексикографічних функцій до цифрової системи тлумачного словника (наприклад, етимологія, синонімія, орфоепія і т.ін.).

Новий “Словник української мови”, як аб ініцію цифрова лексикографічна система, у перспективі буде споряджено й акустичними функціями, тобто засобами озвучення текстового матеріалу. Ілюстративну його частину буде розширено шляхом додавання медіафрагментів — графіки, анімації, відео тощо. Не виключена можливість інтеграції Словника з геоінформаційною системою, що, зокрема, надасть йому ознак цифрового діалектологічного атласу. Таким чином, Словник перетворюватиметься на інтегроване, масштабоване лексикографічне середовище еволюційного типу.

Інформаційно-лінгвістична система Словника еволюціонуватиме і як дослідницька. До неї буде “вмонтовано” низку інструментально-дослідних функцій, зокрема: лінгвістичний корпус української та інших мов з автоматизованою лексичною картотекою обсягом на рівні 100 млн. слововживань; статистику мовних одиниць, функцій, відношень; вивчення різних аспектів таксономії, семантики та ідеографії; автоматизований аналіз та синтез словникових дефініцій; автоматизовану побудову нових типів лексикографічних

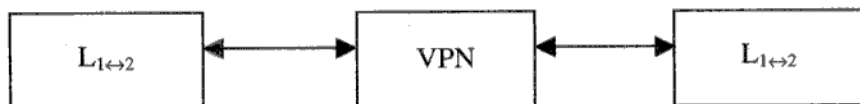
систем (на основі тлумачної); вивчення логічних суперечностей та аномалій в мові й лексикографічних структурах; дослідження з комунікативної, когнітивної лексикографії та культури мови; побудову різних варіантів лексикографічних числень тощо. Реалізація відзначених функцій надасть учасникам проекту потужний інструмент для проведення власних наукових досліджень з української мови на основі дослідницької підсистеми СУМа.

Результати роботи віртуальної лексикографічної лабораторії становитимуть ядро української лінгвістичної системи в мережі Інтернету і, безперечно, будуть надані користувачам. Уже сьогодні тут функціонує в експериментальному режимі Український лінгвістичний портал (адреса сайту HYPERLINK “<http://ulif.org.ua>”, дзеркало якого HYPERLINK “<http://ulif.org.ua>” розташоване на сайті Міністерства освіти та науки України).

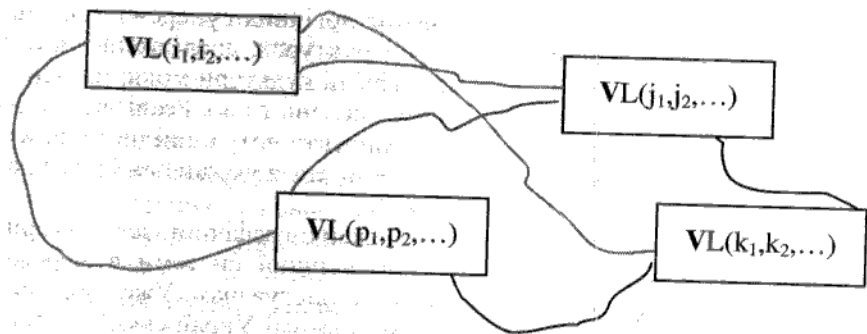
Викладену технологію варто застосувати і до інших лексикографічних проєктів національного рівня; серед таких відзначимо створення віртуальної лексикографічної лабораторії з етимології — цей проєкт розробляється спільно Інститутом мовознавства ім.О.О.Потебні та Українським мовно-інформаційним фондом.

6.4.4. Глобальні системи лінгвістичного діалогу

Якщо описану в попередньому пункті двомовну лексикографічну систему занурити у віртуальне середовище, про яке йшлося вище, тобто піддати процесу віртуалізації, то одержимо віртуальну двомовну лексикографічну лабораторію. Схематично це виглядатиме у такий спосіб:



де символом $L_{1↔2}$ позначено локальну двомовну лабораторію (в смислі системотехніки попереднього розділу), VPN — системотехнічний механізм віртуалізації, за допомогою якого два однорідні локальні об'єкти $L_{1↔2}$ поєднуються в один віртуальний. Якщо таких об'єктів більше двох, позначатимемо таку віртуальну лексикографічну лабораторію символом $VL(i_1, i_2, \dots)$. Як узагальнення описаного процесу отримуємо віртуальне лексикографічне середовище, схематично зображене на рисунку:



де сполучення діють між окремими ВЛЛ, які мають хоч одну спільну мову.

На думку автора, формування такого лексикографічного середовища доцільно було б розпочати зі слов'янських мов, що дозволило би не тільки домогтися прогресу в загальнославистичній проблематиці, але й, завдяки близькості мов, відпрацювати нові системотехнічні рішення, необхідні для ефективного функціонування багатомовних віртуальних систем професійної взаємодії в лінгвістиці.

