

Послідовність подання тегів можлива така: спочатку `<name type=org>`, а потім інші теги залежно від послідовності об'єктів тегування. Наприкінці назви організації після всіх `</>` наводимо `<name type=org>`.

Література

1. Lou Burnard, C.M.Sperberg-McQueen TEI Lite. – Електронний ресурс: <http://www.tei.org>, 1995.

ОСНОВНІ КРИТЕРІЇ ВСТАНОВЛЕННЯ ПАРАМЕТРІВ СИНОНІМІЧНОСТІ СЛІВ В ЕЛЕКТРОННОМУ СЛОВНИКУ СИНОНІМІВ УКРАЇНСЬКОЇ МОВИ

© Олена Устимець, 2007

Український мовно-інформаційний фонд НАН України (Київ)

УДК: 81'374.221

У статті проаналізовано основні критерії встановлення синонімічності слів під час побудови Електронного словника синонімів української мови (ЕСС) на базі комп'ютерної версії „Словника синонімів української мови”. Визначення критерію семантичної подібності слів запропоновано вирішувати через встановлення схожості їхніх тлумачень, спираючись на теорію семантичних станів мовних одиниць.

Проблема лексичних синонімів та синонімії була й залишається актуальною, цікавою й складною для лінгвістів. Особливого значення вона набуває тепер, коли розвиваються нові засоби комунікації (Internet), машинний переклад, системи інформаційного пошуку (індексація за ключовими словами обов'язково передбачає врахування синонімів), створюється новий тип лексичних ресурсів – електронні словники. А це означає, що змінилися практичні вимоги до укладання таких словників, постала потреба в операційному визначенні синонімів.

Як зазначає академік Ю. Д. Апресян [1: 216], проблемі синонімії присвячували й присвячують чимало уваги в лінгвістичних дослідженнях. Однак важливі питання, пов'язані з сутністю цього явища, однозначно все ще не розв'язані. Так, на думку зарубіжних та українських провідних фахівців із питань синонімії, чіткості в проблемі про принципи та критерії синонімічності поки що немає, наявні класифікації синонімів досить умовні, а саме поняття „синонім” надто розпливчате і практично синонімічні відношення між словами під час укладання словників встановлюють інтуїтивно, без урахування чітко випрацюваних наукових положень [1; 2; 3; 9; 10].

Мовознавець С. Г. Бережан взагалі зазначав, що „в сучасній лінгвістиці немає загальноприйнятих основ синонімічності” [3: 35].

Питання теорії синонімії є принциповими для побудови Електронного словника синонімів, ми спробуємо їх розв'язати в пропонуваному дослідженні.

Серед наявних в українському та зарубіжному мовознавстві визначень поняття „синонім” основним критерієм, за яким встановлюють синонімічність, є семантична близькість слів, яку різні науковці інтерпретують по-різному: 1) визначення, в основі яких лежить значення слова (наприклад, А. Б. Шапіро, Р. О. Будагов визначають синоніми як слова, що відрізняються звуковим складом, але близькі за значенням [9: 124]); 2) визначення, в основі яких лежить явище дійсності або поняття, що називається словом (у такому разі синоніми визначають як слова, що означають одне явище дійсності або поняття, але відрізняються відтінками значення, стилістичною належністю, емоційним забарвленням [8: 31; 10 :67]).

Окрім названого вище критерію семантичної подібності, дослідники вводять критерій однакової лексичної сполучуваності. Зокрема, А. П. Євгенєва вважає, що „синоніми – це слова з тотожним або максимально близьким значенням, що можуть бути протиставлені за тонким відтінком у значенні, за вираженням експресії, за емоційним забарвленням, за стилістичною належністю та за сполучуваністю” [7: 11].

Відомий інший підхід до визначення синонімічності слів, а саме через взаємозамінність їх у тексті. У такому разі синонімами вважають тотожні за значенням слова, взаємозамінні в певних контекстах („у тому чи іншому контексті”) без відчуття помітних значеннєвих відмінностей [4: 32; 14: 77; 18: 20; 5: 98]. Через взаємозамінність дає визначення синонімів і Л. О. Новиков [12]. Водночас він чітко ставить проблему співвідношення багатозначності та синонімії. Враховуючи властивість слів бути багатозначними, дослідник вказує на потребу розглядати слова в їхніх конкретних значеннях: „Два (чи більше) ЛСВ (лексико-семантичні варіанти) є синонімами, якщо вони, маючи різні знаки (лексеми), здатні завдяки тотожності або схожості значень заступати одне одного у всіх або певних контекстах, не змінюючи змісту висловлення” [12: 228].

