

РОЗДІЛ II. ТЕРМІНОГРАФІЯ ТА ЕЛЕКТРОННЕ ВПОРЯДКУВАННЯ ТЕРМІНОЛОГІЧНИХ БАЗ ДАНИХ

Вимоги дедалі динамічнішої комунікації та міжнародної інтеграції, приголомшливі темпи розвитку науки і техніки становлять виклик для термінографів, на плечі яких лягає завдання з упорядкування величезної кількості даних. У той час як термінологія вивчає творення, структуру, функціонування та властивості термінів і терміносистем, завдання термінографії на часі такі:

- розробка і оптимізація форматів термінологічних словників;
- розробка алгоритмів і принципів укладання електронних термінологічних словників і баз даних (як людиною, так і машиною);
- стандартизація форми представлення термінів та їхніх параметрів у лексикографічних джерелах;
- аналіз властивостей термінологічних систем, наявних у словниках.

Основні проблеми, з якими зіштовхуються термінологи на практиці – це синонімічна варіантність термінів, яка вважається вкрай небажаною, пошук перекладних відповідників для термінів-неологізмів, впорядкування термінологічних баз даних, що розглядаються далі в розділі.

2.1. Проблеми термінологічної стандартизації в електронних базах даних

Оскільки терміни слугують для комунікації переважно у науково-технічній галузі, у своїй більшості вони мають інтернаціональний характер, тож стандарти комітету ISO/TC-37, заснованого у 1947, становлять надзвичайну вартість для забезпечення уніфікації, обробки та зберігання термінів. Цей комітет займається питаннями комп'ютерної термінографії, лексикографії і розробляє стандарти: ISO/TC 37/SC1 “Принципи і методи” (“Principles and methods”, англ.), ISO/TC 37/SC2 “Термінографія і лексикографія” (“Terminography and lexicography”), ISO/TC 37/SC3 “Комп'ютерне впорядкування термінології” (“Computer applications for terminology”). Стандарти

комітету ISO/TC-37 щодо роботи термінолога публікуються англійською і французькою мовами, хоча інколи вони перекладаються й на інші мови. За перші 50 років існування цей комітет розробив 40 стандартів, деякі з яких, особливо значущі в контексті термінографії, згадаємо нижче.

Уже під час існування комітету ISO/TC-37 у рамках ЮНЕСКО виникла організація ІНФОТЕРМ (1971), що стала найавторитетнішою в світі установою у справах термінології. На базі ІНФОТЕРМ згодом засновано програму ЮНІСІСТ, яка на сьогодні надає термінологічні послуги на міжнародному рівні. Задля збільшення ефективності та поширення своєї діяльності в світі ІНФОТЕРМ підтримує тісний зв'язок з комітетом ISO/TC-37.

Міжнародна організація зі стандартизації (ISO) тримає зв'язок з більш, ніж 1600 організаціями в світі, серед яких ООН (ВООЗ, МАГАТЕ, ФАО, ЮНІДО, ЮНЕСКО), МВФ, МОЗМ, Агенція з ядерної енергії при Організації економічної співпраці й розвитку, Міжнародна асоціація перекладачів конференцій (AIIIC), Міжнародна торгова палата. Серед документів, виданих цією установою, вкрай важливим для термінографії є глосарій Інкотермс (від англійської – “International commerce terms”), перша версія якого датується 1936 роком; дефініції глосарія періодично переглядалися, а окремі позиції додавалися або прибиралися, і, зрештою, у 2010 році документ нараховував 11 термінів. Кожен термін описує певні умови міжнародної торгівлі. Оскільки форма терміна схожа на код, що складається з трьох латинських літер, його може розпізнати фахівець з будь-якої країни, не вдаючись до перекладу [Довідник Інкотермс українською. Особливості використання.].

У той час як комітет ISO/TC-37 займається стандартизацією методів і принципів роботи, згідно з даними ІНФОТЕРМ, близько 200 міжнародних комітетів і приблизно 3 000 національних комітетів виконують роботу зі стандартизації у своїх відповідних галузях науки і техніки [Standardization]. В Україні впорядкуванням термінології опікуються такі установи: Український науково-дослідний інститут стандартизації, сертифікації та інформатики (УкрНДІССІ), Академія наук України, галузеві науково-дослідні інститути, вищі навчальні заклади, Технічний комітет стандартизації науково-технічної термінології. В інших країнах термінологічними питаннями займається також декілька державних установ паралельно. До прикладу,

швейцарська база даних "Termdat" послуговується термінологічними розробками восьми державних закладів цієї країни [Termdat].

