

Глибина кругових визначень у словнику як віддзеркалення властивості лексичної системи

Двомовні перекладні словники заледве розв'язують невелику частину проблем перекладу, перекладачам у професійній діяльності доводиться частіше звертатися до тлумачних словників. Немає потреби спеціально доводити, що пояснення лексичних одиниць у тлумачних словниках замикаються на собі самих, позаяк слова природної мови у словнику тлумачаться теж словами. Тим не менш, тлумачні словники існують тисячоліттями, і користувачам вдається успішно ними послуговуватися. Парадокс цей можна розв'язати, якщо взяти до уваги, що базові поняття рідної мови людина первинно засвоює все ж не шляхом логічних пояснень та аналітичних операцій, а безпосередньо через органи відчуття та почуття у комунікативних ситуаціях. І хоча вважається неприйнятним кругове визначення першого та другого порядку, на кшталт: “млосний” – “той, що викликає млость”, а “млость – іменник від млосний”, кругові пояснення вищих порядків у тлумачних словниках загальної лексики неминучі.

Порядком, глибиною або ступенем кругової дефініції лексикографі називають кількість переходів від однієї словникової статті до іншої, що можна здійснити (в теорії або на практиці) до повтору вихідної лексичної одиниці.

Ось, приміром, який шлях очікує читача Словника Іспанської королівської академії під час знайомства зі вкрай складним для визначення поняття “смысл” (*sentido*):

Sentido: m. *Interpretación de que puede ser objeto un mensaje o una obra. Una novela con múltiples sentidos.*

Mensaje: m. *Trasfondo o sentido profundo transmitido por una obra intelectual o artística.*

Interpretar: 1. tr. *Explicar o declarar el sentido de algo, y principalmente el de un texto* [1].

Можна припустити, що з погляду пересічного користувача, який читатиме дефініцію поняття, необхідна базова умова для розуміння – зрозумілість для нього всіх лексем, які використовуються в поясненні. Доволі рідко виникають ситуації, коли доводиться додатково шукати пояснення слів з самої дефініції, та й мало хто так робить, оскільки це доволі складна розумова задача. Тому не буде перебільшенням вважати, що далі другого-третього порядків пошук пояснень не йде. Відтак, кругові визначення третього порядку і вище можна умовно вважати прийнятними. У такий спосіб лексику можна образно порівняти з величезним багатогранником, виробленим зі зчеплених між собою картонок, у якому кожна грань кріпиться до сусідніх; така конструкція може існувати лише за рахунок зв'язку усієї сукупності, жодна частина не зможе закріпитися, утворити систему або конструкцію самостійно. У такий спосіб яскраво видно, що поняття існує в системі не стільки завдяки поясненню, скільки завдяки тій ніші, які йому залишають суміжні поняття. Подібні спостереження яскраво описує видатний іспанський філософ Х.Ортега-і-Гасет: наш профіль значною мірою залежить від того простору, що його для нас відводять інші особи [2, с. 490]. Ось чому образно можна сказати, що немає у лексики якогось початку або кінця чи навіть основних сем, на яких би трималися складі поняття, так само, як у Всесвіті немає “низу”, “верху” й тих самих “трьох китів”, наразі очевидна істина, до якої людство йшло тисячоліттями.

Упорядники словників цілком свідомі того, що дефініції лексичних одиниць замкнені самі на собі, і найпростіше, чим можуть зарадити користувачеві – це звузити кількість лексем, що використовуються в дефініціях, до списку, за яким в англійській лексикографічній традиції закріпилася назва “defining vocabulary”, термін, що в українській лексикографії часто-густо подається без перекладу; у працях російською мовою вряди-годи з'являється варіант перекладу “дефиниционный словарь”, а вже в зовсім одиничних випадках і український переклад – “дефініційний словник” або “визначальний словник”[3].

