

УДК 808.3-083-55

DOI 10.31651/2076-5770-2019-1-99-103

ТАРАН Алла Анатоліївна,
кандидат філологічних наук, доцент кафедри
українського мовознавства і прикладної лінгвістики
Черкаського національного університету
імені Богдана Хмельницького
ORCID 0000-0001-8091-1477
e-mail: alla__taran@ukr.net

ПОШУКОВА СИСТЕМА Й КОРПУС ЯК ІНСТРУМЕНТ ДОСЛІДЖЕННЯ СЕМАНТИКИ СЛІВ

Нові комп'ютерні технології одержання знань, зокрема й знань про мову, відкривають заманливі перспективи для розв'язання практичних і теоретичних лінгвістичних завдань. У статті розглянуто можливості пошукової системи й корпусу мови як інструменту дослідження семантики слів. Варіанти використання пошукових систем ГУГЛ, МЕТА, ШУКАЛКА, Корпусу текстів української мови на порталі tova.info та національних корпусів слов'янських мов з дослідницькою метою проілюстровано на прикладі семантичних інновацій. Неосемантизми сприяють формуванню нових ділянок у когнітивному просторі сучасної української мови, поглиблюють вербалізацію вже відомих українській мові понять. Унаслідок дослідження запропоновано формат бази даних неосемантизмів.

***Ключові слова:** мовна практика, інтернет, пошукова система, корпус мови, неосемантизм, формат бази даних.*

Постановка проблеми. Зорієнтування україністики на використання нових комп'ютерних технологій одержання знань, нового способу дослідницької праці знаменує якісно новий етап її розвитку. У поле наукового пошуку мовознавців потрапляють бази даних про мовні об'єкти, недоступні в «докомп'ютерну» еру лінгвістики, а новий суб'єкт спілкування – комп'ютер – потребує нового апарату опису мови, нових процедур її моделювання, аналізу й синтезу. Водночас система мови, продукти її реалізації в мовній діяльності сучасного українського суспільства – різноманітні усні та письмові тексти – у комп'ютерному середовищі постають в інших іпостасях, обертаються до дослідника новими своїми гранями, що змушує його замислюватися над досі не уявленими проблемами, шукати нових, інших прийомів репрезентації мовної інформації, розробляти засоби спілкування з новим учасником комунікації – комп'ютером.

Будь-яке мовознавче дослідження розпочинається із добору матеріалу й упорядкування картотеки. Якісні й кількісні параметри зібраного матеріалу значною мірою визначають характер наукової праці. Чим багатше репрезентована екземпліфікація певного мовного явища, тим саме дослідження стає глибиннішим, обґрунтованішим і таким, яке претендує на успішніше досягнення поставлених завдань.

Аналіз останніх досліджень і публікацій. Проблемою аналізу семантики в різний час займалися такі лінгвісти, як Г. Пауль, О. О. Потебня, С. Ульман, В. В. Виноградов, О. І. Смирницький, Д. М. Шмельов, Ю. М. Караулов. В україністиці з'ясуванню природи лексичного значення, системним семантичним зв'язкам присвячено наукові студії Л. А. Булаховського, М. П. Кочергана, О. Г. Муромцевої, Л. С. Паламарчука, В. М. Русанівського, О. О. Селіванової. У процесі вивчення семантичної структури слова посилюється інтерес до смислових зрушень, до змін у семантиці слів. Цінними для осмислення процесів вторинної номінації є публікації Є. А. Карпіловської, Н. Ф. Клименко, Г. Б. Мінчак, І. А. Самойлової, Л. В. Струганець та інших науковців.

Мета статті – простежити відображення зміни семантики слів у пошуковій системі й корпусі мови.

Виклад основного матеріалу. Поява інтернету й створення корпусів текстів різних національних мов значно спростили підготовчий, але надзвичайно значущий для подальшого розгортання дослідження етап. Користувачу достатньо ввести слово в пошукову систему відповідного корпусу, щоб отримати повний індекс його ж слівформ у найближчих і розширених контекстах із відповідною адресацією. Створення емпіричної бази дослідження в такий спосіб забезпечує мінімум витрат зусиль і часу на збирання даних, обсяг яких водночас суттєво збільшиться. З одного боку, це спростило й полегшило, але, з іншого боку, ускладнило процес аналізу й осмислення такої кількості прикладів.