Незважаючи на численні спроби уніфікувати термінологічні формати, вони все ж доволі часто різняться, оскільки специфіка опрацювання термінологічної інформації може бути не завжди однаковою в різних галузях науки і техніки. Якщо йдеться про галузі, які не спираються на культурно марковане онтологічне підґрунтя, такі як фізика, астрономія, математика, медицина, їхні терміни позначають універсальні поняття. А в культурно маркованих науках, таких як історія, право, політологія та інших багато термінів багато термінів мають специфічне локальне, національне, історичне підґрунтя і, відповідно, позначають неуніверсальні поняття. Ця ситуація породжує низку труднощів у міжнародному спілкуванні, і, зокрема, в Євросоюзі намагаються їх подолати шляхом впровадження уніфікованої для всіх країн-членів термінології [Goffin 1994], що в ідеалі буде позбавлена національно маркованого підґрунтя, так званого "євролекту". Однак під час укладання дефініцій термінів виникають супутні проблеми, адже дефініції містять посилання на інші терміни, які, в свою чергу, можуть тлумачитися по-різному в різних країнах.

Такого роду дискусії спричинили необхідність створення додаткових засобів для обговорення нагальних проблем термінології, суперечностей, пропозицій шляхів вирішення. Зокрема, у періодичних виданнях: "Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication" (англійською мовою) [Terminology. International Journal...], "Punto y Coma" (іспанською мовою), в яких детально висвітлюються проблеми гармонізації, адаптації, нормалізації й інтернаціоналізації наявних і новостворених термінологічних одиниць у рамках загальноєвропейського простору, а також різні труднощі перекладу цих термінів. На матеріалі французької та англійської мов проблеми упорядкування перекладу термінів розглядаються у виданні "Traduction, Terminologie, Rédaction" [Traduction, Terminologie, Rédaction].

З огляду на те, що численні термінологічні проблеми виникають у комунікації навіть між країнами, що розділяють спільний адміністративний, законодавчий та почасти фінансовий простір (приміром, у Євросоюзі), тим паче мають виникати труднощі у комунікації між культурно віддаленими спільнотами. Приміром, у радянській науці виокремлювали терміни "материк"

як поняття геології та “континент”, що є суто географічним поняттям. В англійській, іспанській, французькій та інших мовах Євросоюзу такого розрізнення немає, позаяк немає розрізнення в самих науках, і в нинішніх довідкових джерелах зазначається, що з географічної точки зору ці терміни є синонімами, хоча суперечливість не можна зняти остаточно, оскільки й погляди на кількість континентів у світі різняться: за суто геологічними показниками може бути одна класифікація, однак, відповідно до культурно-історичних та політико-економічних критеріїв – інша. В українському сільському господарстві фахівці послуговуються поняттям “баштанні культури”, яке в більшості країн світу не виокремлюється; деякі словники пропонують такий переклад англійською: “cucurbits crop; vine crop; grounds; melons and grounds” [Словарь Мультитран], хоча поняття “cucurbits” (“гарбузові”) виходить за межі множини динь, гарбузів та кавунів, що її позначає термін “баштанні культури”. Однак у певному сільськогосподарському та навіть комерційному контексті їх зручно розглядати разом: гарбузи, дині і кавуни зріють приблизно в однаковий час, зосереджені переважно в тих самих регіонах тощо, однак не обов’язково так само складається торговельно-економічна та сільськогосподарська кон’юнктура. Між поняттями “баштанні” і “cucurbits/гарбузові” немає відношень еквівалентності, гіпо-/гіперонімії. Однак їх можна розглядати як квазісиноніми, оскільки множини позначуваних ними денотатів частково перетинаються. Чимало випадків подібної асиметрії трапляється між видовими поняттями, які у вихідній мові представлені двома або більше термінами з близькими значеннями, а в цільовій мові мають лише один засіб вираження. Наприклад, терміни “новела” і “роман” – в англійській та іспанській мовах передаються одним іменником “novel/novela”; поняття “сон” і “мрія” – словом “dream/sueño” тощо. Інколи певні визначники в контексті (часто – прикметники) можуть знімати двозначність, однак зазначені терміни живаються і без визначників.