Описуючи дефініційний словник, зокрема, редколегія словника *The Longman American Defining Vocabulary* зазначає, що “слова для дефініцій обиралися виключно з заданого списку, а якщо в дефініції й міститься якесь стороннє слово, воно позначалося великими заголовними літерами. Дефініційний словник для тлумачень ретельно відібрано після вивчення таблиці частоти вживання англійських слів. Більше того, слова в дефініціях використовуються в їхньому основному значенні” [4]. У той самий час, Хай Ксю застерігає, що підхід, заснований на використанні найчастіших слів у дефініціях, не позбавлений деяких недоліків через те, що лексикографу для пояснення значення лексем також потрібні узагальнюючі поняття високого рівня абстракції, граматичні терміни. Тому вже в п'ятій версії *Longman Defining Dictionary* список слів, хоча й коротший на 154 одиниці в порівнянні з першою версією і становить 2094 одиниці, він покриває більшу кількість лексики в корпусі текстів, ніж попередній, здавалося б, масштабніший дефініційний словник [5]. І справді, такі слова як “суть”, “властивість”, “специфіка” не є найчастішими у пересічних мовців, однак безсумнівно часто зустрічаються в дефініціях понять. При всьому тому, дослідник вказує на недбалості в використанні як заявленого списку, так і значень, в яких використовуються дефініційні слова, а також частиномовної належності слів; поза тим, окремим проблемним питанням залишається використання в дефініціях словосполучень, що можуть мати самостійне значення, відмінне від значень слів, що входять до його складу [5]. Дійсно, значення сполучення слів на зразок “мати рацію” не так вже й легко вивести зі значень слів, що до їхнього складу входять, тому визначення рамок дефініційного словника виявляється значно складнішим, ніж видається на перший погляд.

Обсяг дефініційного словника теж поки що лишається питанням відкритим. А.П. Кові зауважує, що нічого магічного у числі 2000, адже на практиці кількість використаних для визначень слів суттєво занижується, і насправді потрібно значно більше двох тисяч [6, с. 110].

Мусимо наголосити, що описана проблема стосується словників загальної лексики. Кардинально інакше слід підходити до визначення термінів у спеціалізованих галузях: у них кругові визначення не тільки небажані, а й неприпустимі, адже їхній користувач жодного разу не повинний потрапити у лабіринт замкненого кола перехресних посилань. Крім того, Л. Агілар відзначає, що дефініції в термінологічних словниках мають ґрунтуватися на термінах, у визначеннях яких використовуються ті самі поняття; має уникатися використання синонімів для тлумачення термінів [7, с. 21]. Е.Ф. Скороходько у свою чергу зауважує, що дефініції термінів мають будуватися на тих термінах, що вже визначені в словниках [8, с. 96]. Звичайно, ці думки, цілком слушні, ґрунтуються на тому факті, що користувач послуговується словником, коли вже володіє базовими поняттями мови, які для нього дефініцій не потребують. Відтак, у словнику загальної лексики принаймні частина слів неминуче буде в колі кругових дефініцій.

Попри очевидну практичну значущість та захопливість проблематики, ще не стало спеціальним предметом вивчення питання, якого порядку кругові визначення переважають у загальнолексичних тлумачних словниках, а також не встановлено, для якого лексичного мінімуму неможливо побудувати пояснення без кругових визначень. Цілком можливо, що за рахунок поступової субституції лексем тими визначеннями, на яких вони ґрунтуються, можна суттєво скоротити кількість мінімально потрібних для тлумачень лексичних одиниць, що призведе до громіздкості дефініцій, важко зрозумілих для людини, однак вичерпних для машини. Якщо, як зазначено, у словнику Лонгман дефініційний словник нараховує приблизно 2000 одиниць, чому цих базових слів не може бути менше для використання в галузі штучного інтелекту? І навпаки, як ми можемо упевнитися, що їх достатньо для чітких і наочних пояснень, що потрібні пересічному користувачеві? Це далеко не вичерпний перелік питань, на яких виводить дослідника вивчення характеру та суті кругових дефініцій. Наразі завдання цього дослідження полягають у відповіді на такі питання, яким є в середньому ступінь кругових визначень для слів і якою є їхня роль в інтерпретації властивостей лексики.