ЗАКЛЮЧНІ ПОЛОЖЕННЯ

Потреби практики зумовлюють напрям і темп еволюції комп'ютерних інформаційних систем. Бурхливий, некерований і часто непередбачуваний прогрес веб-простору надав зазначеному розвитку певних рис, які вже дозволяють кваліфікувати четверте покоління інформаційних систем. До першого покоління ми відносимо інформаційні системи, які створювалися ще до появи систем управління (або керування) базами даних (СУБД або СКБД). Друге покоління характеризується активним застосуванням класичних моделей даних та створенням і застосуванням різного роду СУБД. Апофеозом цього напрямку стало формулювання Коддом реляційної моделі даних, розробкою численних реляційних СУБД та створенням розвиненої і стандартизованої мови запитів типу SQL та її різновидів. Третє покоління — постреляційні інформаційні системи характеризуються поєднанням реляційних моделей з об'єктно-орієнтованим підходом до моделювання даних, програмування та застосуванням агентних технологій. Нарешті, четверте покоління інформаційних систем — мовно-інформаційні системи ми пов'язуємо із застосуванням механізмів природної мови. Формальною основою цього підходу, як було продемонстровано у нашій монографії, слугують лексикографічні системи та їх узагальнення (лексикографічні середовища, лексикографічні числення та лінгвістичні системи). Мовно-інформаційні системи є інтелектуально орієнтованими *ab initio*. Адже, як відзначалося в книзі¹ “штучний інтелект є формою індивідуалізації технічних систем, якій притаманний мовний статус”. Вони, очевидно, є орієнтованими на мережеві застосування, оскільки мова, насамперед, є інструментом комунікації. Зрозуміло, що мовно-інформаційні системи роблять поки що тільки перші кроки, адже для завоювання ними веб-простору необхідно розв'язати “Вавілонську проблему”, тобто подолати багатомовність, що тягне за собою вирішення вельми складних завдань як чисто мовознавчих, так і системотехнічних. Адже розвиток суспільства знань усе наполегливіше вимагає систем “управління контентом”. Саме зазначена проблематика, на нашу думку, визначатиме поступ інформаційно-комунікаційних технологій. З неї і впливатимуть ключові завдання прикладної лінгвістики та лінгвістичної технології у найближче десятиліття.

¹ В.А.Широков. Феноменологія інформаційних систем. — К.: Наукова думка. 2004, 326 с.

1. Автоматизация в лингвистике: Сб. статей / Пер. с англ., франц., нем. и чешско-го яз. Ред. и сост. Л.И. Засорина. — М.—Л.: Наука, 1966. — 158 с. — С. 2.
2. Агрикола Э. Микро-, медиа- и макроструктура как содержательная основа словаря // Вопросы языкознания. — 1984. — №2. — С. 72—87.
3. Андреш Й.Ф., Даниленко Л.І., Паламарчук О.Л., Русанівський В.М., Федотюк В.С. Українсько-чеські паралелі в текстах і ситуаціях = Ukrajinško-Ceske textove a situacni paralely: Розмовник / Інститут мовознавства ім. О.О.Потебні НАН України; Український мовно-інформаційний фонд НАН України; Слов'янський ін-т Чеської академії наук. — К.: Довіра, 2002. — 477 с.
4. Апресян Ю.Д., Е. Палл. "Русский глагол — венгерский глагол. Управление и сочетаемость". — Будапешт, 1982.
5. Апресян Ю.Д. Интегральное описание языка и толковый словарь // Вопросы языкознания. — 1986. — № 2. — С. 57—70.
6. Апресян Ю.Д. Избранные труды. Т.1: Лексическая семантика: синонимические средства языка. — 2 изд., испр. и доп. — М.: Школа "Языки русской культуры", 1995. — 472 с.
7. Апресян Ю.Д. Избранные труды. Т.2: Интегральное описание языка и системная лексикография. — М.: Школа "Языки русской культуры", 1995. — 767 с.
8. Бехтерев В. М. Объективная психология. — М., 1991. — 480 с.
9. Бибиц С.П., Ермоленко С.Я., Пустовит Л.О. Словник епітетів української мови / Інститут української мови НАН України. — К.: Довіра, 1998. — 431с.
10. Борисов А.И. Применение лингвистической переменной в системах принятия решений. — М., 1987. — 340 с.
11. Бурячок А.А. Орфографічний словник української мови: Близько 36 000 слів. — 4-ге вид., доопрац. — К.: Наукова думка, 2002. — 460 с.
12. Велика чи мала літера? Словник-довідник. Інститут мовознавства ім. О.О. Потебні/ В.В. Жайворонок. — К.: Наукова думка, 2004. — 203 с.
13. Волькенштейн М.В. Теория информации и эволюция // "Кибернетика живого: Биология и информатика". — М.: Наука, 1984. — С. 45—53.
14. Волькенштейн М.В. Энтропия и информация. — М.: Наука, 1986. — 190 с.
15. Гейзенберг В. Физика и философия // Часть и целое / Пер. с нем. — М., 1989. — С. 191—196.
16. Головацук С.І. Російсько-український словник сталих словосполучень. — К.: Наукова думка, 2001. — 639 с.
17. Головацук С.І. Словник наголосів: Понад 20 000 слів. — К.: Наукова думка, 2003. — 319 с.
18. Головацук С.І. Словник-довідник з українського літературного слововживання. — К.: Наукова думка, 2004. — 448 с.
19. Городецкий Б.Ю. Проблемы и методы современной лексикографии // Новое в зарубежной лингвистике. — М.: Прогресс, 1983. — Вып. XIV. — С. 5—23.

