

УДК 81'374

DOI: 10.18523/2616-8502.2019.2.51-55

Ірина Олександрук

<https://orcid.org/0000-0003-2701-1714>

## ЕКСПЛІКАЦІЯ ЛЕКСИКОГРАФІЧНОЇ ІНФОРМАЦІЇ В ЕЛЕКТРОННИХ СЛОВНИКАХ

*У статті йдеться про роль семантичних досліджень для створення інтелектуальних систем автоматичного опрацювання мовної інформації та розв'язання проблеми експліцитного представлення в них лінгвістичної інформації. Окреслено підхід до створення мовно-інформаційного інструментарію для проведення системного аналізу семантики лексичних одиниць української мови шляхом опису їхніх лексичних значень у поєднанні з описом прояву феноменологічних властивостей мови на лексичному рівні.*

**Ключові слова:** електронний словник, експліцитність, автоматизована лексикографія, тлумачення.

### Вступ

Зростання ступеня інформативності й комп'ютеризації людської діяльності з часом набирає все більших темпів, що своєю чергою загострює проблему підвищення ефективності систем автоматичного оброблення інформації, яка передається засобами природних мов. Необхідним є підвищення ступеня інтелектуалізації систем автоматичної обробки тексту (АОТ) шляхом удосконалення потужних за швидкістю і обсягом перероблюваної інформації лінгвістичних процесорів автоматичного розпізнавання змісту письмових текстів. На сьогодні дедалі більшої актуальності набуває роль електронних словників в інформаційному середовищі, зростає інтерес до опрацювання семантичної інформації програмними засобами.

### Теоретичне підґрунтя

Комп'ютерна лексикографія поставила нові завдання перед лінгвістичною семантикою щодо вибору ознак опису семантичних явищ,

розробки методів аналізу семантики, що забезпечували б представлення результатів цього аналізу у формі, доступній для розпізнавання їх комп'ютером, а також забезпечували б можливість використання відповідної інформації в системах автоматичного оброблювання текстів (АОТ). У вітчизняному мовознавстві новітні дослідження з питань представлення різноманітної лінгвістичної інформації в електронних словниках стали предметом досліджень у працях М. М. Пешак, В. А. Широкова, Т. О. Грязнухіної, Л. Л. Шевченко, Н. П. Дарчук, І. В. Шевченка, О. Г. Рабульця, Н. М. Сидорчук, І. В. Остапової, Т. П. Любченко, М. В. Надутенка, Н. М. Сухарини, Є. В. Купріянова, І. М. Кульчицького та ін. Цілковитим послідовно й достатньо формалізовану теорію електронних словників побудовано в Українському мовно-інформаційному фонді НАН України. Ефективність застосування теорії проілюстровано на різноманітних репрезентативних масивах лінгвістичного матеріалу (Широков, 1998; Широков, 2011).

Метою статті є формулювання принципів семантичної інформації в електронних тлумачних словниках (ЕТЛС), визнаних у лінгвістиці найбільш надійним і репрезентативним джерелом цієї інформації, згідно з головною вимогою до електронних словників щодо подання в них інформації в експліцитному вигляді і структурованій таким чином, щоб була можливість здійснення її оброблення програмними засобами. Без забезпечення цієї вимоги не можна говорити ні про виконання ЕТЛС резидентної функції (коли в системі автоматичного оброблення тексту (АОТ) семантична інформація екстрагується зі словника за допомогою програми і автоматично стає вихідною для роботи іншої програми), ні про використання ЕТЛС як інформаційно-програмного інструментарію семантичних досліджень, здійснюваних в автоматизованому режимі (коли до аналізу автоматично сформованих масивів вихідного лінгвістичного матеріалу залучають різні види семантичної інформації, екстрагованої як із різних зон ЕТЛС (у тому числі з тексту тлумачення), так і з інших лексикографічних джерел, зібраних в інтегрованій лексикографічній системі (ІЛС).

Як показав попередній досвід здійснених в Українському мовно-інформаційному фонді НАН України (УМІФ) досліджень зі створення електронних семантичних словників – синонімічного і словника паронімів, а також електронного словника дієслівного керування, проводити аналіз на великих масивах лінгвістичного матеріалу за багатьма параметрами доцільно шляхом створення спеціальних лінгвістичних баз даних (ЛБД), орієнтованих на опис конкретного семантичного явища із залученням методів комп'ютерної лінгвістики.