Інтернет і його кількісні дані, які супроводжують результати пошуку окремих слів і словосполучень, засвідчують досить об'єктивну картину функціонування мови, оскільки в цих даних і їхніх «квантитативних коефіцієнтах» кожного слова виявляється вибірка вихідного матеріалу для різноаспектних узагальнень, що заслуговують на увагу. Проте постає питання, наскільки показова й збалансована текстова база, з якої ми беремо приклади, наскільки їй можна вірити. Стандартна операція пошуку тексту чи окремого слова в руках філолога може стати корисним інструментом лексикографічного дослідження, що ймовірно закінчиться лексикографічним продуктом – словниковою статтею.

За об'єкт обговорення порушених проблем ми обрали неосемантизми, пов'язані з подальшою аспектуалізацією в українській мові певних понять. Поява в семантичній структурі слова нових значень є сьогодні в українській мові одним з потужних механізмів поповнення його лексикону. Термін неосемантизм є ширшим за обсягом змісту, ніж загальноприйняті вторинна номінація, семантична, або внутрішньослівна деривація, оскільки він охоплює і формування нових значень на власному ґрунті в структурі вже наявних слів, і появу нових запозичених слів-значень (первинних і вторинних), і незалежну міжслівну деривацію з повторним використанням тієї ж моделі словотворення. Тому терміном неосемантизація можна позначити всі результати оновлення семантики слова на певному етапі розвитку мови. Загалом активні семантичні зміни стимульовані потребою української мови здобувати нові поля невідомих досі понять.

Потребою для мовознавця є формування бази даних неосемантизмів, що дає орієнтир для перевірки функціонування таких мовних одиниць у сучасних текстах, для визначення перспектив їхнього усталення в системі української мови. Для потреб вивчення динаміки системи мови в усьому спектрі її виявів, на думку Є. А. Карпіловської, необхідне створення «портретів слів», які подавали б відомості не лише про спектр уживання певного слова в текстах різних функціональних стилів, а й про його формальні, семантичні й функціональні властивості в тексті і в системі мови, уможлиблювали б відстежування динаміки [1, с. 423]. На нашу думку, зручним для систематизації й узагальнення відомостей про неосемантизми буде такий формат бази даних:

1. Реєстрове слово.
2. Нова дефініція
3. Дефініції: за Словником української мови в 11 т. (1970 – 1980), іншими словниками, виданими до 1991 р.

4. Типи семантичних змін. У сучасній лінгвістичній літературі наявні різні класифікації типів семантичних змін, що спричинюють виникнення похідних значень словесних одиниць. Як зазначає О. Я. Бадеева, логічні класифікації зміни значень ґрунтуються на розгляді змін в обсязі поняття слова. Творці генетичних класифікацій враховують причини змін, а в основі психологічних класифікацій змін значень – різні типи психологічних асоціацій [2, с. 13]. Тому цілком очевидно, що мовознавці виокремлюють кілька варіантів смислових змін, по-різному їх називають й інтерпретують, зокрема: розширення (генералізація) значення; звуження (спеціалізація / конкретизація) значення; переосмислення (шляхом метафоризації або метонімізації, приховане запозичування, дія аналогії, унаслідок яких з'являються омоніми). Приховане запозичування – запозичення нового значення для слова, форма якого вже знайома носіям мови-реципієнта. Щоб з'ясувати причину появи неосемантизму, потрібно встановити наявність / відсутність його зв'язку з

уже наявною семантичною парадигмою слова або ж з реаліями української дійсності. Серед напрямів функціонального розвитку лексику виокремлюють детермінування, термінування, зміни функціонально-стилістичної належності лексичних одиниць. На жаль, у лінгвістиці немає єдиної загальноприйнятої універсальної схеми класифікації семантичних змін, що задовольняла б усіх мовознавців.

