

ПРО ФУНКЦІОНАЛЬНІ МОЖЛИВОСТІ ДОСЛІДНИЦЬКОГО ІНСТРУМЕНТАРІЮ ВІРТУАЛЬНОЇ ЛЕКСИКОГРАФІЧНОЇ СИСТЕМИ «СУМ-20»

У статті окреслено загальні тенденції розвитку прикладної лексикографії в Україні, зокрема визначено деякі аспекти автоматизованої лексикографії. Автором проаналізовано основні дослідницькі інструменти та сервісні можливості лексикографічної системи «Словник української мови» в 20-ти томах, а також схарактеризовано специфіку внутрішнього пошуку у відповідній лексикографічній праці.

Ключові слова: прикладна лексикографія, лексикографічна система, внутрішній пошук.

Сучасний стан гуманітарної науки характеризується не тільки розробленням нових напрямів дослідження мови, виявленням її міждисциплінарних зв'язків, істотним удосконаленням і переосмисленням граматичної теорії, гіпотез походження мови тощо, але передусім — посиленою увагою наукової спільноти до інноваційних технологій, поява та стрімке поширення яких має безпосередній вплив і на мовознавство.

Теоретичні та прикладні здобутки прикладної та корпусної лінгвістики суттєво змінюють способи й методи дослідження мови, дають змогу зовсім інакше оцінювати багатоаспектні мовні процеси та явища, по-новому вибудовуючи й інтерпретуючи наукову картину світу. Матеріал для подібних досліджень повинен охоплювати всі рівні мовної системи з усіма можливими девіаціями на кожному з них, тому значення тлумачного загальномовного словника для подібних розвідок важко переоцінити.

Завдяки корпусним технологіям та укладеним базам даних сучасний лінгвіст має змогу працювати з великим масивом мовних даних, вирішуючи низку актуальних питань, адже цифровий формат дозволяє багато разів повертатися до різних об'єктів лінгвістичного опису, працювати дистанційно, отримувати статистичну чи будь-яку іншу інформацію, опрацювання якої ще кілька десятків років тому потребувало в сотні разів більше часу.

У сучасній українській лінгвістиці результатом таких досліджень є загальномовний «Словник української мови» у 20-ти томах, який укладається науковцями Українського мовно-інформаційного фонду та Інституту мовознавства ім. О. О. Потебні НАН України. Перші сім томів цієї лексикографічної праці вже з'явилися у світ у паперовому й електронному варіантах завдяки використанню розробленої в Українському мовно-інформаційному фонді теорії лексикографічних систем, заснованої на феноменології лексикографічного ефекту, застосування якої стало поштовхом до здійснення необхідних системних уза-

гальнень і створення методології побудови лексикографічних моделей. Ця теорія також передбачає виділення в досліджуваній мовній системі підсистеми її елементарних інформаційних одиниць і визначення множини їхніх системно-структурних параметрів¹.

Розроблене в Українському мовно-інформаційному фонді НАН України спеціальне програмне забезпечення для укладання нового академічного «Словника української мови» в 20-ти томах має ряд переваг: дистанційний доступ через мережу Інтернет, постійне оновлення даних Словника, відслідковування внутрішніх зв'язків між словниковими статтями, забезпечення їх уніфікованості, вільний доступ до електронної версії «Інструкції до «СУМ-20»» і «Українського правопису» та багато іншого. Завдяки унікальному набору інструментів віртуальна лексикографічна система (ВЛС) «СУМ-20» має багатоцільове призначення: від отримання найрізноманітніших статистичних даних, дослідження різновидів статей, способів і засобів сучасного лексикографування до визначення принципів і законів семантизації, відслідковування системних зв'язків у мові в цілому. Віртуальна лексикографічна лабораторія (ВЛЛ), основне призначення якої — бути робочим місцем сучасного лексикографа, також має розвинений інструментарій, правильне використання якого може стати в нагоді не лише в укладанні відповідного тлумачного словника, а й у безпосередньому дослідженні мовних феноменів.

