

КОМП'ЮТЕРНА ПРОГРАМА «GLOSSA» - ЗАСІБ СИСТЕМАТИЗАЦІЇ ТА АНАЛІЗУ СТИЛЬОВО-СТИЛІСТИЧНИХ РЕМАРОК

У статті висвітлено процедуру створення та принципи роботи комп'ютерної програми «Glossa». Наведені приклади застосування цієї програми. Проведено порівняльний аналіз маркування сліву досліджуваних лексикографічних працях.

Ключові слова: комп'ютерна програма, словник, ремарка, маркування.

Починаючи з другої половини ХХ ст. і до сьогодні, все більше й більше зростає увага до використання інформаційних технологій у лінгвістиці, зокрема у галузі лексикології та лексикографії. В українському мовознавстві досвід залучення комп'ютерних програм для створення словників та опрацювання лінгвістичних даних висвітлено у працях С. А. Карпіловської, Н. Ф. Клименко, Л. С. Ковтуненко, М. М. Пецак, В. М. Широкова, Л. В. Струганець, В. Г. Волошина, Н. П. Дарчук. Ці праці дають загальне уявлення про основи та розвиток прикладної лінгвістики, засвідчують різнобічність застосування лінгвінформаційних підходів.

Метою нашого дослідження є аналіз та систематизація стильово-стилістичних ремарок, зафіксованих у словниках української мови, з використанням методики комп'ютерного програмування. *Стильово-стилістичні ремарки* – важливі позначки у словникових статтях, що вказують на сферу функціонування слова, його експресивне чи емоційне забарвлення [1, с. 93]. Ми виокремлюємо власне стильові («*розмовне слово*»), «*книжне слово*», «*поетичне слово*») та стилістичні ремарки («*вульгарне слово*»), «*іронічне слово*», «*урочисте слово*» тощо).

Одним із основних завдань сучасної комп'ютерної лінгвістики є створення баз даних для пошуку та опрацювання конкретної інформації. Створення автоматизованих систем інформаційного пошуку дозволяє виявляти фактичні відомості за заданими ознаками [2, с. 43]. Однак опис функціонування ремарок у словниках та їх дослідження за допомогою сучасних інформаційних технологій залишається на периферії уваги українських мовознавців. Чи не єдина заторкує цю проблему Л. В. Струганець у своїй праці «Динаміка лексичних норм в українській лексикографії ХХ століття». У цій роботі висвітлено використання комп'ютерних технологій для

аналізу лексико-семантичної системи української літературної мови. Авторка створила персональну автоматизовану інформаційну систему «Lingua», яка дозволяє порівняти реєстри загальномовних, перекладних, орфографічних словників української мови. Застосування цієї комп'ютерної програми дає змогу дослідити лексичні норми сучасної української літературної мови у синхронії та діахронії.

Вихідними матеріалами дослідження інформаційної системи «Lingua» є кілька баз даних. Кожній лексемі присвоюється цифровий код. Такими кодами представлені й усі відомості баз даних вихідних матеріалів дослідження, параметри вибірки також задаються цифровими кодами [6, с. 83].

Безперечно, перевагою створеної інформаційної системи «Lingua» є те, що вихідні дані цієї системи відображають багатоаспектне вивчення лексики на основі лексикографічних джерел. Незважаючи на те, що «Lingua» дозволяє порівняти реєстри словників та зміни стилістичного маркування лексем, ця система не орієнтована безпосередньо на вивчення стильово-стилістичних ремарок. Її основу створюють кілька баз даних, які орієнтовані на комплексне дослідження граматичних, семантичних, структурних, стилістичних характеристик слів. Крім того, застосування цифрових кодів для вибірки ремарок та вихідних даних дослідження стилістичних характеристик слів, на нашу думку, не надто практичне для потенційного користувача програмою (наприклад, у базі даних <Стилістична позначка> для ремарки «розмовне» передбачено код 83).