5. Синтагматичні відношення. За нашими спостереженнями, вторинні значення мають більшу синтагматичну зумовленість, але менш закріплені парадигматично.

6. Епідигматичні й парадигматичні відношення.

Варіанти використання пошукової системи й корпусів текстів з дослідницькою метою проілюструємо на прикладі неосемантизму *синдром*, що зафіксовано зі значеннями: «комплекс характерних для певного захворювання симптомів» [СУМ, IX, с. 181]; синдром [франц. syndrome грец. Syndrome – той, що разом біжить] – мед. сполучення ознак (симптомів якогось захворювання) [ССІС, с. 625]; СНІД – синдром набутого імунodefіциту [r2u.org.ua]. Сьогодні це слово почало частіше вживатися в мові політикуму зі значенням «комплекс ознак, вирізняльних для певного явища, об'єкта». Тип семантичної зміни – детермінування. Синтагматичні відношення найменувань синдрому виявлено у формуванні складених номінацій: «**Синдром Насірова**»: Вакарчук поставив діагноз державній системі (УП, 11.12.2018); ГПУ звинуватила Денисенку у «**синдромі Парасюка**» і завела справу (УП, 27.12.2015); Росія: **синдром Людоїдки Еллочки** (УП, 28.07.2014); «**Донбаський синдром**». Скільки людей з ним зараз в Україні? (УП, 24.05.2016); **Артемівський синдром**. Прифронтова (УП, 04.08.2014); «**Синдром героя**»: як повернутися до мирного життя (УП, 28.10.2014); «**Синдром завтра**»: чому Порошенко відклав ключові податкові реформи (УП, 13.03.2018); **Кінець ери долара? Китайський синдром** (26.11.2015) **Підкуп на виборах: як подолати синдром 205 округу** (УП, 17.10.2015). Подібні «синдроми» вирізняють явища, небажані для носіїв української мови, які вони оцінюють як загрозу їхньому мирному, стабільному життю; **синдром черствості** (УП, 21.06.2016); **синдром відсутності культури** (УП, 14.08.2015); **синдром стурбованості** (УП, 14.09.2015).

Поява нового значення зумовлена контекстом функціонування слова й внутрішніми семантичними можливостями слова, адже стрижень значення – комплекс ознак чогось – залишається в усіх словосполучах. Суспільний контекст надає об'єкти з такими комплексами ознак.

Ми протестували пошукові системи ГУГЛ, МЕТА і ШУКАЛКА на формування нового значення слова *синдром*. ГУГЛ за 52 секунди запропонував 19 000 000 результатів пошуку, проте перші 10–15 сторінок не репрезентують нових контекстів функціонування слова. Як зазначає Сідні Лендау, загальна проблема, що постає під час роботи з цитатною картотекою, – це брак прикладів конкретного слововжитку; загальна проблема під час роботи з корпусом – надмір прикладів (“влучень” hits) [3, с. 32].

Можливості розширеного пошуку ГУГЛ передбачають уведення словосполучки: зі словосполучкою *донецький синдром* зафіксовано 35 результатів пошуку за 23 с. Як бачимо, особливого полегшення пошуку немає – система унаочнює вже наявні знання про мову, а отже, не передбачає виявлення в ній нових явищ. Таке знання мовознавець уже повинен мати, створюючи запит для пошуку інформації про слово. Українська пошукова система ШУКАЛКА показує результати лише російською мовою із відомим значенням. МЕТА не фіксує статистики і контекстів із новим значенням.

І лише Корпус текстів української мови (КТУМ) на порталі mova.info у підкорпусі «публіцистичні тексти» відображає окремі словосполучки на зразок *трамвайний синдром*, *пострадянський синдром*.