Сучасне прикладне словникарство в Україні поступово набуває рис автоматизованої лексикографії, що передбачає вироблення алгоритмів удосконалення вже існуючих лексикографічних праць, а також створення нових із застосуванням сучасних технологій і науково-практичних напрацювань. Така тенденція цілком закономірна й може бути пояснена стрімким розвитком сучасного суспільства, яке, на думку В. А. Звегінцева, тяжіє до «машинізації», одним з аспектів якої є прагнення науковців автоматизувати практично всі види людської діяльності, зокрема ті, що потребують участі інтелекту².

Останнім часом особливої значущості та стрімкого поступу набувають інтеграційні процеси в лексикографії, коли для створення зручного робочого місця лінгвіста та спрощення його роботи лексикографічні й довідкові праці різних типів і років об'єднуються завдяки спеціальному програмному забезпеченню, яке гарантує одночасний доступ до кількох праць завдяки відповідному інтерфейсу та інструментарію. Тому однією з нагальних потреб сучасної лексикографії є необхідність автоматизованого трансформування одного лексикографічного продукту в інший. На думку Л. Беляєвої, такий підхід до створення словників передбачає формування й використання корпусу текстів, який може розглядатися як база даних не лише для розв'язання дослідницьких проблем, а й для досягнення практичних лексикографічних завдань³.

Безумовно, для побудови досконалих мовно-інформаційних артефактів необхідно мати експериментальні масиви лінгвістичних даних, які б відображали всі мовні феномени⁴. Академічний «Словник української мови» в 20-ти томах,

¹ Широков В. А., Білоноженко В. М. та ін. Лінгвістичні та технологічні основи тлумачної лексикографії.— К., 2010.— С. 58, 60.

² Звегінцев В. А. Теоретическая и прикладная лингвистика : Учеб. пособие.— 2-е изд.— М., 2007.— С. 272.

³ Беляева Л. Автоматизированное рабочее место лексикографа: состав и структура // Прикладна лінгвістика та лінгвістичні технології: MegaLing-2011 : 36. наук. пр. / Редкол.: В. А. Широков та ін.— К., 2012.— С. 35.

⁴ Широков В. А., Бугаков О. В. та ін. Корпусна лінгвістика.— К., 2005.— С. 5.

представлений спеціальною лексикографічною системою, і є таким масивом лінгвістичних даних.

Тлумачний словник як компонент інтегрального лінгвістичного опису містить два типи лексикографічної інформації — класифікаційну та операційну⁵. Спеціальний набір дослідницьких інструментів ВЛС «СУМ-20», яка входить до Віртуальної лексикографічної лабораторії, дозволяє по-різному відтворювати цю інформацію, розвиваючи у такий спосіб своєрідне лексикографічне моделювання. При цьому «Словник української мови» у 20-ти томах може бути вихідним продуктом для інших лексикографічних праць (надалі під Словником розуміємо відповідну лексикографічну систему). Усе це стає можливим тільки завдяки певному набору системно-структурних ознак, що визначають словник як абстрактний мовно-інформаційний об'єкт⁶.

Серед сервісних можливостей ВЛС «СУМ-20», поряд з адмініструванням, комунікуванням лексикографів, моніторингом, редагуванням та ін., є вибірка, яка забезпечує багатопараметричний доступ до зазначених статей та їхніх компонентів. Вибірка призначена для формування й подальшого використання цільових наборів словникових статей шляхом комбінування та встановлення різних умов, яким мають відповідати відібрані структурні елементи статей⁷.

Здатність ВЛС виконувати пошуковий запит за вказаними параметрами дозволяє укладачеві здійснювати комплексну перевірку опрацьованих статей, їх уніфікування й подальше наукове дослідження словникового складу мови. Вибірка як спеціальний інструмент інтерфейсу призначається для внутрішньої навігації та подальшого найрізноманітнішого оброблення (аналізу, порівняння, уніфікування, систематизування, синтезування і т. ін.) отриманих результатів пошуку за вказаними параметрами.

У пункті меню «Вибірка» діалогового вікна ВЛС «СУМ-20» є особливий набір пошукових параметрів, які можуть додатково використовуватися лексикографом (див. рис. 1). Деякі сервісні можливості цього пункту лексикографічного опрацювання реєстрових одиниць допомагають пришвидшити виконання запланованих завдань.