Матеріалом дослідження обрано електронний тлумачний словник англійської мови обсягом 110 674 словникових статей, приведений до формату, зручному для опрацювання в базах даних, що знаходиться у відкритому доступі на ресурсі *Gutenberg Project*, заснований на виданні *US Webster's Unabridged Dictionary 1913* року [9].

Словник опрацьовувався програмою, яку з огляду на інтерес до проекту люб'язно зголосився укласти програміст М.С. Іванов мовою *Java*. Для спрощення проведення експерименту й уникнення зайвого негативного матеріалу під час обробки словника програмою не приймалися до уваги лексеми, довжина яких коротша за 4 символи; не приймалися до уваги псевдокругові визначення першого порядку, оскільки в дефініціях могли міститися однокореневі слова або навіть слова з леми в ілюстративних прикладах; для досягнення точнішого результату від слів у дефініціях відкидалися флексії або псевдофлексії -ing, -ed, -s.

Ми свідомі того, що, як і будь-яка формалізована обробка, позбавлена людського смислового аналізу, запропонована тут схема має низку недоліків: під визначення однакових слів потраплятимуть омоніми і омографи ("потяг до культури" і "потяг до Відня"; "дорóга" і "дорогá" тощо), і це призводить до певної похибки. Втім, для первинної розвідки з метою віднаходження проблемних аспектів, дискусійних завдань, практичних труднощів та неочікуваних парадоксів зазначених умов цілком достатньо.

Результати перших експериментів дозволяють зробити певні висновки, а також окреслити перспективи нововиявлених проблемних питань:

1. У словнику програма нараховувала 110674 лем; після відкидання флексій і псевдофлексій – 100728; у дефініціях у цілому використовується 71125 словоформ. Відтак, кількість словоформ у дефініціях менша, ніж кількість лем. Це цілком логічно, з огляду на вимоги щодо стислості дефініційного словника, однак значно перевищує прийняту на сьогодні міру порядку 2000 лексем.

2. Наразі не вдалося встановити максимальну глибину кругових дефініцій. Оскільки кількість оброблюваних слів зростає у геометричній прогресії, кількість здійснюваних машиною порівняльних операцій становитиме X^n , де X – середній обсяг дефініції, а n – порядок пошуку кругової дефініції. Більше того, оскільки теоретично лабіринт перехресних посилань від однієї статті до іншої може задіяти всі абсолютно статті, теоретично можлива глибина дорівнює

$$\sum_{i=1}^{vol} = X_1 + \dots + X_{vol}$$

де vol – обсяг словника; у такому випадку можна приблизно поррахувати, що кількість здійснюваних для цього операцій може дорівнювати числу, що складається зі ста тисяч цифр. Якщо виходити з того, що сучасні комп'ютери наближуються до швидкості 10^{18} операцій в секунду, то для проведення 10^{100000} операцій знадобиться 10^{99982} секунд, що приблизно становить $3 \cdot 10^{99974}$ років (йдеться про час розрахунку тільки останнього члена у цьому багаточлені), в той час як вік Всесвіту за нинішніми розрахунками $3 \cdot 10^{10}$ років. Однак цей проміжний умовляний висновок заслуговує на окрему детальнішу розвідку, як у теоретичному, так і практичному планах. У теоретичному аспекті необхідно показати у термінах теорії множин або формальної

логіки, чи можлива глибина кругової дефініції більша, ніж кількість лем у словнику; на практиці, якщо це завдання взагалі зрима, необхідно за допомогою комп'ютерних програм віднайти максимально глибокі дефініції у тлумачних словниках різних мов. Для уточнення розв'язання цієї задачі розраховано середню довжину дефініцій у словниках, яка становить у середньому приблизно 11,1 слів.