Сказане свідчить, що сучасні терміносистеми різних мов характеризуються суттєвою асиметрією, що може спричиняти і ситуації непорозуміння, і труднощі в перекладі. Це означає, що термінологи мають здійснити чимало зусиль щодо поняттєвої уніфікації, зокрема: гармонізації, інтернаціоналізації, нормалізації та формального впорядкування терміносистем. Розроблені на сьогодні формати, зручні для укладання термінологічних баз даних, словників та модулів комп’ютерної

перекладацької пам'яті, розглядаються у подальших підрозділах.

2.2. Характеристики термінологічних словників

Перша риса, що спадає на думку у зв'язку з термінологічними, (спеціалізованими), словниками є їхня точність, відсутність розбіжностей і непорозумінь. Така вимога, значно менш категорична по відношенню до загальних словників, вимагає певних пояснень. А. Мелбі, Г. Будін, С. Райт стверджують, що відмінності між термінологічними та загальними словниками не обмежуються ступенем точності, а й стосуються також структурних характеристик: об'єктом термінологічних словників є поняття, а не лексеми [Melby 1993, с. 89].

Окрема перевага спеціалізованих словників у тому, що вони зосереджені на певній галузі знань і не містять обтяжливої кількості омонімічних лексем, численних відтінків і нюансів значень з інших спеціалізованих сфер, тож користувачеві значно легше знаходити в них відповідний лексико-семантичний варіант, ніж у словниках загального користування. З іншого боку, ступінь деталізації словникової термінологічної статті може варіюватися. Наприклад, С.В. Гриньов наголошує на важливості 19 параметрів опису термінів, що їх розробив В.Ю. Городецький, які включають етимологію, семантичні зв'язки з іншими термінами, металінгвістичними даними (як, наприклад, історія дослідження денотата) тощо. У свою чергу Ю. Караулов виокремлює 68 таких параметрів, тоді як формат обміну термінологічними даними ФОЛИТ охоплює 168 характеристик [Гринев 1990, с. 25-40], що значно детальніше, ніж більшість сучасних комп'ютерних форматів. Отже, існують серйозні розбіжності в оформленні описів термінів. Представлення термінів відповідно до заданих параметрів з метою їх систематизації і зберігання у базах даних зазвичай називають термінологічними картками. Більше відомостей щодо опису термінів у термінологічних картках можна знайти в підрозділі 2.5.

Хоча двомовні перекладні термінологічні словники організовані за алфавітом, тематичне впорядкування термінології здійснити не так важко, як для загальної лексики в цілому. З огляду на деякі об'єктивні чинники, як, наприклад конвенційність і системність термінології, на відміну від

загальної лексики, термінологія легко структурується за синоптичними ієрархічними схемами. Серед найяскравіших прикладів слід згадати біологічну таксономію, згідно з якою весь світ розподіляється на царства, класи, ряди, родини, сімейства, роди, види. Така чітка структурованість дає змогу упорядникам словника організовувати поняття не лише за алфавітом, а й за відповідними групами, класами, тобто, не лише семасіологічно, а й ономасіологічно. Це наочно ілюструє, що ономасіологічний підхід до опису терміносистеми більш виправданий і доречний, ніж для загальної лексики. Серед прикладів ономасіологічних термінологічних словників слід згадати "Тезаурус з комп'ютерної лексикографії" [Тезаурус з комп'ютерної], упорядкований О.Б. Сірук. На жаль, ономасіологічні термінологічні словники обмежуються поодинокими розробками.

Часто терміни створюються спочатку в одній мові, а потім поширюються і запозичаються іншими мовами. У таких випадках необхідно, щоб термінологічні словники включали також початкову форму мовою оригіналу, в якій вони виникли. Подібний підхід використовується, зокрема, при описі ботанічних та зоологічних видів, наукові назви яких первинно створюються латиною з системою подвійних назв, започаткованою Карлом Ліннеєм.

Цей принцип можна поширювати й на юридичні науки, зокрема, на термінологію, яка походить з римського права. У цьому зв'язку згадаємо багатомовний юридичний словник-довідник за редакцією І.О. Голубовської [Багатомовний юридичний 2012]).