Перед нами постало завдання створити комп'ютерну програму, яка б на основі однієї бази даних давала цілісне уявлення про використання стильово-стилістичних ремарок у словниках та була б зручною у застосуванні для пересічного користувача. Зважаючи на позитивний досвід використання інформаційної системи «Lingua» для дослідження лексичного складу української мови на основі лексикографічних джерел, автором було розроблено власну методика залучення інформаційних технологій для аналізу використання стильово-стилістичних ремарок та маркованих ними слів у словниках української мови. На основі цієї методики аспірантом кафедри інформаційних технологій НаУКМА В'ячеславом Горборуковим було створено комп'ютерну програму «Glossa». Оскільки доцільність використання ремарок неможливо оцінити без їх конкретного застосування до лексики, дослідження ремарок здійснювалось у зв'язку з відповідною лексикою.

Лінгвістичну базу даних програми було створено на основі словників української мови, а саме: «Російсько-українського словника» у 3-х тт. за редакцією А. Кримського (1924–1933), «Російсько-українського словника» у 3-х тт. (1969), «Словника української мови» в 11-ти тт. (1970–1980) – СУМ та «Словника іншомовних слів» за редакцією С. Я. Сроменко (2006). З огляду на те, що важлива динаміка стилістичного забарвлення слів у діахронному аспекті, ми будемо аналізувати словники різних часових зрізів.

Для зручності інформаційного пошуку та представлення його результатів у вигляді таблиці, кожному словнику було присвоєно кодові назви – РУС24 (Рус24), РУС69 (Рус69), СУМ (SUM), СІС06 (SIS06).

До бази даних програми також внесено список стильово-стилістичних ремарок. Як відомо, візуалізація ремарок у різних словниках неоднакова. Це чітко про-

стежується у словниках різних років. Так, у «Словнику української мови» в 11-ти томах укладачі дотримувались послідовності у скороченні ремарок – «ірон.», «пестл.», «жарт.». Проте у «Російсько-українському словнику» за редакцією А. Кримського застосовувались паралельно такі стилістичні позначки: «ірон., іронич.», «ласк., ласкат. назв.», «насм. насмеш., насмешл.», тощо. Наприклад: **Белотелый** – (ірон.) білотелесий; **аршинник** – 2) (іронич.) шахрай [4]. Як відомо, комп'ютерні програми дуже чутливі до будь-яких відмінностей у структурі знаків. Для таких програм порівнювані скорочені слова – вже є різними одиницями. Крім того, у перекладних словниках (РУС24, РУС69) ремарки подаються російською мовою. Тому для зведеної таблиці, яка відтворює маркування слів у різних словниках, було запрограмовано, аби комп'ютерна програма подавала такі ремарки українською мовою.

Отже, щоб уникнути максимальних похибок у процесі порівняльного аналізу маркування слів, було задано такі формули: *арх.* = *архаич.*; *дит.* = *дет.* = *детс.* = *детск.* тощо.

Інструментальна панель програми «Glossa» складається з кнопок «Шукати» і «Зберегти» та полів «Пошук в тексті», «Слово», «Значення слова», «Словники», «Фільтр ремарок» (Див. мал. 1).

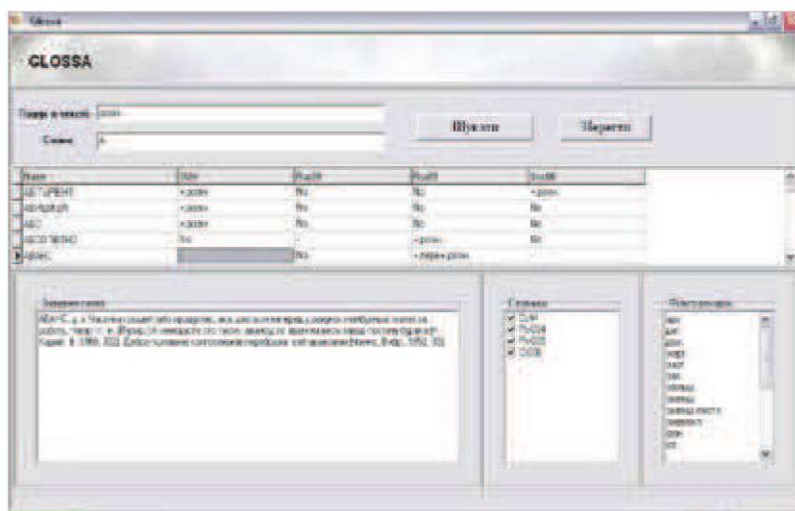


Рис. 1. Інструментальна панель комп'ютерної програми «Glossa»