Завдання нашого дослідження – створення мовно-інформаційного інструментарію для проведення аналізу семантики лексичних одиниць, тлумачення значень яких у Тлумачному словнику української мови подано за відсильними формулами, які лише встановлюють зв'язок між семантичними властивостями конкретного слова, що має похідну основу, і лексичним значенням слова, від якого слово утворено. Наприклад: *АВАНГА#РДНИЙ*, а, е. 1. Прикм. до авангард. *АБСУ#РДНИЙ*, а, е. Стос. до абсурду; безглуздий, нісенітний (Український мовно-інформаційний фонд НАН України (УМІФ), 2001–2018).

## Методи та матеріал дослідження

Зазначені дослідження здійснюються на матеріалі ~19 тисяч відносних прикметників, автоматично екстрагованих з Електронного тлумачного словника української мови засобами Віртуальної лексикографічної лабораторії «Словник української мови у 20 томах» (ВЛЛ СУМ-20). У словнику тлумачення цих прикметників представлені формулами тлумачення «Прикметник до...» (*АВІА#ТОРСЬКИЙ*, а, е. Прикм. до авіатор.) і «Стосовно до...» (*АКЦІОНЕРНИЙ*, а, е, фін. Стос. до акції (див. а#кція) (Віртуальна лексикографічна лабораторія (далі – УМІФ), 2001–2018).

Семантичні репрезентації такого типу в (Широков, 2012) кваліфіковано як *формули квазісемантики*. На відміну від поширених в ТЛС відсильних формул «те саме, що...», «див. ...», які дають змогу програмно експлікувати значення слів за допомогою операції перенесення інформації із зони «тлумачення» слова, записаного після «що» чи «див.» у конкретній формулі, в зону «тлумачення» відповідного слова, «Прикметник до...» і «Стосовно до...» тільки вказують на один із компонентів лексичного значення відносних прикметників, а саме на його зв'язок із семантикою слова, від основи якого походить прикметник.

Мотивування значення відносного прикметника семантикою словотвірного слова, від якого цей прикметник утворено, за допомогою певних суфіксів, є визначальним моментом у семантиці певних лексичних груп також й інших лексико-граматичних класів слів. До них належать:

- віддієслівні іменники з формулою тлумачення семантичної теми (СТ) в ІЛС «Дія за значенням...» *ДОБИВА#ННЯ*, я, с. Дія за знач. *добивати* (УМІФ, 2001–2018); «Дія і стан за значенням» *ВИЖИВА#ННЯ*, я, с. Дія і стан за знач. *виживати* 1, 2, 4 (УМІФ, 2001–2018); відприкметникові іменники – СТ «Властивість і якість за значенням» *МОНУМЕНТА#ЛЬНІСТЬ*, ності, ж. *Властивість і якість за знач. монумента#льний* 2, 3 (УМІФ, 2001–2018); «Абстрактний іменник до...» *ВИДОЗМІ#НЮВАНІСТЬ*, ності, ж. *Абстр. ім. до видозмі#нюваний* (УМІФ, 2001–2018); «Властивість і стан за значенням» *БЕЗЖУ#РНІСТЬ*, ності, ж. *Властивість і стан за знач. безжурний*; «Якість за значенням» *ЧОРНОТА#*, и, ж. 1. *Якість за знач.*

чорний 1; чорний, темний колір чого-небудь (УМІФ, 2001–2018); «Стан за знач.» БЕЗ-ШЛЮ#БНІСТЬ, ності, ж. Стан за знач. безшлюбний (УМІФ, 2001–2018);

- дієприкметники з формулами «Дієприкметник пасивний до...» АВІЗО#ВАНИЙ, а, е, фін. Дієпр. пас. до авізувати (УМІФ, 2001–2018); «Дієприкметник активний до...» ВІДМЕ#РЗЛИЙ, а, е. Дієпр. акт. до відме#рзнути, відме#рзти (УМІФ, 2001–2018);
- прислівники з формулами «Прислівник до...» АДЕКВА#ТНО. Присл. до адекватний (УМІФ, 2001–2018).