Пункт меню «Вибірка» також передбачає постійне комунікування в колективі укладачів Словника. Цей аспект у ВЛС «СУМ-20» реалізовано завдяки спеціальному підпункту меню «Коментарі», де завдяки функціям «Додати», «Обробити», «Записати» укладач має можливість зафіксувати побажання, зауваження, пропозиції щодо розроблення тих чи інших статей, а також висловити власні міркування щодо доцільності / недоцільності окремих пропозицій. Такий своєрідний лінгвістичний діалог потрібен на всіх етапах роботи зі Словником, тому вимагає посиленої уваги та контролю. Інтерфейс ВЛС передбачає швидкий відбір словникових статей з коментарями за допомогою підпункту меню «Коментарі», де серед параметрів пошуку є можливість указати автора чи адресата, а також дату та відпрацьованість / невідпрацьованість відповідних зауважень лексикографом.

⁵ Плющ М. Я. Лексикографічне тлумачення дієслівної лексеми // Українська і слов'янська тлумачна і перекладна лексикографія : Леонідові Сидоровичу Паламарчукові. — К., 2013. — С. 196.

⁶ Широков В. А. Комп'ютерна лексикографія. — К., 2011. — С. 15.

⁷ Манако В., Широков В., Якименко К. Пользовательский интерфейс виртуальной лексикографической лаборатории «Словарь украинского языка» // Прикладна лінгвістика та лінгвістичні технології. — С.193, 197.

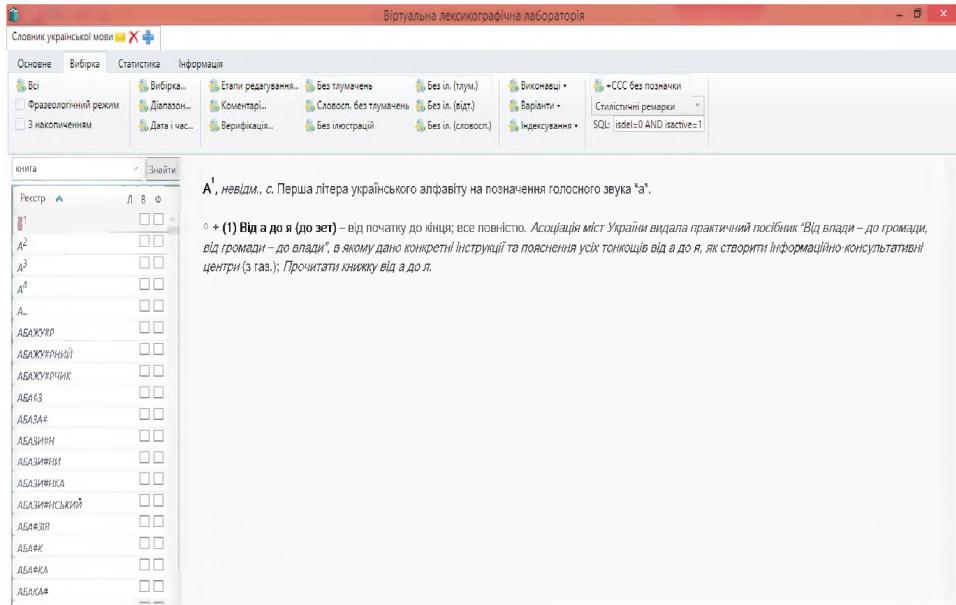


Рис. 1. Основне діалогове вікно ВЛС «СУМ-20»

Важливим елементом взаємодії лексикографів є написання й опрацювання коментарів із зауваженнями та пропозиціями до окремих структурних частин словникових статей. Додатковою функцією такого сортування є отримання статистичних даних щодо створення й опрацювання коментарів лексикографами (див. рис. 2).