Численні терміни упорядковуються міжнародними установами, тому не дивно, що інколи їм надаються також закодовані назви, що складаються зі сполучень, як правило, латинських літер та арабських цифр. Такий код може виявитися дуже цінним знаряддям, адже він по суті є екстралінгвістичною позначкою термінологічного поняття, яка вільна від проблем омонімії і багатозначності. Цей код слугує позамовним і безпомилковим засобом номінації поняття. Закодовані назви використовуються, до прикладу, у таких терміносистемах: Міжнародна класифікація хвороб ВООЗ або Універсальна десятикова класифікація галузей знань, що, згідно з даними Іспанської агенції з нормалізації (AENOR), нараховує понад 70 000 назв. Оскільки умовний код уже не прив'язаний до певної мови, він може легко оброблятися машиною; він доступний для розуміння

широкому колу фахівців, на відміну від номінацій природною мовою, якої фахівець може й не знати.

Попри відомі вимоги щодо однозначності найменувань термінів та одно-однозначних зв'язків (адже одна форма має асоціюватися з одним поняттям), усе ж, велике число термінів не позбавлене омонімії та синонімії. Приміром, у сучасних філологічних працях широко використовується поняття “дискурс”, яке доволі багатозначне й охоплює такі тлумачення: “1) зв'язний текст у контексті численних супровідних фонових чинників – онтологічних, соціокультурних, психологічних тощо; текст, ‘занурений’ у життя; 2) замкнута цілісна комунікативна ситуація (подія), складниками якої є комуніканти й текст як знаковий посередник, зумовлена різними чинниками, що опосередковують спілкування й розуміння (соціальними, культурними, етнічними і т. ін.); 3) стиль, підмова мовного спілкування; 4) зразок мовленнєвої поведінки в певній соціальній сфері, що має певний набір змінних” [Комунікативна ситуація 2008, с. 567-568]. Інші значення дискурсу трапляються значно рідше і переважно в нетермінологічному вживанні. Як видно, третє значення корелює з терміном “підмова”. Отже, термінологічні словники мають охоплювати й випадки полісемії, синонімії і омонімії, явищ, вкрай небажаних, однак об'єктивно присутніх у терміносистемах.

Таким чином, під час розробки термінологічних словників необхідно враховувати, що їхнім об'єктом виступають не лексеми, а поняття. Тому й ономазіологічний підхід (від поняття до форми) для укладання термінологічних словників виявляється набагато перспективнішим, ніж для загальної лексики, особливо ж, в електронній формі, яка додатково забезпечує широкі можливості формального пошуку як за повними, так і за частковими даними, дозволяє автоматично знаходити лему за дефініцією або ієрархічно класифікованими метамовними дескрипторами тощо. Оскільки об'єктом термінологічного словника є поняття, а не лексеми, суттєвим компонентом тлумачення має стати встановлення зв'язків з іншими поняттями, укладання тезауруса. У термінологічних базах даних, на відміну від звичайних перекладних словників, терміни мають супроводжуватися дефініціями. Існують різні підходи і вимоги до формулювання дефініцій, що докладніше розглядаються у наступному підрозділі.

2.3. Вимоги до дефініцій термінів у спеціалізованих словниках

Методологічно правильне і обґрунтоване впорядкування дефініцій є запорукою успішного використання термінологічних карток у будь-якому електронному ресурсі. Існує два поширених типи дефініцій: дефініції за інтенсіоналом і дефініції за екстенсіоналом. Дефініція за інтенсіоналом зазвичай передбачає використання узагальнюючого терміна, гіпероніма з уточненням вирізняльних сем-характеристик для окреслення поняття на тлі інших понять даного роду, виду, типу. У визначенні специфічних рис поняття використовуються диференційні семи. У цьому зв'язку слід згадати лексикографа й перекладознавця В.Г. Гака, який розподіляє диференційні семи на описові та відносні [Гак 1977, с. 14-15].