Утім, у лексичному значенні прикметники з формулами тлумачення «Прикметник до...» і «Стовсно до...» на відміну від зазначених вище лексичних груп із відсильними формулами тлумачення, якими передбачено визначення значення слова через семантичні властивості однокореневих із ним слів, виділяють ще один базисний мотивувальний фактор, пов'язаний із специфічним реляційним характером ознаки, позначуваної саме відносними, а не якісними прикметниками. Експліцитним представленням реляційної ознаки в тексті тлумачення є семантичні відношення\*, які можуть встановлюватися між референтами, позначуваними прикметником та синтаксично зв'язаними з ним іменниками у сполученнях «відносний прикметник + іменник».

Методи дослідження – теорія семантичних станів, компонентний аналіз, елементи кількісного аналізу, корпусні технології.

### Результати дослідження

У дослідженні для опису лексичного значення слів застосовано формалізм теорії семантичних станів (ТСС) мовних одиниць, розробленої В. А. Широковим (Широков, 2018, с. 16–20), яка апробована в дослідженнях семантики прийменників (Бугаков, 2006), при створенні електронного словника синонімів. З позиції феноменологічного підходу до вивчення мови, описаного в книзі В. А. Широкова (Широков, 2008, с. 16–20), є цілком правомірним віднесення семантичних відношень до числа диференційних ознак, за якими визначаємо семантичний стан відносних прикметників.

\* Семантичні відношення визначаємо як виражені мовними засобами відношення між предметами об'єктивної дійсності.

Детермінований цими відношеннями мовний контекст інтерпретуємо через лексичну сполучуваність ад'єктива з іменником, представлену в семантичних класах мотивувальної основи прикметника і зв'язаного з ним в тексті іменника.

Кількість значень (відтінків у межах значення) конкретного прикметника перебуває в прямій залежності від кількості семантичних відношень, які реалізуються ними в сполученнях з іменником.

Зокрема, прикметнику БАГАТИ#РСЬКИЙ, представленому в ТЛС одним значенням: «прикметник до багатир», у ЕТЛС відповідатиме значення: «прикметник від багатир у значенні “той, хто має багатство (майно, гроші)”», з відтінками.

У сполученні з іменником на позначення родинних зв'язків (онук, син, дочка, племінник) – семантичне відношення «генеративне». В ЕТЛС відтінок 1.1: «який перебуває у родинних зв'язках з багатієм».

У сполученні з іменником на позначення видів майна (земля, обійстя, хата) – семантичне відношення «посесивне». В ЕТЛС відтінок 1.2: «який належить багачеві».

У сполученні з іменником на позначення групи людей (група, купа, бік) – семантичне відношення «партитативне». В ЕТЛС відтінок 1.3: «який має у своєму складі багачів».

У сполученні з іменником на позначення дії – семантичне відношення «суб'єктне». В ЕТЛС відтінок 1.4: «який виконується, здійснюється багачем».

Прикметник БОГАТИ#РСЬКИЙ, який є паронімом до багатир, у ТЛС представлено одним значенням: «прикметник до богатир», у ЕТЛС відповідатиме значення: «прикметник від богатир у значенні «кремезна, працююча, відважна людина», з відтінками.

У сполученні з іменником на позначення дії – семантичне відношення «суб'єктне». В ЕТЛС відтінок 1.1: «який має суб'єктом богатиря».

У сполученні з іменником на позначення часу (минуле) – семантичне відношення «темпоральне». В ЕТЛС відтінок 1.2: «який є часовою характеристикою існування богатирів».

У сполученні з іменником на позначення жанрів, видів (епос, билина, сюжет) – семантичне відношення «інформативне». В ЕТЛС відтінок 1.3: «який є носієм інформації про богатирів».

У сполученні з іменником на позначення предметів, зброї (знаряддя, меч, щит) – семантичне відношення «посесивне». В ЕТЛС відтінок 1.4: «який належить богатиреві».

У сполученні з іменником на позначення групи людей, колективу (рать) – семантичне відношення «партитативне». В ЕТЛС відтінок 1.5: «який має у своєму складі богатирів».