Однак ВЛС «СУМ-20» є не тільки багатфункціональним робочим місцем лексикографа, але й може бути матеріалом для різних лінгвістичних досліджень, а головне — незамінним інструментом для мовознавчих та інших розвідок. При цьому вкрай важливим для дослідників мови є не стільки весь реєстр (макроструктура) загальнономовного словника, скільки його внутрішня організація, зокрема побудова та експлікація лінгвістичних відношень між словниковими статтями, їхнє об'єднання в поля та групи на основі семантичних, тематичних, асоціативних та інших ознак. Відповідну структуру словникових статей і способи розкриття семантики реєстрових одиниць, ієрархію їхніх значень реалізовано на мікро- та медіоструктурному рівнях⁸. Розглядаючи вслід за відомими науковцями мову як систему (Ю.Д. Апресян, В.А. Звегинцев, І.А. Мельчук, М. В. Нікітін, М. М. Пешак, В. А. Широков та ін.⁹), можна стверджувати, що всі аргументи, які доводять системний характер мови, приховані в її внутрішній структурі. В. А. Широков наголошує, що тлумачний словник містить у собі велику кількість неявних системних зв'язків мовної системи, які формують його семантичний фрейм¹⁰.

⁸ Широков В. А. Зазнач. праця. — С.17.

⁹ Апресян Ю. Д. Исследования по семантике и лексикографии : В 2 т.— М., 2009.— Т. I. — 568 с.; Звегинцев В. А. Зазнач. праця; Толково-комбинаторный словарь современного русского языка : Опыт семантико-синтакс. описания рус. лексик / Мельчук И. А., Жолковской А. К. и др.— Вена, 1984.— 992 с.; Никитин М. В. Курс лингвистической семантики —2-е изд., дополн. и исправл.— СПб., 2007.— 819 с.; Пешак М. М. Нариси з комп'ютерної лінгвістики: текст.— Ужгород, 1999.— 199 с.; Широков В. А. Зазнач. праця.— 351 с.

¹⁰ Широков В.А. Системна семантика тлумачних словників // Акцентологія. Етимологія Семантика : До 75-річчя академіка НАН України В. Г. Скляренка. — К., 2012. — С. 510.

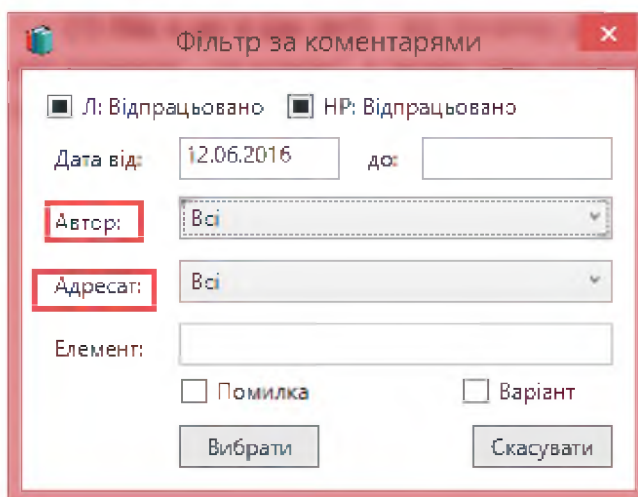


Рис. 2. Фільтр словникових статей за коментарями

Дослідження таких лінгвістично-когнітивних аспектів може бути здійснене інструментами ВЛС «СУМ-20», у якій передбачено особливий підпункт «Вибір-ка» в однойменному пункті меню «Вибір-ка» діалогового вікна лексикографічної системи. Важливість та унікальність цього підпункту зумовлені здатністю системи сортувати весь масив словника за різними заданими користувачем параметрами реєстрової та інтерпретаційної частин. За допомогою цього пункту дослідник має доступ до окремих підсистем мови, які представляють ті чи ті лексикографічні ефекти і, як наслідок, — доводять феноменальність мови як системи в цілому. Зазначимо, що всі ці сервісні можливості реалізовані завдяки дискретному характеру матеріалів, адже саме дискретність є основною системно-структурною ознакою сучасних словників¹¹.