Для порівняння протестували базу даних Neomat, укладену працівниками відділу сучасної лексикології та лексикографії при Інституті чеської мови Академії наук Чеської Республіки. У статтях Neomat не подано тлумачення нового слова *синдром*, однак контекст уживання лексичної одиниці відображає її нову лексичну або синтаксичну сполучуваність, що дає змогу сформулювати дефініцію слова. У Чеському національному корпусі на основі

формул [tag="A.*"] [lemma="syndrom"] та [lemma="syndrom"] [tag="N.*"] визначаємо контексти вживання лексеми. Результати пошуку програма репрезентує у вигляді конкордансу – типу словника, що подає до кожної реєстрової одиниці (слова чи словоформи) усі або вибіркові контексти її вживання, тобто уривки тексту, де вона функціонує. Контекст, який конкорданс подає до кожної лексеми, виконує дуже важливі функції: він потрібний для вибору та актуалізації певного значення; модифікує смисл у межах одного значення; уточнює значення (розширюючи чи звужуючи клас його денотатів); формує оказіональні значення; необхідний для десемантизації тощо.

У Національному корпусі польської мови також за контекстами можна набагато швидше з'ясувати нове значення слова. Перші результати пошуку вже засвідчують *китайський синдром* із новим значенням.

Висновки та перспективи дослідження. Одержання сучасної моделі системи української мови передбачає не лише аналіз й опис форми та семантики лексичних одиниць, їхнє функціонування в текстах, а й створення надійних інструментів для відстежування перебігу її динаміки. Як бачимо, важливою особливістю даних Інтернету як лінгвістичного джерела є можливість безпосереднього звернення до контекстів, у яких зафіксовані вживання потрібного слова. Це важливо для переходу від кількісних характеристик до якісних, від абстрактного інваріанта, представленого в квантитативному коефіцієнті слова, до його семантичних, стилістичних й інших варіантів. Проте пошукова система й корпуси мови не можуть претендувати на повноту відображення динаміки мови, змін у ній, якими б потужними вони не були. На сьогодні, на жаль, пошукові системи потребують удосконалення, адже не відомо, як упорядкувати результати пошуку, було би добре, якби функціонували фільтри того, що ми шукаємо. Або ж мовознавцям треба створювати свої, спеціалізовані пошукові системи, зорієнтовані на виконання наукових завдань.

Список використаної літератури

1. Карпіловська Є. А., Кислюк Л. П., Клименко Н. Ф., Критська В. І., Пуздирєва Т. К., Романюк Ю. В. Вплив суспільних змін на розвиток української мови : Монографія / Відп. ред. Є. А. Карпіловська. – Київ : Видавничий дім Дмитра Бураго, 2017. – 444 с.
2. Бадеева Е. Я. Метонимия имени в общественно-политической лексике английского языка: когнитивный и прагматический аспекты: автореф. дис. ... канд. филол. наук. – Москва, 2005. – 23 с.
3. Сидні І. Лендау Словники: мистецтво та ремесло лексикографії / Лендау Сидні І. – К.: К.І.С., 2012. – 480 с.
4. Роль человеческого фактора в языке: Язык и картина мира / Б. А. Серебренников, Е. С. Кубрякова, В. И. Постовалова, В. Н. Телия, А. А. Уфимцева. – М.: Наука, 1988. – 216 с.
5. Селіванова О. О. Проблема значення в мовознавстві / О. О. Селіванова // Вісник Черкаського університету. Сер. Філологічні науки. – Черкаси, 2018. – № 2. – С. 3–10.
6. Телия В. Н. Русская фразеология. Семантический, прагматический и лингвокультурологический аспекты / В. Н. Телия. – М.: Школа “Языки русской культуры”, 1996. – 288 с.
7. Тараненко О. О. Полісемічний паралелізм і явище семантичної аналогії / О. О. Тараненко. – К.: Наукова думка, 1980. – 115 с.
8. Пауль Г. Принципы истории языка / Г. Пауль. – М.: Изд-во иностр. лит-ры, 1960. – 500 с.
9. Ульманн С. Семантические универсалии: пер. с англ. // Новое в лингвистике. – Вып. V: Языковые универсалии. – М.: Прогресс, 1970. – С.250 – 299.