Ю. Д. Апресян до основних ознак синонімів, крім „однакової предметно-поняттєвої віднесеності” та „приблизно однакової сполучуваності”, відносить ще й належність слів до тієї самої частини мови [1: 223]. Отже, синонімами можна вважати слова, які: 1) мають однакове чи максимально наближене тлумачення (однакову предметно-поняттєву віднесеність); 2) є взаємозамінними в тексті, що не призводить до істотної зміни змісту цього тексту (мають подібну лексичну сполучуваність); 3) належать до тієї самої частини мови.

З усіх названих критеріїв синонімічності слів найважче формально описати перший, тому в нашому дослідженні завдання операційного визначення критерію схожості синонімів є одним з основних. Принципово пропонуємо розв'язувати його через встановлення схожості тлумачень синонімів, спираючись на теорію семантичних станів мовних одиниць [19]. Детально про це мова йтиме далі.

Установлення семантичної схожості синонімів передбачає розгляд такого важливого питання, як співвідношення багатозначності та синонімії. Ця проблема не може залишатися поза увагою й укладачів синонімічних словників, і дослідників синонімії як мовного явища взагалі. Чимало лінгвістів [3; 9; 10; 12; 20] вказують на необхідність принципу встановлення синонімічних відношень на рівні конкретних значень (лексико-семантичних варіантів (ЛСВ)) багатозначного слова.

Ми також керуємося цим принципом у визначенні синонімічних груп у процесі укладання Електронного словника синонімів української мови, який також виступає у функції основної лексикографічної бази, що використовується для формування української зони автоматичного багатомовного перекладного словника (АБПС) як частини інтегрованої лексикографічної системи, що розробляється в Українському мовно-інформаційному фонді [6; 17]. Основне завдання перекладного словника полягає в тому, щоб точно передати значення слів однієї мови засобами іншої. В автоматичному багатомовному перекладному словнику цього можна досягти шляхом супроводження перекладних еквівалентів словника відповідними їм синонімами в мовах-учасниках перекладу, а також через мітки щодо належності тексту, який перекладають, до конкретної підмови, сфери вживання слова у своєму конкретному значенні, через систему міток щодо лексико-граматичних властивостей перекладних одиниць: для дієслів – це вид, стан, число, рід, схеми керування; для іменників – рід, число, відмінок; для ад'єктивів – число, відмінок.

Крім того, інтегрований у загальну лексикографічну систему, ЕСС є одним з інструментів семантичного маркування Українського національного лінгвістичного корпусу текстів у встановленні синонімічних відношень між текстовими одиницями. Саме орієнтація на розв'язання цих завдань зумовила „строгий” підхід до визначення параметрів робочого процедурного визначення основних одиниць словника – синонімів та синонімічних груп.

Експліцитне представлення конкретного значення, за яким слова об'єднуються в синонімічні групи, виявляється важливим у семантичному маркуванні Корпусу текстів за умови пошуку синонімів. Програма розмітки тексту на основі існуючого Електронного словника синонімів української мови, яку створено в Українському мовно-інформаційному фонді, звісно, є надзвичайно корисною, такою, що забезпечила принципову можливість

різноманітних досліджень явища синонімії на справді репрезентативному текстовому матеріалі. Проте відсутність індексування синонімів за значеннями дає лише загальне уявлення про потенційні можливості однозначного слова або лексико-семантичних варіантів (ЛСВ) багатозначних слів вступати в синонімічні відношення з ЛСВ інших слів.

Це є принципово важливим, коли маємо справу з багатозначними словами, особливо якщо їх кілька в одній синонімічній групі. Адже коли в тексті синоніми багатозначних слів не проіндексовані за номерами відповідних їм синонімічних груп, то користувачеві буде важко визначитися, за яким конкретно значенням виділено кожний із відзначених у тексті синонімів. Так, наприклад, під час маркування синонімів для слова *підводити* в Корпусі українських текстів обсягом 42 млн. слововживань за ЕСС такими було визначено *підводити, зводити, піднімати, підбурювати, підраховувати, спрямовувати, фарбувати, ставити, підсумовувати, підбивати*. Із уведенням індексів номерів синонімічних груп – аналогів значень, за яким це слово вступає в синонімічні відношення з іншими словами, експліцитною буде інформація, що маємо справу з входженням слова *підводити* в синонімічні групи 5, 7, 10, 100, 105, 300, 350, 400, 500, 700. Аналогічно буде представлена й решта з виділених синонімів: *піднімати* (5, 7, ...); *ставити* (7...); *фарбувати* (300...); *підсумовувати* (10...); *підбурювати* (105...); *зводити* (5, 6, ...); *спрямовувати* (100...); *підбивати* (10...); *підраховувати* (10...). Відповідні значення індексів: **5 – споруджувати, зводити будівлю; 7 – переміщувати знизу вгору; 10 – рахуючи, визначати кількість чогось; 300 – підновлювати забарвлення; 105 – спонукати когось до певних дій, вчинків; 100 – надавати якогось напряму діяльності** і т. ін. Тоді за індексами, що збігаються у слів визначеної множини, й будуть визначатися конкретні ЛСВ слів, які вступають у синонімічні відношення. Так, у наведеному прикладі індекс „5” у слів *підводити, піднімати, зводити* вказує на те, що вони визначені як синоніми за значенням „споруджувати, зводити будівлю”, індекс „10” – *підсумовувати, підбивати, підраховувати* – за значенням „рахуючи, визначати кількість чогось” і т. ін.