Кнопка «Шукати» дає змогу пошуку запити, що вводиться у полі «Пошук в тексті» або «Слово». Якщо потрібно знайти всі слова із заданою ремаркою на окрему літеру, в полі «Пошук в тексті» вводимо ремарку, а в полі «Слово» – пошукову літеру (наприклад, А). Також може бути використаний складний пошук.

Знайдені програмою «Glossa» слова, що марковані за нашим запитом, на екрані репрезентує порівняльна таблиця. В ній наочно представлено слова та їх маркуван-

ня у різних словниках. Слова автоматично розміщено в алфавітному порядку. Під полем з кодовою назвою словника подано інформацію про маркування слова. Якщо слово марковане за нашим запитом, воно позначається знаком «+» (плюс) та відповідною ремаркою або ремарками. Якщо маркування немає, тобто слово є стильово або стилістично нейтральним, цей факт відображає знак «-» (мінус). Проте якщо при слові у реєстрі словника фігурує ремарка, але вона відмінна від тієї, яку ми задали у полі пошуку, то таке слово позначено знаком «-» та ремаркою, з якою воно подано у конкретному словнику. Знак «No» означає, що дане слово не було знайдено у реєстрі словника. Це істотно відрізняє створену нами програму «Glossa» від інформаційної системи «Lingua», у якій дані щодо стилістичної характеристики слів були репрезентовані у вигляді цифрових кодів.

Поле «Словники» відображає кодові назви, які було присвоєно для словників у базі даних. За його допомогою ми можемо вибрати, у якому саме словнику здійснювати пошук. Поле «Значення слова» візуалізує знайдене слово, його тлумачення, маркування та ілюстративний матеріал (відповідно до вибраного словника). Його активуємо натисканням на клітинку з позначками «+» або «-» у порівняльній таблиці результатів пошуку.

Поле «Фільтр ремарок» дає змогу обирати ремарки, що потрапляють до порівняльної таблиці. Клавішу «Зберегти» використовуємо для візуального представлення матеріалів дослідження. За її допомогою порівняльну таблицю відкриваємо у «Microsoft Excel» для подальшого аналізу та опрацювання результатів.

Таким чином, комп'ютерна програма «Glossa» дозволяє отримати максимально точну інформацію про використання стильово-стилістичних ремарок у словниках, а порівняльна таблиця отриманих даних демонструє зміну стилістичного забарвлення слів на синхронному та діахронному зрізі. Це дало змогу порівняти лексикографічні реєстри та систематизувати відомості про використання досліджуваних ремарок у словниках.

У межах даної роботи ми обмежили дослідження слів з ремаркою «розмовне» («розм.») на літери А, Б, В. Марковані таким чином слова ми виявили завдяки запити в полі «Пошук в тексті» позначки «розм.» та у полі «Слово» літер А, Б, В. У полі «Словники» ми обрали СУМ та РУС69. Отримані результати дають підстави зробити такі висновки:

1. Загальна кількість лексичних одиниць на літери А, Б, В, позначені ремаркою «розм.», у СУМі становлять 1355 слів. Це близько 7% усієї лексики на ці літери.
2. Ремарка «розмовне» може стосуватися як заголовного слова загалом, так і окремого значення, а також кількох значень слова або й фразеологічної одиниці (ремарку «розмовне» щодо фразеологічної одиниці у нашому дослідженні до уваги не беремо). Такі випадки використання ремарки Ф. П. Філін слушно називає стилістичною позицією [5, с. 151]. У реєстрі СУМа на літери А, Б, В нами зафіксовано 1440 стилістичних позицій ремарки «розмовне» (найчастіше ремарка «розмовне» стосується одного значення заголовного слова).
3. Чимало слів, поряд з ремаркою «розмовне», мають й інші ремарки. Такі ремарки називаємо комбінованими. Наприклад: *безбатченко* – 1. *розм.*, *заст.* Нешлюбна дитина. Комбіновані стилістичні ремарки вказують на те, що слово, залеж-

но від мовного вжитку, може набувати то одного, то іншого стилістичного забарвлення, або активізувати обидва одночасно.