3. Кількість величин глибини кругової дефініції розподілилася таким чином:

Глибина дефініції	Кількість випадків
1	Не приймалися до уваги
2	17 116
3	3 641
4	3 072
5	2 334
6	1 164
7	499
8	234
9	77
10	36
11	10
12	6
13	2
14	0
15	0
16 та більше	0

Як видно з таблиці, глибше дванадцятого порядку кругові дефініції практично не зустрічаються. Більшість кругових дефініцій у словнику мають другий порядок, що, на перший погляд може викликати подив і обурення: як у принципі можна користуватися словником, якщо вже на другому порядку користувач може потрапити в замкнене коло? Але це не суперечить можливості розуміти тлумачення лексем. По-перше, в основі створення, засвоєння та розуміння багатьох понять лежить бінарність, і “день та ніч”, “добро і зло”, “сум і радість” тощо пізнаються тільки в сукупності, а не послідовно по черзі. Поза тим, дефініції другого порядку часто виникають через те, що дефініції надто деталізовані, і після пояснення значення надається також список синонімів лем, для точніших результатів ці додаткові пояснення в ідеалі необхідно видалити. У тлумаченні лексеми “accuracy” у списку синонімів міститься “precision”, а в тлумаченні “precision” міститься “accuracy”. Це похибки формальної обробки, що призводять до пошуку певної кількості псевдокругових дефініцій. У той самий час нерозуміння одного слова в масштабній дефініції ще не означає нерозуміння читачем тлумачення. Людській мові властива інформативна надмірність, за підрахунками Р.Г. Піотровського, становить 70-80% [10, с. 58].

Окрім впорядкування бази словника для отримання та належної інтерпретації точніших результатів перспективними найближчими кроками є порівняння з іншими тлумачними словниками: 1) англійської мови інших років та видань, зокрема також тих, в яких дефініції побудовано на обмеженому дефініційному словнику; 2) словниками інших мов, які вже переведено і зручний для комп'ютерної обробки формат.

Література

1. Diccionario de la Real Academia Española. – [Електронний ресурс]. Режим доступу: www.rae.es
2. Ortega y Gasset J. Sobre el fascismo / J. Ortega y Gasset // Obras completas. Tomo III. – Madrid: Alianza editorial, Revista de occidente, 1983. – P. 489-497.
3. Пшенична Л., Шевченко В, Шишкіна Н. Термінологічна культура фахівця – освітньо-навчальний проєкт / Л. Пшенична, В. Шевченко, Н. Шишкіна. – [Електронний ресурс]. Режим доступу: <http://ena.lp.edu.ua:8080/bitstream/ntb/15263/1/8-34-41.pdf>
4. The Longman American Defining Dictionary. – [Електронний ресурс]. Режим доступу: http://www.longmandictionariesusa.com/res/shared/vocab_definitions.pdf
5. Hai Xu. A Critique of the Controlled Defining Vocabulary in Longman Dictionary of Contemporary English / H. Xu. – Centre for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou, China. – [Електронний ресурс]. Режим доступу: <http://lexikos.journals.ac.za/pub/article/view/1013/530>

6. Cowie A.P. English dictionary for foreign learners: a history / A.P. Cowie. Oxford: Clarendon Press, 1999. – 235 p.
7. Aguilar L. Lexicografía y terminología aplicadas a la traducción. Curso práctico de introducción / L. Aguilar. – Bellaterra: Servicio de Publicaciones de la Universidad Autónoma de Barcelona, 2001. – 86 p.
8. Скороходько Э.Ф. Семантические сети и автоматическая обработка текста / Э.Ф. Скороходько. – К.: Наукова думка, 1983. – 128 с.
9. Gutenberg Etext of Webster's Unabridged Dictionary – [Электронный ресурс]. Режим доступа: <http://www.mso.anu.edu.au/~ralph/OPTED/>
10. Пиотровский Р.Г. Информационные измерения языка / Р.Г. Пиотровский. – Ленинград, 1968. – 116с.