Зроблене в такий спосіб маркування синонімів у Корпусі збільшує інформаційні можливості як Корпусу, так і ЕСС. Наприклад, стає можливим формування контекстів синонімів за визначуваним індексом значення у дослідженні контекстних синонімів, аналізі синонімів за параметром взаємозамінності їх у тексті, у визначенні для багатозначного слова співвідношення між значеннями та відтінками значень слів-синонімів.

Названі напрями використання інформації про синонімічні відношення в системах машинного перекладу, на етапі автоматичного семантичного маркування Корпусу, в лінгвістичних дослідженнях семантики багатозначних

слів ми враховували у створенні Електронного словника синонімів української мови.

Базову основу укладання ЕСС становить Словник синонімів української мови (ССУМ) [15]. В електронному представленні ССУМ – це словникові статті, що являють собою синонімічні ряди, у яких слова в межах одного ряду мають схоже значення та об'єднані спільною доміантою. Виділені синонімічні ряди стають об'єктом аналізу щодо семантичної схожості їхніх елементів.

Як було зазначено, практичне призначення ЕСС потребує особливої уваги в питанні основних параметрів, за якими слова об'єднуються в синонімічні ряди, а саме: семантичної схожості синонімів та їхньої взаємозамінності в тексті. Орієнтація на розв'язання означених прикладних завдань зумовила певне переосмислення засад, на яких формувався базовий Словник синонімів української мови. Так, зокрема, вимога взаємозамінності спричиняє вилучення із синонімічних рядів тих лексем, що перебувають у родо-видових відношеннях. Потрібно зазначити, що такі слова в ССУМ нерідко подані в одній словниковій статті через крапку з комою (;). Наприклад, синонімічна група іменника базового Словника синонімів: **чоботи** (взуття з високими халявами); **сап'янці** (з сап'яну); **піми** (хутрянні); **півчобітки, полуботки** (з коротшими халявками). Такі синонімічні ряди будуть перебудовані або взагалі вилучені зі словника, якщо вони складаються лише з двох елементів.

Процедурне визначення семантичної подібності слів у межах синонімічних рядів пропонуємо здійснювати через зіставлення їх тлумачень, поданих у Словнику української мови (ТЛС) [16]. Проте відразу зазначимо, що практично сформувати синонімічні ряди досить непросто, зважаючи на відмінності між семантикою тлумачних словників і семантикою базового ССУМ.

Зіставлення тлумачень елементів синонімічних рядів відбувається за допомогою створеного в Українському мовно-інформаційному фонді автоматизованого інструментального комплексу, що являє собою лексикографічну систему, яка реалізує інтеграцію електронних версій ТЛС і ССУМ. Завдяки цій інтеграції дослідник в автоматизованому режимі отримує змогу переглядати в ТЛС тлумачення та ілюстративний матеріал аналізованих слів, водночас записуючи їх у відповідне поле словникової статті Електронного словника синонімів для їхнього подальшого аналізу. Враховуючи те, що слова української мови є переважно багатозначними, розуміємо труднощі встановлення потрібного значення (адже кількість значень і відтінків лише одного слова може іноді сягати до 30 – 50 варіантів). Тому особливої актуальності набуває питання встановлення критеріїв семантичної подібності слів за тлумаченням у межах синонімічного ряду. Формальними критеріями семантичної схожості синонімів, що визначається з

тлумачних статей, можуть бути: 1) тлумачення значення слова через мітку „те саме, що”; 2) через синонім. Наприклад:

- 1 хвалько** розм. Той, хто любить хвалити себе, хизуватися, хвастати чим-небудь; **хвастун**
- 2 хвастун** підсил. розм. Хвастовита людина; **хвалько**
- 3 хваст** підсил. розм. Те саме, що хвастун
- 4 вихваліяка** підсил. розм. Той або та, що вихваляється
- 5 самохвал** розм. Той, хто сам себе вихваляє; **хвалько**
- 6 самохвалець** розм. Те саме, що самохва́л

За такої тотожності значень слів не виникає проблем у визначенні належності їх до одного синонімічного ряду. Однак варто зазначити, що таких синонімічних рядів небагато, більше тих, для яких дуже важко встановити синонімічність їхніх компонентів. Формальних критеріїв для встановлення синонімічності слів таких рядів недостатньо, тому ми пропонуємо використати новий підхід до розв'язання цієї проблеми.