Діалогове вікно підпункту меню «Вибір-ка за елементами статей» має п'ять вкладок: «Реєстр», «Значення», «Ілюстрації», «В-Значення» та «В-Ілюстрації» з відповідними полями, заповнення яких дозволяє вводити умови пошукового запиту з одним чи кількома параметрами (див. рис. 3). У вкладці «Реєстр» можна задати умови для пошуку за реєстровим словом або навіть дериватом, лівою частиною словникової статті, зв'язком реєстрового слова з відсильними статтями. Також одразу з'являється можливість порівняти реєстр «Словника української мови» в 20-ти томах та 11-томного видання, прослідкувавши помітне його (реєстру) оновлення та розширення у двадцятитомному виданні.

Вкладка «Реєстр» має декілька вільних полів (див. рис. 4), заповнення яких передбачає відбір словникових статей, що відповідають заданим параметрам пошуку. Наприклад, у полі «Частина мови» з фіксованого списку користувач може обрати необхідний частиномовний клас. Важливо, що пошукові умови задаються не ізольовано, їх можна комбінувати, доповнюючи одна одну. Для цього у підпункті меню «Вибір-ка» до кожного поля передбачено спеціальні умови: «->» (пошукова умова не враховується); «->та->» (перетин умов пошуку); «-або->» (операція об'єднання)¹². Таким чином, користувач має можливість додатково редагувати умови пошуку: отримавши відібрані словникові статті якоїсь частини мови, він може працювати з усією підсистемою цього класу слів або відсор-

¹¹ Широков В.А. Комп'ютерна лексикографія.— С. 15.

¹² Манако В., Широков В., Якименко К. Зазнач. праця.— С. 198.

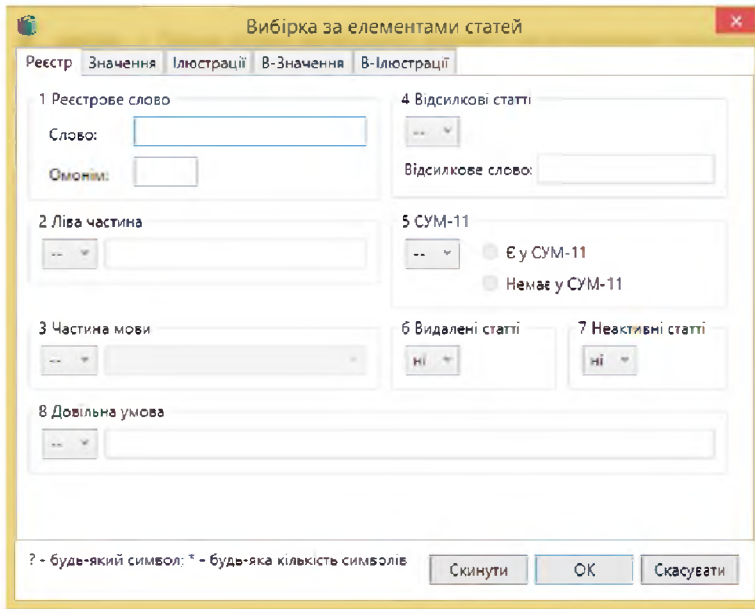


Рис. 3. Діалогове вікно «Вибірка за елементами статей»