Наприклад, дефініцією за інтенсіоналом поняття "сторони світу" можна вважати таку: "кожна з чотирьох частин, на які ділиться горизонт відносно північного полюсу". Натомість дефініція за екстенсіоналом полягає у перерахуванні сукупності денотатів, що підпадають під певну дефініцію. Визначення за екстенсіоналом терміна "сторони світу" матиме такий вигляд: "північ, південь, схід, захід". Саме таке визначення подається у Тлумачному словнику української мови "СУМ" [СУМ]. Відповідно до міжнародного стандарту "ISO-1087", пункт дефініцій в термінологічній картці має включати дефініцію за інтенсіоналом та дефініцію за екстенсіоналом.

Коли йдеться про мало відомі лексичні одиниці, бажано та навіть необхідно, щоб терміни супроводжувалися відповідними дефініціями. Існує низка вимог, яких слід дотримуватися під час формулювання дефініцій: 1) дефініції термінів мають утворювати систему понять; вони мають охоплювати ті характеристики, які є суттєвими для заданої галузі знань; 2) дефініції мають ґрунтуватися на наявних термінах, визначення яких спирається на уніфіковану базу понять; 3) слід уникати використання синонімів у дефініціях; 4) вихідний термін не може входити до складу дефініції (виняток становлять терміни-словосполучення) [Aguilar 2001, с. 21]. Крім того, у дефініціях не повинно бути надмірних характеристик. Наприклад, для визначення поняття "дерево" зайвим було б зазначати, що у нього є "листя та коріння" або що "воно виростає з насіння", оскільки ці

характеристики наявні також у найближчому гіперонімі цього поняття – “рослина” [Troiano J.A., 10]. У словнику Королівської академії іспанської мови пропонується таке визначення: “planta perenne, de tronco leñoso y elevado, que se ramifica a cierta altura del suelo” [DRAE] (“багатолітня рослина з дерев’яним високим стовбуром, що на певній висоті над ґрунтом розділяється на гілки”). Визначення будується на основі гіпероніма – “рослина”, та супроводжується додатковими диференційними семами. За необхідності до дефініції шляхом наслідування можна було б додати також усі ознаки поняття рослини. Цей процес, можливо, позбавлений сенсу для людини, може відігравати важливу роль при автоматичній обробці тексту, зокрема, під час аналізу вторинних номінацій. Один і той самий денотат, скажімо, ‘дерево’, у контексті може бути номінований гіперонімом (“рослина”), гіпонімом (“дуб”), меронімом (“гілля”), і для аналізу й розпізнавання предметного змісту й смислу тексту потрібне врахування широкої кількості ознак, відносних і описових сем, архісем.

У свою чергу Е.Ф. Скороходько додає, що дефініції мають ґрунтуватися на лексемах, які містяться у словнику, інакше станеться помилка, яка називається “перервана дефініція” (користувач словника зіштовхнеться у тлумаченні з лексемою, дефініцію якої не зможе знайти у словнику); одна й та сама дефініція не може повторюватися в різних словникових статтях (інакше різні поняття підпадатимуть під однакову дефініцію); значення слів, використаних у дефініції, не повинні суперечити одне одному [Скороходько 1983, с. 96]. Якщо словник представити у вигляді семантичної сітки, незалежні помилки знайдуть формальне вираження, і їх можна буде виявляти автоматично [Скороходько 1983, с. 96].

Звичайно, лексична система замкнена сама на собі, тому якщо в словнику й не буде перерваних дефініцій, у ньому неминуче з’являться кругові дефініції. Натомість у термінологічних словниках цього можна уникнути, адже терміносистеми є свідомо створеним середовищем, заснованим на консенсусі про те, що базовий термінологічний апарат має ґрунтуватися на лексемах загального словникового складу, а вторинні, похідні термінологічні поняття спиратимуться на базові.

Значна кількість намагань учених використати тлумачні словники у системах штучного інтелекту наштовхувалася на перешкоди саме у зв’язку з недосконалістю словникових

дефініцій. Однак дефініції – дуже потужне джерело для створення цінних комп'ютерних знарядь: приміром, наявність впорядкованих стилістичних та соціолокальних поміток дає змогу генерувати нові словники на основі вже існуючих; окрім того, Т. Фонтанелле вказує, що можливо давати визначення семантичних функцій, виходячи зі структури речення дефініції [Fontenelle 1998]; В.С. Перебийніс зазначає, що зі структури дефініції можна автоматично виводити також окремі метамовні дескриптори [Перебийніс 2009, с. 208]. У свою чергу зазначимо, що, в тлумачних словниках неминуче є певні недоліки в дефініціях, через що нам видається раціональнішим і перспективнішим використання словників синонімів та антонімів в автоматичному опрацюванні значень слова, адже в них робиться акцент на зв'язку лексеми з іншими, її місці в лексичній системі мови. Безумовно, ця ідея потребує експериментальної перевірки.