Установлення подібності значень елементів синонімічного ряду рекомендуємо здійснювати за допомогою методу компонентного аналізу, тобто на основі виявлення спільних семантичних компонентів у тлумаченнях, залучаючи теорію семантичних станів мовних одиниць, розроблену В. А. Широковим [19]. Згідно з цією теорією, будь-яке слово (мовна одиниця) у контексті або в мовному потоці перебуває в певному семантичному стані, який для одиниць лексичного рівня є сумою ознак граматичної та лексичної семантики, тобто „лексемі x та y перебувають у відношенні синонімії $S: xSy$, якщо існують семантичні стани, а саме, стани лексичної семантики – $c(x)$ та $c(y)$, що відповідають певним лексичним значенням розглядуваних лексем, які є близькими для всіх граматичних значень цих лексем: $c(x) \approx c(y)$ ” [19:53]. Визначення семантичного стану слова здійснюється через тлумачення словникової статті, тобто за схожістю семантики слова у відповідних тлумаченнях. Для цього потрібно структурувати тлумачення за компонентами семантичного стану, до яких належать: 1) семантична тема; 2) диференційні семи. Потрібно зазначити, що диференційні семи є специфічними для окремої частини мови. Наприклад, для дієслів це може бути суб'єкт, об'єкт, адресат, спосіб, характер, час, мета, причина та ін., для іменників – належність до певного семантичного класу – істота / неістота, пристрій, деталь, абстрактне поняття та ін., для прикметників – інтелектуальна ознака, ступінь вияву ознаки тощо. Схожість тлумачень встановлюємо за тотожністю лексичного представлення цих семантичних компонентів. Розглянемо для прикладу синонімічний ряд із дієслівною групою синонімів:

1 укладати – *складати, упорядковувати (словник, збірник і т. ін.), створювати що-небудь певним чином;*

2 упорядковувати – систематизувати, складати, підбирати в певному порядку (який-небудь матеріал);

3 складати – відбираючи і розташовуючи певним чином матеріал, створювати що-небудь ціле (про збірник, словник і т. ін.).

Тлумачення цих слів мають спільні компоненти: 1) семантичну тему – створювати, 2) диференційні семи: а) об'єкт – словник, збірник, який-небудь матеріал; б) характер дії – певним чином; у певному порядку. Тотожність виведеної формули семантичного стану домінанти з іншими елементами цієї групи та останніх між собою встановлюємо за збігом (тотожністю) лексем, відповідних виділених семантичних компонентів. На основі подібності семантичного стану ці слова об'єднуємо в один синонімічний ряд.

Операційним визначенням параметра „взаємозамінність” є встановлення подібної лексичної сполучуваності претендентів на роль синонімів у тексті. Системний опис цього параметра забезпечує сучасний стан корпусної лінгвістики, тобто можливість здійснити аналіз мовних явищ (зокрема, встановлення лексичної сполучуваності) на репрезентативних корпусах текстів.

Зазначимо, що спроби встановити синонімічність слів на підставі аналізу фактів сполучуваності (або взаємозамінності) слів у текстах робили й раніше. Наприклад, ще наприкінці ХХ ст. А. Є. Супрун запропонував один із різновидів дистрибутивно-статистичного аналізу слів. Дослідник указував, що під час сполучуваності реалізуються семантичні властивості слів, тому при збільшенні подібності сполучуваності заданих слів із іншими словами збільшується ступінь їх синонімічності [13: 196]. Але зрозуміло, що цю роботу проводили як експеримент, „вручну”. Сьогодні наявність Корпусу текстів (як інформаційної системи з автоматизованим пошуком за визначуваними користувачем певними параметрами) дає лінгвістам також можливість системно перевірити критерій взаємозамінності слів.