20. Горинич В.О. Словник географічних назв України (топоніми та відтопонімі прикметники): Близько 25000 слів / НАН України; Інститут української мови. — К.: Довіра, 2001. — 526 с.
21. Гринев С.В. Введение в терминологическую лексикографию. — М.: МГУ, 1986. — 106 с.
22. Грязнухіна Т.О. “Мовознавство”. — 2001. — № 5.
23. Девкин В. Д. О неродившихся немецких и русских словарях // Вопросы языкознания. — 2001. — № 1. — С. 85—97.
24. Денисов П.Н. Лексика русского языка и принципы ее описания. — М.: Рус. яз., 1980. — С. 217.
25. Денисов П.Н. Системность и связанность в лексике и система словарей // Проблематика определенных терминов в словарях разных типов / Ред. кол. С.Г.Бархударов и др. — Л.: Наука, ЛО, 1976. — С. 63—73. — С. 68—69.
26. Дзюбишина-Мельник Н.Я., Дужик Н.С., Ермоленко С.Я., Ленець К.В., Пустовіт Л.О. Культура мови на щодень / НАН України; Інститут української мови / С.Я. Ермоленко (ред.). — К.: Довіра, 2000, 2002. — 170 с.
27. Дубичинский В.В. Теоретическая и практическая лексикография. — Вена, Харьков: Wiener Slavistischer Almanach Sonderband 45, 1998. — С.57.
28. Дьячков М. В. Миноритарные языки в полиэтнических (многонациональных) государствах. — М., 1996. — 116 с.
29. Економіка знань: виклики глобалізації та Україна / Під заг.ред. А.С.Гальчинського, С.В.Львовичкіна, В.П.Семиноженка. — К.: Національний інститут стратегічних досліджень, 2004. — 261 с.; Україна на шляху до суспільства знань. — К., Прайвесі Юкрейн, 2005. — 69 с.
30. Етимологічний словник української мови: У 7 т. / НАН України; Інститут мовознавства ім. О.О.Потебні / О.С. Мельничук (гол. ред.), Р.В. Болдирев (уклад.) — К.: Наукова думка, 2003. — Т. 4: Н — П. — 653 с.
31. Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений. — М., 1976. — 879 с.
32. Зализняк А.А. Грамматический словарь русского языка: Словоизменение. 3-е изд. стереотип. — М.: Рус. яз., 1987. — 879 с.
33. Золотова Г. А. Синтаксический словарь: Репертуар элементарных единиц русского синтаксиса. — М., 1988. — 440 с.
34. Казакевич О.А. Автоматизация лексикографических работ: Автоматические словари. // НТИ. Сер. 2. 1985. — № 9. — С. 25—29.
35. Калашник В.С., Колоїз Ж.В. Словник фразеологічних антонімів української мови / НАН України; Український мовно-інформаційний фонд; Харківський національний ун-т ім. В.Н.Каразіна. — К.: Довіра, 2001. — 284 с.
36. Караулов Ю.Н. Лингвистическое конструирование и тезаурус литературного языка. — М.: Наука, 1981. — 367 с.
37. Караулов Ю. Н., Молчанов В. И. и др. Анализ метаязыка словаря с помощью ЭВМ. — М., 1982. — С. 6.
38. Клименко Н.Ф., Пецак М.М., Савченко І.Ф. Формалізовані основи семантичної класифікації лексики. — К.: Наук. думка, 1982. — 252 с.
39. Коваленко А. Вероятностный морфологический анализатор (стеммер): <http://linguist.nm.ru/ling/>.
40. Ковалик І.І. Зауваження до “Словника української мови” // Інформаційні матеріали наукової ради з проблеми “Закономірності розвитку національних мов у зв’язку з розвитком соціалістичних націй”. — К.: Наукова думка, 1974. — Вип. 16. — С. 50—53 та ін.