Отже, ретельне впорядкування дефініцій сприятиме уніфікації науково-технічних даних у світі, якіснішому використанню тлумачних термінологічних словників у системах штучного інтелекту. Не виключаємо і зворотні процеси, а саме, можливість генерування дефініцій машиною за допомогою контекстних зв'язків, параметрів частиномовної та семантичної розмітки, наслідування ознак гіперонімів.

2.3.1. Проблема кругових визначень у дефініціях

Кругове визначення поняття – це таке визначення, що в дефініції повторює форму поняття, яке підлягає визначенню. Кругові визначення можуть мати декілька рівнів. Кругова дефініція першого порядку містить повтор лексики безпосередньо в тексті дефініції; дефініції вищих порядків містять повтор уже не в самій дефініції, а в дефініціях її складових. Наприклад: "сонет – віршова форма, що підпорядковується правилам оформлення сонету" або "Диверсифікація – (лат. *diversus* – різний і лат. *facere* – робити) процес розвитку діяльності підприємства, пов'язаний із збільшенням діапазону видів та проникненням в нові сфери діяльності, освоєнням нових виробництв, розширенням асортименту товару, що включає не тільки диверсифікацію товарних груп, але й розповсюдження підприємницької діяльності на нові та не пов'язані з основними видами діяльності фірми [Бутенко Н.В. Диверсифікація

виробництва: цілі та стратегії реалізації / Н.В. Бутенко // Економіка АПК. – 2003. – №7 – с. 109-114]”. Подібні дефініції однозначно неприйнятні для тлумачних словників, і помилки подібного роду насправді трапляються в опублікованих словниках украї рідко. Однак цілком можливо, що в дефініції лексеми міститимуться такі лексеми, дефініції яких включатимуть вихідну лексему. Таке явище можна назвати круговою дефініцією другого порядку. Кругові визначення другого і третього порядку значно поширеніші, позаяк вони не такі помітні на перший погляд. Ось спрощений приклад кругового визначення другого порядку: у дефініції іменника “child” згадується лексема “son” та *viceversa*:

CHILD: Young boy or girl, Son or daughter of any age, Descendant or follower, Person or thing as the product of particular circumstances or influences, Childish person

SON: Male child, Son-in-law, Form of address to a boy or young man, (cap.) Second member of the Trinity; Jesus Christ

[LingvoSoft Online Explanatory Dictionary]

Хоча в дефініціях допускається певний повтор, його випадки обмежені чіткими принципами. За логікою речей, можна ґрунтувати дефініції на однокореневих словах, якщо текст дефініції спирається на семантично простіше слово для тлумачення складнішого [Скороходько 1983, с. 8]; при укладанні дефініції діє також етимологічний принцип: новостворені слова тлумачаться на основі утворених раніше. Однак у поданому вище прикладі не одразу очевидно, яка з лексем виникла в англійській мові раніше, “son” чи “child”.

Як уже сказано, кругові дефініції неминучі. Проаналізуємо значення іспанського іменника “sentido” (“смысл”), тлумачення якого викликає особливі труднощі і фактично має передавати не лише якийсь сенс, а виступає масштабним метапоняттям. У словнику Королівської академії іспанської мови воно тлумачиться так:

Sentido: *m. Interpretación de que puede ser objeto un mensaje o una obra.*

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 [DRAE].*

Як можемо переконалися, для розуміння поняття “sentido” звернення до лексем “mensaje”, “interpretar”, що містяться в його дефініції, у свою чергу, вимагають розуміння вихідного поняття “sentido”, і тому дефініції лексем “mensaje”, “interpretar” становлять кругові визначення другого порядку.