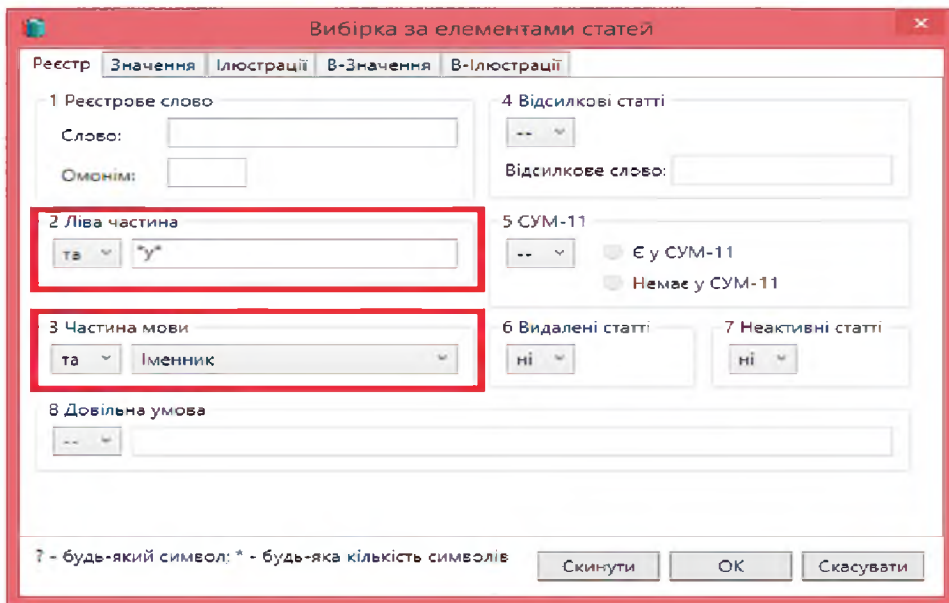


Рис. 4. Вкладка «Реєстр» діалогового вікна «Вибірка за елементами статей»

тувати їх за додатковим параметром, додавши у поле «Ліва частина» інформацію, наприклад, про відповідну словозмінну парадигму.

Важливим елементом діалогового вікна підpunkту «Вибірка» є вкладка «Значення», за допомогою якої можна здійснювати пошукові запити за різними структурними елементами інтерпретаційної частини описової словникової статті, зокрема безпосередньо за дефініціями чи їхніми елементами (див. рис. 5). Сортування в такому випадку можна виконувати за певними семами чи ключовими словами, заповнивши відповідне поле «Значення».

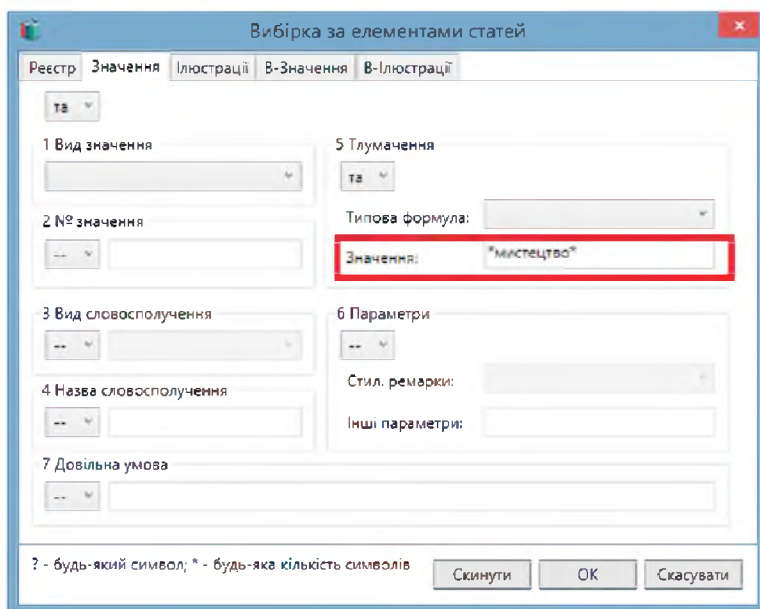


Рис. 5. Вкладка «Значення» діалогового вікна «Вибірка за елементами статей»

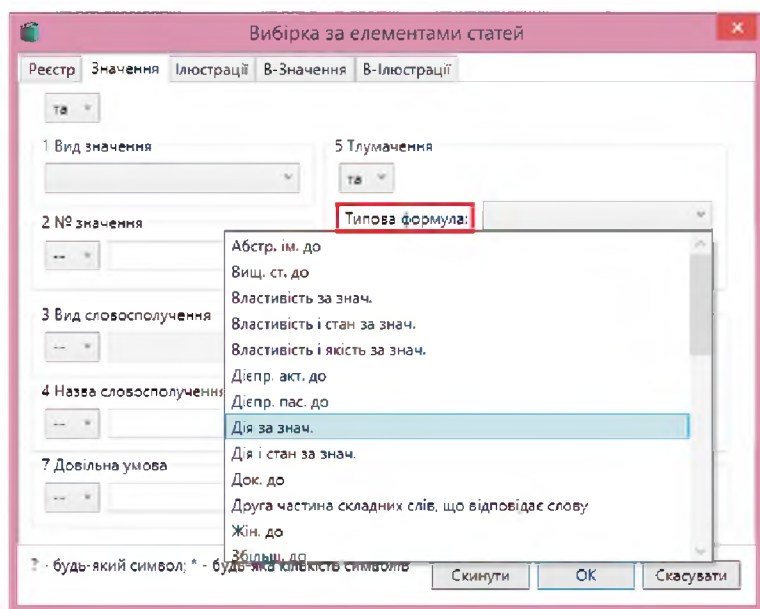


Рис. 6. Перелік типових формул тлумачень, використовуваних у ВЛС «СУМ-20»