41. *Кодоухов В. И.* Контекст как лингвистическое понятие // Языковые единицы и контекст. — Л., 1973. — С. 7—32
42. *Колесников Н. Ф.* Семонимические словари. — Ростов, 1981. — С. 5.
43. *Колмогоров А. Н.* Три подхода к определению понятия “количество информации” // “Теория информации и теория алгоритмов”. — М.: Наука, 1987. — С. 213—223.
44. *Колшанский Г. В.* Контекстная семантика. — М., 1980. — 149 с.
45. *Коннолли Томас, Бегг Каролин, Страчан Анна.* Базы данных: проектирование, реализация и сопровождение. Теория и практика. — 2-е изд.: Пер. с англ. — М.: Издательский дом “Вильямс”, 2000. — 1120 с.: ил.
46. *Кочерган М. П.* Слово і контекст: Лексична сполучуваність і значення слова. — Львів, 1980. — 183 с.
47. *Кустова Г. И., Паучева Е. В., Рахилина Е. В., Родина Р. И., Филипенко М. В., Якубова Н. М., Янко Т. Е.* Словарь как лексическая база данных: об экспертной системе “Лексикограф” // НГЛ. Сер. 2. — 1993. — №11. — С. 18—20. Лингвистические исследования. Вопросы лексикологии, лексикографии и прикладной лингвистики. / Отв. ред. Р. П. Рогожников. — М., 1976. — 232 с.
48. *Ленець К. В., Ставицька Л. О.* Короткий російсько-український словник контрастивної лексики: Повал 7000 слів / НАН України; Інститут української мови. — К.: Довіра, 2002. — 286 с.
49. *Лучик А. А.* Російсько-український та українсько-російський словник еквівалентів слова / НАН України; Український мовно-інформаційний фонд. — К.: Довіра, 2003. — 495 с.
50. *Лучик А. А.* Семантика прислівникових еквівалентів слова української і російської мов. — К., 2001. — 218 с.
51. *Маказонов П., Александров М., Гельбух А.* Формулы проверки подобия слов с обучением на примерах: построение и применение // Труды Международной конференции “Корпусная лингвистика — 2004”. Санкт-Петербург. — С. 239—255.
52. *Мельничук О. С.* Словник української мови в 11 томах // Вісник АН УРСР. — 1984. — № 3. — С. 100—103.
53. Орфографічний словник української мови: Близько 120 000 слів. Український мовно-інформаційний фонд. Інститут української мови / С. І. Головашук (уклад.). — К.: Довіра, 1994.
54. Орфографічний словник української мови: Близько 125 000 слів / НАН України; Український мовно-інформаційний фонд. Інститут української мови / С. І. Головашук (уклад.). — 2 вид., випр. і доп. — К.: Довіра, 1999. — 989 с.
55. Орфоепічний словник української мови: В 2 т / НАН України; Український мовно-інформаційний фонд. Інститут української мови; Інститут мовознавства ім. О. О. Потебні / М. М. Пещак (ред.). — К.: Довіра, 2001. Т. 1. — К., 2003. — Т. 2.
56. *Падучева Е. В.* Выказывание и его соотносительность с действительностью. — М., 1985. — 271 с.
57. *Паламарчук Л. С.* Новий академічний словник // Мовознавство. — 1980. — № 5. — С. 3—9.
58. *Паламарчук Л. С.* Про принципи добору лексичного інвентаря до загальномовного словника // Мовознавство. — 1973. — № 3. — С. 3—11.
59. *Паламарчук Л. С.* Тлумачний словник української мови в колі слов'янських словників цього типу. Доповідь на VII Міжнар. з'їзді славистів / Варшава, серпень 1973 р. — К.: Наукова думка, 1973. — 20 с.
60. *Паламарчук Л. С.* Українська радянська лексикографія: питання історії, теорії і практики. — К.: Наукова думка, 1978. — 203 с.; Слово і фразеологізм у словнику: 36. наук. праць / Відп. ред. Л. С. Паламарчук. — К.: Наукова думка, 1980. — 250 с.

61. Паламарчук Л.С. Українська радянська лексикографія: питання історії, теорії і практики. — К.: Наукова думка, 1978. — 203 с. — С.131.
62. Пецак М.М., Клименко Н.Ф., Ярун Г.М., Карпіловська Є.А. Лексична семантика в системі “людина — машина”. — К.: Наук. думка, 1986. — 282 с. — С. 276.
63. Пецак М.М., Клименко Н.Ф., Карпиловская Е.А., Шкурфов В.А., Цимбалюк И.В. Украинский семантический словарь: Проспект. — К.: Наукова думка, 1990. — 264 с.
64. Погрібна О.О. Дис. на здобуття наукового ступеня канд. філол. наук. На правах рукопису. — К., 2005.
65. Полога Л.М. Словник антонімів української мови / НАН України; Інститут українознавства ім. І.Крип'якевича Український мовно-інформаційний фонд / Л.С. Паламарчук (ред.). — 2. вид., доп. і випр. — К.: Довіра, 1999. — 275 с. 2 вид., доп. і випр. — К.: Довіра, 2001. — 276 с.
66. Полога Л.М. Словник синонімів української мови / НАН України; Інститут українознавства ім. І.Крип'якевича; Український мовно-інформаційний фонд. — К.: Довіра, 2001, 2004. — 477 с.
67. Полога Л.М., Калашник В.С., Колоїз Ж.В. Словник антонімів і фразеологічних антонімів української мови, / Інститут українознавства ім. І.Крип'якевича НАН України; Український мовно-інформаційний фонд НАН України; Харківський національний ун-т ім. В.Н.Каразіна / Л.С. Паламарчук (ред.). — 2 вид., доп. і випр. — К.: Довіра, 2004. — 275, 284 с.
68. Потєбня А. А. Из записок по русской грамматике.— М., 1958.— Т. 1 — 2.— С. 42, 47—48
69. Рабулець О. Г., Сухарина Н. М., Широков В. А., Якименко К. М. Дієслово в лексикографічній системі, — К.: Довіра, 2004.
70. Рабулець О.Г. Інтегровані лексикографічні системи. Дисертація на здобуття ученого ступеня кандидата технічних наук. На правах рукопису. — Київ, 2002.
71. Рабулець О. Г. Інтегровані лексикографічні системи : Автореф. дис. ... канд. техн. наук. — К., 2002. — 18 с.
72. Ревзина О. Г. Контекст // Русский язык: Энциклопедия / Гл. ред. Ю. Н. Караулов.— М., 1998.— С. 198.
73. Рогоженникова Р. П. Словарь эквивалентов слова: наречные, служебные, модальные единства.— М., 1991. — 254 с.
74. Роменская В.Ф. О классификационной схеме понятия “тип словаря” в информационном тезаурусе // Структ. и приклад. лингвистика. — Л., 1978. — Вып.1. — С. 181—187.
75. Російсько-український словник / НАН України; Інститут мовознавства ім. О.О.Потєбні; Інститут української мови / — К.: Наукова думка, 2003. — 1216 с.
76. Російсько-український словник: Близько 45 000 слів / НАН України; Інститут мовознавства ім. О.О.Потєбні; Український мовно-інформаційний фонд / Г.П. Півторак (ред.), О.І. Цимбалюк, Т.В. Скопненко (уклад.). — К.: Довіра, 2002. — 631 с.
77. Російсько-український словник-довідник / Упор. О.І.Скопненко, Т.В.Цимбалюк; за ред. В.М.Бріцина. — К.: Довіра, 2005. — 942 с.
78. Русанівський В. М. Структура лексичної та граматичної семантики. — К., 1988. — 240 с.
79. Русанівський В.М., Широков В.А. Інформаційно-лінгвістичні основи сучасної тлумачної лексикографії // Мовознавство.— 2002.— № 6.— С. 7—48.
80. Русский семантический словарь. Толковый словарь, систематизированный по классам слов и значений / Под общ. ред. Н. Ю. Шведовой. — М., 1998. — Т. I. — 832 с.