В окремому дослідженні, проведеному на матеріалі словника англійської мови “Gutenberg Etext of Webster's Unabridged Dictionary” [Фокін 2018] шляхом автоматизованого опрацювання словника, здійсненого М.С. Івановим, показано, що більшість кругових дефініцій у словнику мають другий порядок. На перший погляд, такий кричущий факт не може не викликати подиву і обурення: як у принципі можна користуватися словником, якщо вже на другому порядку користувач потрапляє в замкнене коло? Але це не суперечить можливості розуміти тлумачення переважної більшості 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 окремо. А поза тим, при автоматичному пошуку дефініції другого порядку знаходяться через те, що вони надто деталізовані, і після пояснення значення надається також список синонімів леми: у тлумаченні англійської лексеми “assigasy” у списку синонімів міститься “precision”, а в тлумаченні “precision” використано синонім “accuracy”.

У той самий час нерозуміння одного слова в масштабній дефініції ще не означає нерозуміння читачем тлумачення в цілому, тому користувачі словників відносно рідко наштовхуються на кругові дефініції вище другого порядку. Людській мові властива інформативна надмірність, яка, за підрахунками Р.Г. Піотровського, становить 70-80% [Піотровский 1968, с. 58], тож можна розуміти текст навіть за умов розуміння лише певного відсотка його лексичних складових.

2.4. Методи пошуку перекладу терміна за допомогою комп'ютерних технологій

Незважаючи на величезну кількість двомовних перекладних словників у найрізноманітніших мовних парах, величезна частина термінів до них не потрапляють, отже перекладачеві необхідно володіти прийомами пошуку перекладу терміна у довідкових джерелах та методикою встановлення еквівалентних відношень між різномовними термінами. Необхідно це і лексикографам при упорядкуванні сучасних комп'ютерних словників.

Слід визнати, що певною мірою самі білінгви можуть спиратися на власний мовленнєвий досвід для встановлення відношень еквівалентності між одиницями тих мов, якими добре володіють. У такому разі провідний критерій встановлення еквівалентності між словами – онтологічний, тобто, коли один і той самий предмет, явище або характеристика номінується різними мовами. Не відкидаючи такий, безумовно вартісний, метод, мусимо визнати, що він не позбавлений суб'єктивності. З іншого боку, часто потрібно знаходити еквіваленти для тих характеристик та явищ, які білінгв у силу свого досвіду може відповідно номінувати лише в одній мові або ж не знає його взагалі й не може назвати жодною з мов. Тоді неможливо встановити еквівалентність емпіричним шляхом. Тому перекладачам потрібні об'єктивні методологічні знаряддя, щоб встановлювати відношення еквівалентності між одиницями різних мов. Зокрема, з цією метою можна використовувати такі методи: семантичний аналіз і синтез, пошук у паралельних текстах, калькування, транскрипцію, голонімічний пошук, гіперонімічний пошук, дистрибутивний аналіз, використання мови-посередника. Розглянемо детальніше кожен з цих методів.

Семантичний аналіз і синтез. Метод семантичного аналізу є найнадійнішим способом встановлення відношень еквівалентності між різномовними термінами. У разі наявності одномовних спеціалізованих тлумачних словників двома мовами доволі ефективним способом встановлення еквівалентних відношень між термінами обох мов буде пошук збігів у дефініціях.

Після аналізу та розуміння значення певної лексики під час перекладу необхідно віднайти спосіб її вираження іншою мовою. Останній можна розглядати як синтез семантичних характеристик лексики, які в дефініції розкладені на компоненти. Наприклад, відносно легко перейти від дефініції "перебування обвинуваченого в момент злочину в іншому місці як доказ його

непричетності до злочину” до лексеми – “алібі”. Річ у тім, що процес такого синтезу заледве піддається формалізації, і відбувається він переважно інтуїтивно. Здійснити аналіз терміна оригіналу на рецептивному етапі перекладу, тобто, перехід від слова до компонентів його значення можна здійснити елементарно, звернувшись до його дефініції у тлумачному словнику. Натомість на репродуктивному етапі перекладу синтез цілого значення на основі його компонентів складніший і частково можливий за допомогою ономасіологічних словників.