Джерелом для встановлення лексичної сполучуваності є текст. Процес визначення текстових контекстів вживання словосполучень «А + І» у «ручному режимі» надзвичайно трудомісткий. У нашому дослідженні, яке ми проводили у межах Інтегрованої лексикографічної системи, компонентом якої є Український національний корпус обсягом понад 190 мільйонів слововживань, вибір контекстів здійснюється автоматично із використанням програм автоматичного морфологічного аналізу (АМА) тексту та лематизації. Інструментом АМА і лематизації в ІЛС слугує Електронний граматичний словник української мови (Шевченко, Рабулець, Широков, 2005, с. 124).

Далі подано опис зон ЛБД «Семантика відносних прикметників», побудованої як інформаційно-програмний інструмент для дослідження лексичного значення відносних прикметників із похідною основою, що здійснюється за такими параметрами:

Зона 1. Лексема, лексико-семантичні варіанти (ЛСВ) з одноцифровим індексом на позначення номера значення для багатозначних слів, відтінків у межах ЛСВ із двоцифровим індексом, що вказує номер значення і записаний через крапку номер відтинку.

Зона 2. Словотворчий формант (СФ) прикметника.

Зона 2.1. Семантика словотворчого форманта.

Зони орієнтовано на використання ЛБД для дослідження ролі словотвірної семантики у формуванні лексичного значення слів із похідною основою.

Зона 3. Мотивувальна основа: іменник, від якого утворено прикметник. Іменник записують, формують з індексами, зазначеними у формулі тлумачення прикметника в ТЛС; ці індекси вказують у випадку багатозначного іменника на конкретне значення (відтинки значення), які є визначальними. Наприклад: *АВТОМАТИ#ЧНИЙ, а, е. 1. Який діє, працює механічно; самодіючий. БРА#ТНІЙ, я, е. 1. Прикм. до брат 1, 2, 5.*

*БА#ЛОВИЙ, а, е. «Прикм. до бал 2.2».* (УМІФ, 2001–2018). У випадку лексичних омонімів або багатозначного іменника індекси слугують формальними показниками конкретних значень мотивувальної основи прикметників, які мають бути експліковані в тлумаченнях Електронного тлумачного словника. Якщо в ТЛС словникова стаття має вигляд: *АГІТА#ТОРСЬКИЙ, а, е. Прикм. до агітатор*, то в ЕТЛС вона набуває вигляду: *АГІТА#ТОРСЬКИЙ, а, е. прикметник від агітатор у значенні «стосовно особи, що проводить агітацію».* Відсутність індекса у багатозначних іменників означає необхідність аналізувати всю його тлумачну частину словникової статті.

Зона 4. Тлумачна стаття з відтинками значень. *БОГАТИ#РСЬКИЙ, а, е. 1. Оспіваний у народній творчості герой, хоробрий воїн – людина-велетень із надзвичайною силою та відвагою. 2. перен. Про дужу, кремезну, працювиту, відважну людину. 2.1. Про велику, могутню тварину, рослину, машину і т. ін.* (УМІФ, 2001–2018).

Зона 5. Семантична тема. Лексичне значення похідної основи.

Зона 6. Стилїстична позначка.

Зона 7. Тлумачення з ТЛС з вказівкою на значення похідної основи. *БОГАТИ#РСЬКИЙ, а, е. Стос. до народних героїв – богатирів* (УМІФ, 2001–2018).

Зона 8. Семантичний клас іменника (у сполученні А + І). Наприклад, СК: літературний твір (епос, билина, пісня, легенда, дума); СК: предмет (зброя, знаряддя); СК: час (епоха, минуле); СК: дія (битва, двобій).

Зона 9. Семантичні відношення у сполученні А + І: R [A ↔ I; (A + I)].

Зона 10. Ілюстрація з УНЛК до кожного значення та відтинку значення.

Зона 11. Коментар.

## Висновки

Роль семантичних досліджень для створення інтелектуальних систем автоматичного опрацювання мовної інформації є надзвичайно важливою. Завдання створення мовно-інформаційного інструментарію для проведення аналізу семантики лексичних одиниць української мови можна вирішити через застосування методів комп'ютерної лінгвістики в семантичних дослідженнях на шляху побудови інтегрованих з електронним тлумачним словником лексико-

графічних баз даних для багатопараметричного опису семантичних властивостей слів. Здійснення зазначених видів аналізу на матеріалі 20-томного Словника української мови

сприятиме розв'язанню однієї з основних проблем комп'ютерної лексикографії – експліцитного подання інформації в електронних словниках.