81. Словарь української мови: В 2 т. / Борис Дмитрович Грінченко (упоряд. з дод. власного матеріалу) — К.: Довіра, 1995.
82. Словник іншомовних слів: Близько 10000 слів / Київський національний ун-т ім. Тараса Шевченка; Український мовно-інформаційний фонд НАН України / Є.І. Мазніченко (ред.) С.М. Морозов (уклад.), Л.М.Шкарапута (уклад.). — К.: Наукова думка, 2000. — 662 с.
83. Словник синонімів української мови: У 2 т. / НАН України; Інститут мовознавства ім. О.О.Потебні / А.А. Бурячок (ред.) — К.: Наукова думка, 1999, 2001.
84. Словник українських наукових і народних назв судинних рослин: Близько 40 тис. назв / НАН України; Інститут екології Карпат / Михайло Голубець (відп. ред.), Юрій Кобів (уклад.). — К.: Наукова думка, 2004. — 799 с.
85. Словник української мови / І.К. Білоділ (гол. ред) та ін. — Т. 1—11. — К.: Наукова думка, 1970—1980.
86. Словник фразеологізмів української мови / НАН України; Інститут української мови; Український мовно-інформаційний фонд / В.О. Винник (відп. ред.), В.М. Білоноженко (уклад.). — К.: Наукова думка, 2003. — 1098 с.
87. *Соломонович А.* Семіотика и лингвистика. — М.: Молодая гвардия, 1995. — 352 с.; ил.
88. *Стратонович Р.Л.* Теория информации. — М.: Сов. радио, 1975. — 423 с., гл. 12.
89. *Сухарина Н. М.* Граматична та лексична семантика українського дієслова в лексикографічній системі. Автореферат к.ф.н., — К., 2003.
90. Сучасна українська літературна мова. Вступ. Фонетика (відп. ред. М.А.Жовтобрюх) і “Сучасна українська літературна мова. Морфологія” (відп. ред. В.М.Русанівський), К.: Наукова думка, 1969.
91. Сучасний словник-мінімум іншомовних слів. Вид. 3-є, випр. і доп. — К.: Довіра, 2005.
92. Сучасний словник-мінімум іншомовних слів: Близько 6 тисяч слів. — К.: Довіра, 1999. — 368 с. (Словники України).
93. Сучасний словник-мінімум іншомовних слів: Близько 6000 слів. — 2. вид., випр. і доп. — К.: Довіра, 2002. — 447 с.
94. Теория и практика современной лексикографии: Сб. научных трудов. / Отв. ред. Р.П.Рогожника. — М.: Наука, Ин-т русс. яз., 1984. — 183 с.
95. Глумачний словник-мінімум української мови: Близько 6 тис. сл. / О.М. Єфімов (відп. за вип.). — К.: Довіра, 1999. — 447с.
96. Глумачний словник-мінімум української мови: Близько 7,5 тис. слів / Л.О. Ващенко (уклад.), О.М. Єфімов (уклад.). — 2 вид., виправ. і доп. — К.: Довіра, 2000. — 536 с.
97. Глумачний словник-мінімум української мови: Близько 7500 слів / Л.О. Ващенко (уклад.), О.М. Єфімов (уклад.). — 3. вид., виправ. і доп. — К.: Довіра, 2001. — 535 с.
98. Український орфографічний словник: Орфографічний словник української мови. Близько 143000 слів / Український мовно-інформаційний фонд; Інститут мовознавства ім. О.О.Потебні / М.М.Пешак (уклад.). — 3-є вид., перероб. і доп. — К.: Довіра, 2002. — 1006 с.
99. Український орфографічний словник (Орфографічний словник української мови): Близько 155000 слів / Український мовно-інформаційний фонд; Інститут мовознавства ім. О.О.Потебні / В.В. Чумак (уклад.). — 4-є вид., перероб. і доп. — К.: Довіра, 2005. — 1069 с.