«Словник української мови» в 20-ти томах передбачає не лише описові тлумачення, а й відсилні, що зумовлено особливостями метамови цієї лексикографічної праці. Такі дефініції представлені стандартним набором формул тлумачення, своєрідних семантичних кліше, які, на думку В.А. Широкова, маркують тип відношення між означуваним словом та його відповідником у дефініції¹³. Цей особливий спосіб дефініювання дозволяє швидко сортувати весь реєстр за значеннями, адже в підпункті меню «Вибірка» крім різних пошукових парамет-

¹³ Широков В.А. Системна семантика тлумачних словників. — С. 500.

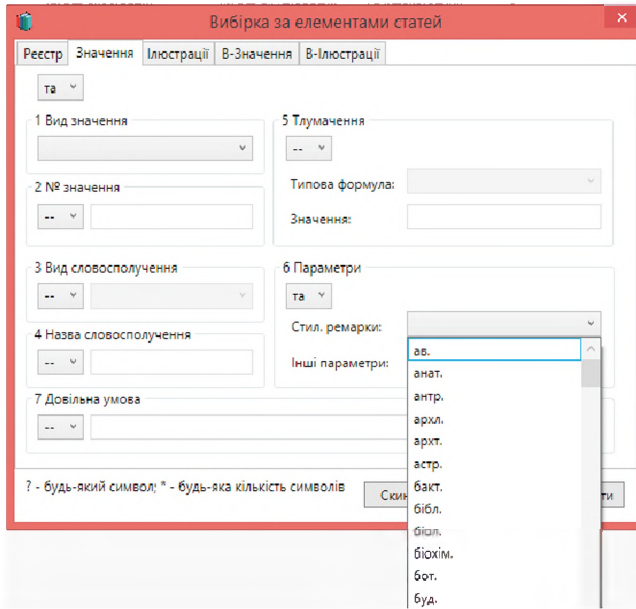


Рис. 7. Перелік стилістичних і галузевих ремарок, використовуваних у ВЛС «СУМ-20»

рив є і повний перелік типових формул тлумачень, або, за термінологією В.А. Широкова, формул квазісемантики¹⁴. Для виконання необхідного пошукового запиту варто лише з фіксованого списку обрати необхідну дефініцію (див. рис. 6).

Крім описаних вище можливостей у підпункті меню «Значення» передбачені й інші параметри, вибір яких забезпечує сортування всього реєстру Словника за стилістичними та галузевими ремарками (див. рис. 7).

Отже, дослідницькі можливості ВЛС «СУМ-20» не обмежуються лише пошуком за реєстром, а передбачають внутрішню навігацію за структурними елементами словникової статті, яку додатково перевіряють та уніфікують. Завдяки цьому ВЛС «СУМ-20» можна вважати своєрідним новітнім інструментарієм сучасних лінгвістичних досліджень, який забезпечує швидкий доступ до різних мовних підсистем.

V. V. SYVOKOZOVA

ON THE FUNCTIONAL POTENTIAL OF THE EXPLORATORY TOOLS OF THE VIRTUAL LEXICOGRAPHICAL LABORATORY «SUM-20»

The article defines the main tendencies of the development of Ukrainian applied linguistics and the main aspects of the automatized lexicography. The author analyzes the lexicographic system «The Dictionary of Ukrainian language in 20 volumes», its main investigative instruments and service opportunities. Also the specificity of the internal research in this lexicographic work is given.

Key words: the applied lexicography, lexicographical system, internal search.

¹⁴ Там же. — С. 501.