4. Ми виявили спільне та відмінне у маркуванні лексики ремаркою *«розмовне»*:

- **спільне** у використанні ремарки *«розмовне»* у різних словниках свідчить про сталість стильового навантаження відповідної лексики в системі мови. Функціонування такої лексики в основному зосереджено в сфері літературного спілкування, а власне: її використання обмежується розмовно-побутовим стилем. Ми підтримуємо думку Н. Я. Дзюбишиної-Мельник та пропонуємо залишити ремарку *«розмовне»* для теоретичних курсів, а в тлумачних та перекладних словниках замінити ремаркою *«розмовно-побутове»* (у тих випадках, коли слово вживається у розмовно-побутовому стилі) [3]. Це слова на кшталт: *азукати, аліментник, бідак, багатотиражка, базарування, бородач, бозна-хто, вишняк*;

- **відмінності** в маркуванні можуть свідчити про незадовільність на даному етапі розробленості критеріїв оцінки лексики для маркування. Наприклад: слово *білявка* РУС69 виправдано подає без жодних ремарок, а у реєстрі СУМа його подано з ремаркою *«розмовне»*. У СУМі слова *бабій, брехня, балайда, вискотня* марковані ремаркою *«розмовне»*, а РУС69 подає ці слова з ремаркою *«просторічне»*.

Такий стан речей потребує перегляду зазначених ремарок, з тим, щоб у майбутніх виданнях по можливості уникнути неузгодженості у словниках.

Таким чином, комп'ютерна програма «Glossa» відкриває можливість швидко виявити марковані слова та якісно здійснювати порівняльний аналіз слів на синхронному зрізі мови. В подальшому залучення до бази даних програми словників української мови різного часу дасть змогу простежити динаміку стилістичного забарвлення слів у діахронії. Одержані, завдяки цій програмі, дані уможливили корекцію ремарок, покликані сприяти нормалізаційним процесам у мові, у тому числі усувати чужорідні елементи.

Література

1. Бурячок А. А. Функціонально-стильова оцінка слів у словнику української мови / А. А. Бурячок // Лексикологія та лексикографія. – К., 1969. – С. 78–95.
2. Волошин В. Г. Комп'ютерна лінгвістика : [навч. посіб.] / В. Г. Волошин. – Суми : Університетська книга, 2004. – 382 с.
3. Дзюбишина-Мельник Н. Я. Відображення мовної картини світу в сучасному словнику / Н. Я. Дзюбишина-Мельник // Українське мовознавство. – К., 2008. – Вип. 38. – С. 70–84.
4. Російсько-український словник / [ред. А. Кримський та С. Єфремов]. – К. : Держ. вид-во України – УАН, 1924–1933. – Т. 1–3. Електронна версія: (підготував О. Телемко). – К. : К.І.С., 2007.
5. Склярєвська Г. Н. Лексикографическая стилистика: состояние и проблемы / Г. Н. Склярєвская // Словарные категории. – М., 1988. – С. 150–155.
6. Струганець Л. В. Динаміка лексичних норм української літературної мови ХХ століття : [монографія] / Л. В. Струганець. – Тернопіль : Астон, 2002. – 352 с.

Zaitseva O. V. A computer program «GLOSSA» as method of analysis of stylistic labels

The article covered how to create and main principles of the computer program «Glossa». Examples of application of this program are given. A comparative analysis of marked words in the studied lexicographical works have been analyzed.

Key words: computerprogram, dictionary, linguistic label, marking.

Зайцева О. В. Компьютерная программа «GLOSSA» как средство систематизации и анализа стилистических помет

В статье освещена процедура создания и принципы работы компьютерной программы «Glossa». Приведены примеры применения данной программы. Проведен сравнительный анализ маркировки слов в исследуемых лексикографических трудах.

Ключевые слова: компьютерная программа, словарь, словарная помета, маркирование.