Перевірку синонімів за визначальним параметром взаємозамінності їх у тексті в нашій роботі виконано на основі аналізу їхнього контекстного оточення в текстах Українського національного лінгвістичного корпусу обсягом 42 млн. слововживань. Для виділення контекстів використовуємо компонент програмного інструментарію семантичної розмітки текстів, розроблений в Українському мовно-інформаційному фонді для здійснення автоматичного маркування синонімів. За його допомогою на Корпусі формуються конкорданси синонімів завдовжки в одне речення. На виділених контекстах визначаємо списки семантичної сполучуваності кожного з претендентів-синонімів. Суттєва кількісна відмінність цих списків або наявність в них лише 1–2 спільних елементів слугують підставою для перебудови відповідних синонімічних рядів. Прикладом є вилучення із ряду {виконувати, нести – „проводити в життя певне завдання, доручення, наказ,

задум”} слова *нести*, яке траплялося 43 рази з цим значенням лише в сполученні з лексемою *служба* (*несуть військову службу, несуть службу в прикордонних військах та ін.*), тоді як у *виконувати* відповідний ряд, крім слова *служба*, містить 19 лексем і, крім того, слово *служба* зустрілося лише 1 раз. Можна стверджувати, що значення лексеми *виконувати*: „нести відповідальність”, „нести військову службу” в системі семантики лексеми *нести* взагалі фразеологізується.

Також сумнівними видаються нам синоніми вихідного словника, значення яких у тлумаченні розкриваються через лексичну сполучуваність на зразок *глухий* (*про ліс, сад* тощо), *франтівський* (*про одяг, вигляд* і т. ін.). У процесі перевірки на взаємозамінність таких слів-претендентів у текстах синонімічні ряди здебільшого руйнуються. Так, наприклад, можна сказати про *ліс* – *глухий, дикий, незайманий*, але не *здичавілий, здичілий* (у значенні „позбавлений догляду”), ці два синоніми швидше співвідносяться зі словом *сад*.

У наш час електронні словники є одними із найпоширеніших типів ресурсів, які використовують у сфері інформаційних технологій. Електронний словник синонімів української мови поєднує риси довідникової системи та інструмента для проведення лінгвістичних досліджень. Особливість цього словника потребує ретельного підходу до встановлення основних параметрів синонімічності слів та відкриває широкі можливості його використання.

Література

1. Апресян Ю. Д. Лексическая семантика. Синонимические средства языка. – М., 1974. – 367 с.
2. Апресян Ю. Д. Новый объяснительный словарь синонимов. – М., 1995. – 560 с.
3. Бережан С. Г. Семантическая эквивалентность лексических единиц. – Кишенёв, 1973. – 371 с.
4. Булаховський Л. А. Нариси з загального мовознавства. – К.: Рад. школа, 1955. – 307 с.
5. Горнунг Б. В. О природе синонимии в языке и теоретических предпосылках составления синонимических словарей // Вопросы языкознания. – 1965. – №5. – С. 95 – 99.
6. Грязнухіна Т. О. Система багатомовного машинного перекладу // Мовознавство. – 2001. – №5. – С. 14 – 25.
7. Евгеньева А. П. Введение // Словарь синонимов русского языка. – М. – Л., 1970. – С. 5 – 19.
8. Жовтобрюх М. А., Кулик Б. М. Курс сучасної української літературної мови. – К., 1965. – Ч. 1.
9. Звягинцев В. А. Теоретическая и прикладная лингвистика. – М., 1967. – 337 с.
10. Лисиченко Л. А. Лексикология сучасної української мови: Семантична структура слова. – Х., 1977. – 116 с.
11. Нечитайло О. І. Синоніми в лексикографії. – К., 1987. – 132 с.
12. Новиков Л. О. Семантика русского языка. – М., 1982. – 272 с.

13. Общее языкознание / Под ред. А. Е. Супруна. – Минск, 1983. – 456 с.
14. Русанівський В. М. Структура лексичної і граматичної семантики. – К.: Наукова думка, 1972. – 243 с.
15. Словник синонімів української мови: В 2 т.– К., 1999–2000.
16. Словник української мови: В 11 т.– К., 1970 – 1980.
17. Устимець О. В. Формування електронного словника синонімів української мови // Учёные записки Таврического национального университета им. В. И. Вернадского. Серия „Филология”. Том 20 (59).– Симферополь, 2007. – № 4.– С.57 – 61.
18. Фаворин В. К. Синонимы в русском языке. – Свердловск, 1953. – 71 с.
19. Широков В. А. Семантичні стани мовних одиниць та їх застосування в когнітивній лексикографії // Мовознавство. – 2005. – №3 – 4.– С. 47 – 62.
20. Шмелёв Д. Н. Проблемы семантического анализа лексики. – М.: Просвещение, 1973. – 280 с.