References

1. Influence of social changes on the development of the Ukrainian language (2017). [resp. ed. I. A. Karpilovska]. Kyiv: Vidavnichiy Dim Demetra Burago (in Ukr.).
2. Badoeva, E. J (2005). *Metonymy of the name in the social and political vocabulary of English : cognitive and pragmatic aspects*. Thesis of Doctor. Philology 10.02.01 «Russian languages». Moscow (in Russ.)
3. Sidney I. Landau (2012). *Dictionaries: The Art and Craft of Lexicography*. Kyiv: K.I.S. (in Ukr.).
4. *A role of human factor is in language: Language and picture of the world* (1998) / of B.A.Serebrennikov, E. S. Kubriakova, V. I. Postovalova, V. N. Teliya, A. A. Ufymceva. M.: Science (in Russ.).
5. Selivanova, O. O. (2018). Problems of meaning in linguistics. *Visnyk Cherkaskoho universytetu (Bulletin of the University of Cherkasy)*, 2, 3–11. doi: 10.31651/2076-5770-2018-2 (in Ukr.)
6. Teliya, V. N. (1996). Russian phraseology. Semantic, pragmatic and linguocultural aspects. M.: School «Languages of the Russian culture». (in Russ.).
7. Taranenko, O. O. (1980). *Polysemic parallelism and the phenomenon of semantic analogy*. K.: Scientific Opinion (in Ukr.).

8. Paul, G. (1960) *Principles of the History of Language*. Moscow: Publishing house of foreign literature (in Russ.).
9. Ulmann, S. (1970). *Semantic universalies: Trudged. with an eng // New in linguistics. Prod. V: Language universalies. M.: Progress, (in Russ.)*.

TARAN Alla Anatoliyivna,

Candidate of Philology, Associate Professor of Ukrainian and applied linguistics, The Bohdan Khmelnytsky National University of Cherkassy

e-mail: alla_taran@ukr.net

SEARCH SYSTEM AND POWER AS AN INSTRUMENT FOR EXPLORING THE SEMANTIC OF WORDS

Abstract. Introduction. *The emergence of the Internet and the creation of the body of texts of different national languages significantly simplified the preparatory, but extremely important for further deployment of the study stage. It is enough for the researcher to enter the corresponding word body into the search engine in order to obtain the full index of its own word forms in the nearest and extended contexts with the corresponding addressing. Creating an empirical research base in this way provides a minimum of effort and time to aggregate data, which at the same time will increase significantly. On one hand, it simplified and facilitated, but, on the other hand, it complicated the process of analysis and comprehension of such an amount of examples.*

Purpose. *Find out the capabilities of the search engine and the body of the language in the study of changing the semantics of words.*

Methods. *According to research material in the scientific studies, the following methods are used: descriptive (linguistic techniques of observation and generalization), component analysis method (including analysis of dictionary definitions), method of contextual analysis and method of comparative analysis.*

Results. *An important feature of the Internet as a linguistic source is the ability to directly refer to the contexts in which the use of the right word is recorded. This is important for the transition from quantitative characteristics to qualitative, from the abstract invariant, represented in the quantitative coefficient of the word, to its semantic, stylistic and other variants. However, for today, unfortunately, search engines need to be improved, since it is not known how to streamline the search results, it would be nice if the filters of what we are looking for would work. Or linguists need to create their own, specialized search engines, oriented to the fulfillment of scientific tasks.*

Originality. *The variants of the use of search engines GOOGLE, META, SHUKALKA, the Body of texts of the Ukrainian language on the portal mova.info and the buildings of Slavic languages or research purposes are illustrated by the example of semantic innovations. As a result, the format of the database of neosemanticisms is proposed.* **Conclusion.** *Перспективу подальших наукових пошуків убачаємо в дослідженні функціонального потенціалу семантичних інновацій.*

Conclusion. *The prospect of further scientific research is seen in the study of the functional potential of semantic innovations.*

Key words: *language practice, internet, search engine, language body, neosemanticism, database format.*

Надійшла до редакції 16.01.19

Прийнято до друку 21.01.19