#### Список використаної літератури

- Бугаков О. В. Функціонування прийменників в українському тексті: морфологічний та семантико-синтаксичний аспекти : дис. ... канд. філол. наук: 10.02.01 / О. В. Бугаков. – Київ, 2006.
- Віртуальна лексикографічна лабораторія «Словник української мови». Український мовно-інформаційний фонд НАН України, 2001–2018 [Електронний ресурс]. – Режим доступу: <http://lcorp.ulif.org.ua/LSlist/>. – Назва з екрана.
- Грязнухина Т. А. Эксплицитное представление семантической информации в Электронном толковом словаре украинского языка (синонимическая и паронимическая параметризация) / Т. А. Грязнухина, Т. П. Любченко, К. М. Якименко // Прикладная лингвистика и лингвистические технологии: MegaLing-2012 : сборник научных трудов. – Киев, 2012. – С. 79–91.
- Корпусна лінгвістика : [монографія] / В. А. Широков, О. В. Бугаков, Т. О. Грязнухіна та ін. – Київ : Довіра, 2005. – 470 с.
- Лінгвістичні та технологічні основи тлумачної лексикографії: [монографія] / В. А. Широков, В. М. Білоноженко, О. В. Бугаков та ін. – Київ : Довіра, 2010. – 293 с.
- Шевченко И. В. Электронный грамматический словарь украинского языка / И. В. Шевченко, А. Г. Рабулец, В. А. Широков // Труды Международ. конфер. «Прикладная лингвистика в поиске новых путей». – Санкт-Петербург : Изд-во Осипов, 2005. – С. 124–29.
- Широков В. А. Інформаційна теорія лексикографічних систем : [монографія] / В. А. Широков. – Київ : Довіра, 1998. – 331 с.
- Широков В. А. Семантичні стани мовних одиниць та їх застосування в когнітивній лексикографії / В. А. Широков // Мовознавство. – 2005. – № 3–4. – С. 47–62.
- Широков В. А. Комп'ютерна лексикографія : [монографія] / В. А. Широков. – Київ : Наукова думка, 2011. – 352 с.
- Широков В. А. Системна семантика тлумачних словників / В. А. Широков // Акцентологія. Етимологія. Семантика. До 75-річчя В. Г. Скляренка. – Київ : Наукова думка, 2012. – С. 487–510.
- Широков В. А. Граматичні системи: феноменологічний підхід : [монографія] / В. А. Широков, Т. П. Любченко, І. В. Шевченко, К. В. Широков. – Київ, 2018. – С. 16–20.

*Iryna Oleksandruck*

## EXPLICATION LEXICOGRAPHIC INFORMATION IN ELECTRONIC DICTIONARIES

**Background.** *The level of informativeness and computerization in human activity exacerbates the problem of increasing the efficiency of automatic processing systems and explicit representation of linguistic information in them, which is transmitted by means of natural languages. Today, the role of electronic dictionaries in the information environment is increasing; thus, the interest in the processing of semantic information by software is enhancing.*

**Purpose.** *The purpose of the article is to formulate the principles of semantic information in electronic interpretative dictionaries, recognized in linguistics as the most reliable and representative source of this information, according to the main requirement, to electronic dictionaries for the presentation of information in them in an explicit form and structured in such a way that it was possible to carry out its processing software tools.*

**Methods.** *The method of component analysis, the theory of semantic states, corpus technologies, and elements of quantitative analysis are applied.*

**Results.** *The constructed LBD “Semantics of relative adjectives” combines the function of the automatic workplace (ARM) of the researcher and the programmatic and informational tools for conducting semantic studies of adjectives.*

**Discussion.** *The task of creating a language and information tool for semantic analysis of the semantics of lexical units of the Ukrainian language will be solved using methods of computer linguistics in semantic research on the way of constructing a lexicographic database integrated with the electronic explanatory dictionary for a multiparameter description of the semantic properties of words.*

**Key words:** electronic dictionary, explicitness, semantics, automated lexicography, interpretation.

*Матеріал надійшов 30.12.2018*