100. Український орфографічний словник (Орфографічний словник української мови). Близько 165000 слів / Український мовно-інформаційний фонд; Інститут мовознавства ім. О.О.Потебні / В.В.Чумак (уклад.). — 5-е вид., перероб. і доп. — К.: Довіра, 2005. (у друці)
101. Українсько-македонський розмовник. Македонсько-український розмовник / О. Онищенко, Б. Ристовські, В. Русанівський та ін. — К.: Довіра, 2005. — 287 с.
102. Українсько-російський словник / НАН України; Інститут мовознавства ім. О.О.Потебні; Інститут української мови / Л.Г. Скрипник (відп. ред.) Г.П.Бжакевич (упоряд.), Т.К. Черторизька (відп. ред.). — 2 вид., доопр. і доп. — К.: Наукова думка, 2004. — 1008 с.
103. Українсько-російський словник: Близько 40 тис. слів. — К.: Довіра, 2003. — 704 с.
104. Українсько-російський словник: Близько 40 тис. слів. — К.: Довіра, 2000. — 704 с.
105. *Ульман Дж.* Основы систем баз данных. — М.: Финансы и статистика, 1983. — 334 с.
106. *Успенский В. А.* К определению падежа по А. Н. Колмогорову / Бюллетень Объединения по проблемам машинного перевода. — № 5. — М., 1957. — С. 11—18.
107. *Успенский В.А.* Теория алгоритмов: основные открытия и приложения. — М.: Наука, 1987. — 298 с.
108. Учебный словарь сочетаемости слов русского языка / Под ред. П. Н. Денисова и В. В. Морковкина. — М.: “Русский язык”, 1978.
109. Філіпчук Г.Г. Словник термінів з питань екології та безпеки життя і діяльності: Навч. посібник / Київський міжнародний ун-т; Український мовно-інформаційний фонд; Інститут педагогіки АПН України; Всеукраїнська екологічна ліга; Інститут безпеки суспільства і людини / Б.Г.Чижевський (ред.), І.Б.Жиляєв (ред.). — Чернівці: Зелена Буковина, 2003. — 752 с.
110. Фразеологічний словник української мови у 2-х кн. — К.: Наукова думка, 1993; та 2-ге вид. — 1999.
111. *Чейф У.* Значение и структура языка. — М., 1975. — 432 с.
112. *Шведова Н.Ю.* Об активных потенциях, заключенных в слове // Слово в грамматике и словаре. — 1984. — С. 11.
113. *Шевченко В.М.* Словник-довідник з релігієзнавства. — К.: Наукова думка, 2004. — 560 с.
114. *Шевченко І.В.* Алгоритмічна словозмінна класифікація української лексики // Мовознавство. — 1996. — № 4—5. — С. 40—44.
115. *Шевченко І.В.* Автоматизована дистрибуція наголосів у словозмінній парадигмі української лексики. — Мовознавство. — 2001. — № 5. С. 26—30.
116. *Шевченко І.В.* Моделі та алгоритмічно-програмне забезпечення лексикографічних систем. Дис. на здобуття наук. ступеня канд. техн. наук. — К., 2000.
117. *Шенк Р., Бирнбаум Л., Мей Дж.* К интеграции семантики и прагматики // Новое в зарубежной лингвистике: Вып 24. Компьютерная лингвистика. — М., 1988. — С. 33.
118. *Шилейко А.В., Кочнев В.Ф., Химушин Ф.Ф.* Введение в информационную теорию систем. — М.: Радио и связь, 1985. — 278 с.
119. *Широков В.А., Воронько М.П., Костышин А.М.* Структура и информационная модель толкового Словаря украинского языка. Труды научной конференции // “Проблеми створення машинних фондів мов”. — Київ, 1991. — С. 74—76.
120. *Широков В.А., Пещак М.М.* Структурна модель реєстрової частини Словника української мови. Зб. наук. праць // “Національна архівна інформаційна сис-

- тема "Архівна та рукописна Україніка і комп'ютеризація архівної справи", вип. 1: "Інформатизація архівної справи в Україні: сучасний стан та перспективи". — К., 1996. — С. 154—174.
121. Широков В.А. Інформаційна теорія лексикографічних систем. — К.: Довіра, 1998. — 331 с.
 122. Широков В.А. Строение лексикографических систем // Математические машины и системы. — 1999. — № 2. — С. 83—104.
 123. Широков В.А. Інформаційна теорія та системотехнічні засади комп'ютерної лексикографії: Автореф. дис. ... д-ра тех. наук. — К., 1999. — 32 с.
 124. Широков В.А. Гуманітарна традиція і технологічний статус мови. — Мовознавство. — 2001. — № 3.
 125. Широков В.А., Манак В.В. Організація ресурсів національної словникової бази // Мовознавство. — 2001. — № 5. — С. 3—13.
 126. Широков В.А., О.Г. Рабулець. Формалізація в галузі лінгвістики. Зб. пр. Актуальні проблеми української лінгвістики. — 2002. — Вип. V. — С. 3—28.
 127. Широков В.А., Рабулець О.Г., Костишин О.М., Шевченко І.В., Якименко К.М. Технологічні основи сучасної тлумачної лексикографії // Мовознавство. — 2002. — № 6. — С. 49—86.
 128. Широков В.А. Феноменологія лексикографічних систем. — К.: Наукова думка, 2004. — 326 с.
 129. Широков В.А. Семантические состояния языковых единиц. ББК 81. Труды международной конференции "MegaLing'2005. Прикладная лингвистика в поиске новых путей" / Отв. ред. В.П. Захаров, С.С. Дикарева. — СПб.: Издательство "Олеипов", 2005. — 180 с.
 130. Широков В.А. Семантичні стани мовних одиниць та їх застосування в когнітивній лексикографії. — "Мовознавство". — 2005. — №№ 3—4.
 131. Широков В.А., Рабулець О.Г., Шевченко І.В., Костишин О.М., Якименко К.М. УС "Словники України": Парадигма. Транскрипція. Фразеологія. Синонімія. Антонімія. Електронний ресурс. Версія 1.03 / НАН України; Український мовно-інформаційний фонд. ISBN 966-507-149-1 (в-во "Довіра"). — К., 2003.
 132. Широков В.А., Рабулець О.Г., Шевченко І.В., Костишин О.М., Якименко К.М. УС "Словники України": Парадигма. Транскрипція. Фразеологія. Синонімія. Антонімія. Електронний ресурс. Версія 1.04 / НАН України; Інститут мовно-інформаційних досліджень. — К.: Довіра, 2004.
 133. Широков В.А., Рабулець О.Г., Шевченко І.В., Костишин О.М., Якименко К.М. УС "Словники України": Парадигма. Транскрипція. Фразеологія. Синонімія. Антонімія. Електронний ресурс. Версія 1.05 / . ISBN 966-507-182-3 (в-во "Довіра"). НАН України; Інститут мовно-інформаційних досліджень. — К., 2005.
 134. Широков В.А., Шевченко І.В., Рабулець О.Г., Костишин О.М., Пецак М.М. УС "Словники України": Парадигма. Транскрипція. Фразеологія. Синонімія. Антонімія. Електронний ресурс / НАН України; Український мовно-інформаційний фонд. — К., 2001. — 1 (CDR) — (Словники України). — ISBN 966-507-122-X.
 135. Широков В.А. Всеукраїнський лінгвістичний діалог у контексті теорії лексикографічних систем // Мовознавство. — 2003. — № 6. — С. 3—7.
 136. Шерба Л. В. Языковая система и речевая деятельность. — Л., 1974. — 428 с.
 137. ANSI/X3/SPARK DBMS study group interim report. //FDT-Bull. ACM SIGMOD. — 1975. — V. 7. — №2. — 140 p.
 138. CODASYL DBTG 1971 [CODASYL Data Base Task Group April 71 Report. ACM-New York, 1971]

139. DNS и BIND. Пол Альбиц и Крикер Ли, O'Reilly, 2002.
140. European Knowledge Society www.eurofound.eu.int/areas/industrialchange/knowledgesociety/ <http://www.meaningprocessing.com/personalPages/tuomi/articles/TheFutureOfTheEuropeanKnowledgeSociety/> etc.
141. *Frege G.* Begriffsschrait, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens. Halle, 1879.
142. <http://cogsci.princeton.edu/~wn/>.
143. <http://www.w3.org/MarkUp/>
144. Longman Dictionary of Contemporary English. — Longman, 1978
145. *Porter M.* An Algorithm for Suffix Stripping. //Program.1980. № 14. P. 130—137.
146. *Selfridge O.G.* Pandemonium: a paradigm for learning. In: Mechanisation of thought processes. London. HMSO., 1959. — P. 511—531.
147. *Szillard L.* Über die Entropievermindung in einem Thermodynamischen System bei Eingriffen intelligenter Wesen. — Zs. f. Phys., 1929. — 53, 11—12, Heft. — P. 840—856.
148. The BBI Combinatory Dictionary of English: A Guide to Word Combinations by M. Benson, E. Benson, and R. Ilson. — John Benjamins Publishing Company, Amsterdam (Philadelphia), 1986.
149. The Oxford English Dictionary. — Oxford, 1970.— V. 1: A — B.— P. VII.
150. *Tsichritzis D.* and *Klug A.* (eds.) The ANSI/X3/SPARK Framework, AFIPS Press, Nontvale, N. J., 1978.
151. Webster's New World Dictionary and Student Handbook, Nashville, 1972.
152. *Zgusta L.* Manual of Lexicography. Praha: Nakl. Ceskosl. akad. ved., 1971.

ВСТУП.....	3
<i>Розділ 1</i>	
ТЕОРЕТИЧНІ ЗАСАДИ ФУНДАМЕНТАЛЬНОЇ ЛЕКСИКОГРАФІЇ	
1.1. Місце і роль лексикографії в лінгвістичній науці.....	12
1.2. Системна лексикографія та інтегрований опис мови.....	15
1.3. Означення, функції та загальна структура словника.....	18
1.4. Теорія лексикографічних систем.....	28
1.4.1. Лексикографічний ефект в інформаційних системах.....	28
1.4.2. Структура і архітектура лексикографічних систем.....	40
1.4.3. Лексикографічні структури і словники.....	57
1.4.4. Лексикографічні середовища.....	63
1.4.5. Інтегровані Л-системи та методика їх побудови.....	65
<i>Розділ 2</i>	
ЕЛЕКТРОННІ ГРАМАТИЧНІ СЛОВНИКИ	
2.1. Моделювання словозмінної системи.....	70
2.1.1. Морфологічний аналіз.....	70
2.1.2. Формальні засади побудови ЕГС (Морфологічна модель словозмінної флективної мови).....	72
2.2. Структура бази даних ЕГС.....	80
2.2. Зовнішній рівень представлення ЕГС.....	86
<i>Розділ 3</i>	
ЛЕКСИКОГРАФІЧНА СИСТЕМА СЛОВНИКА УКРАЇНСЬКОЇ МОВИ	
3.1. Словник української мови в 11-ти томах (1970—1980 рр.).....	94
3.2. Представлення семантичних структур СУМа засобами теорії лексикографічних систем.....	98
3.3. Концептуальні засади нового академічного тлумачного Словника української мови.....	103
3.3.1. Мікроструктура реєстрових частин словникових статей СУМа.....	104
3.3.1.1. Дієслово. Структура.....	104
3.3.1.2. Дієслово. Класифікація.....	119
3.3.1.3. Іменник. Структура та класифікація.....	131
3.3.2. Мікроструктура інтерпретаційних частин словникових статей СУМа.....	161
3.4. Системотехнічні засади побудови лексикографічної бази даних СУМа.....	182
3.4.1. Структура лексикографічної бази даних СУМа.....	184
3.4.2. Програмний комплекс редагування ЛБД СУМа.....	197

Розділ 4

ІНФОРМАЦІЙНО-ЛЕКСИКОГРАФІЧНЕ МОДЕЛЮВАННЯ ІНТЕГРОВАНІХ СЛОВНИКОВИХ СИСТЕМ

4.1. Загальні положення.	202
4.2. Будова парадигматично-орфоепічної Л-системи.	203
4.3. Будова фразеологічної Л-системи.	207
4.4. Будова синонімічної Л-системи.	208
4.5. Будова антонімічної Л-системи.	211
4.6. Інтегрована лексикографічна система “Словники України”	215

Розділ 5

СЕМАНТИЧНІ СТАНИ МОВНИХ ОДИНИЦЬ ТА ЇХ РОЛЬ У МОДЕЛЮВАННІ МОВИ

5.1. Мовна реальність.	224
5.2. Нарис теорії семантичних станів.	226
5.3. Семантичні стани у формалізмі нечітких множин.	229
5.4. Формальні аспекти відношення синонімії та аналогія з визначенням перекладних еквівалентів.	232
5.5. Семантичні стани та лексикографічні числення.	235
5.5.1. Лексикографічні числення.	235
5.5.2. Узагальнена лінгвістична змінна.	247

Розділ 6

БАГАТОМОВНІ ЛЕКСИКОГРАФІЧНІ СЕРЕДОВИЩА

6.1. Представлення перекладних еквівалентів у багатомовному перекладному словнику (БПС)	254
6.1.1. Принципи побудови БПС у концепції семантичних станів ..	254
6.1.2. Приклади представлення перекладних еквівалентів у зонах БПС.	257
6.2. Інтегрована система синонімічної та тлумачної Л-систем.	258
6.2.1. Принципи побудови WordNet.	258
6.2.2. Л-система Словника синонімів української мови.	265
6.2.3. Інтеграція синонімічної ЛБД з ЛБД СУМа.	272
6.3. Системотехніка двомовної лексикографії.	279
6.4. Системи лінгвістичного діалогу	281
6.4.1. Всеукраїнський лінгвістичний діалог в контексті теорії Л-систем	281
6.4.2. Віртуальні лексикографічні системи	283
6.4.3. Віртуальна лексикографічна лабораторія	287
6.4.4. Глобальні системи лінгвістичного діалогу	291
ЗАКЛЮЧНІ ПОЛОЖЕННЯ.	293
ЛІТЕРАТУРА.	294

Наукове видання

ШИРОКОВ Володимир Анатолійович

ЕЛЕМЕНТИ ЛЕКСИКОГРАФІЇ

Відповідальні редактори *Т. О. Грязнухіна, О. Г. Рабулець*

Коректор *О. В. Романенко*

Комп'ютерна верстка *В. М. Виноградова*

Підписано до друку 22.12.05. Формат 60×84¹/₁₆.

Папір офсетний № 1. Гарнітура Тайме. Друк офсетний.

Умовн. друк. арк. 17,67. Умовн. фарбовідб. 18,14. Обл.-вид. арк. 19,22.

Наклад 1000 прим. Замовлення 05-92.

Видавництво "Довіра"

вул. Кіквідзе, 2/34, Київ-103, 01103

Свідцтво про внесення до Державного реєстру видавців
серія ДК № 669 від 14.11.2001 р.

ТОВ Університетське видавництво "Пульсари"
вул. Межигірська, 7/16, Київ-71, 04071.

Широков В. А.

Ш64 **Елементи лексикографії.** — К.: Довіра, 2005. — 304 с.: іл. —
Бібліогр.: с. 296–303.

ISBN 966-507-187-4

У монографії викладаються результати багаторічної діяльності Українського мовно-інформаційного фонду Національної академії наук України зі створення національної словникової бази. Подано нову концепцію та комп'ютерну технологію лексикографічної бази.

ББК 81.2УКР-4с