Однак власне процес синтезу значення багато в чому залежить від інтуїції, для нього немає чіткого алгоритму, і успіх синтезу залежить від здібностей творчого мислення й натхнення перекладача. Наприклад, дефініція “частина корпусу струнних та музичних інструментів, що відбивають і посилюють звук” описується поняттям “дека”. В ономасіологічному словнику цей термін може потрапити до різних категорій, відповідно до його гіпероніма та гіпоніма: гітара, звук, музика тощо. Оскільки процес синтезу не позбавлений суб’єктивності, його слід поєднувати з семантичним аналізом цільового терміна для перевірки його еквівалентності вихідному терміну. Якщо здійснити синтез не вдається, можна звертатися по допомогу до галузевих фахівців, які здатні довести процес синтезу до завершення.

Використання паралельних текстів. Щоразу більшою поширеністю характеризуються конкорданси паралельних текстів, в яких зручно шукати і перевіряти перекладні еквіваленти. До таких ресурсів належить словник *Glosbe*, який заснований саме на принципі конкордансу і містить українськомовні тексти у парі з англійською, іспанською, французькою та деякими іншими мовами. Безумовно, користуючись підказками таких корпусів, перекладач не може безапеляційно приймати варіант перекладу, використаний іншим перекладачем; до того ж, такого роду ресурси охоплюють далеко не всі мовні пари та предметні галузі. За умов відсутності відповідного паралельного корпусу текстів можна застосувати співвідносні або квазіпаралельні тексти. Наприклад, для перекладу цитати з Біблії з англійської мови на українську можна цілком використати безпосередньо український переклад Біблії, не вдаючись до пошуку оригіналу, першоджерел латиною, арамейською мовою. Оскільки англійський текст Біблії і український не перебувають у відношеннях оригінал-переклад, а є перекладами з одного й того самого джерела, ми такі тексти не

можемо назвати паралельними, і пропонуваній нами термін для їх найменування – квазіпаралельні тексти.

Калькування. Цей метод надзвичайно вартісний для перекладу багатокомпонентних термінів, і полягає у він відтворенні семантики кожного з цих компонентів іншою мовою. Наприклад, термін “підвищувальний трансформатор” на деякі романські мови перекладається еквівалентною калькою: “transformateur élévateur” (фр.), “transformador elevador” (ісп.), “trasformatore elevatore” (італ.), хоча в англійській мові номінація цього поняття інша: “step-up transformer”.

Наявність скалькованого в цільовій мові терміна ще не є гарантією семантичної еквівалентності: подібні за структурою і семантикою терміни можуть позначати різні поняття: іспанське словосполучення “estudios económicos” може позначати й дослідження в галузі економіки, і економічну освіту; “human population” (англ.), “población humana” (ісп.), “population humaine” (фр.) позначає не “гуманне населення”, а просто “населення”, “населення світу”, прикметник “людське” буде у цьому сполученні інформативно надмірним. Отже, якщо скалькований термін є широковживаним у цільовій мові, для встановлення еквівалентності з вихідним терміном необхідно здійснити семантичний аналіз.

На практиці здійснена калька нерідко призводить до помилкового перекладу або навіть неіснуючих лексем. Наприклад, було б неправильно перекладати калькою такі терміни: “друга негативна група крові”, яка французькою та англійською мовами позначається так: “A negative”. У країнах Західної Європи групи крові прийнято позначати не за нумерацією, а за типом білка; аналогічно не існує в зазначених мовах таких метафоричних термінологічних номінацій як “морська свинка”, яку правильно номінувати так: “guinea pig” (англ.), “conejillo de indias” (ісп.). Порівняймо з хибними перекладами: “*the second negative blood group” (англ.) / “*el segundo grupo sanguíneo negativo” (ісп.), “*sea pig” (англ.)/ “*cerdo del mar” (ісп.) та ін. Саме тому після здійснення кальки необхідно переконатися, що створений калькою термін справді існує в цільовій мові, хоча влучні варіанти трапляються не так вже й рідко: “глушник” – “silenciador” (ісп.), “silencieux” (фр.), “silenziatore” (італ.), “łumik” (польськ.); “спідометр” – “velocímetro” (ісп.); “вихлопна труба” – “tubo de escape” (ісп.), “pot d’échappement” (фр.), “tubo di scappamento”; “вихлопний газ” –

“gas de escape” (ісп.); “морський коник” – “caballito de mar” (ісп.)¹. Після перевірки наявності в цільовій мові скалькованого терміна слід також перевірити його значення для уникнення так званих “хибних друзів” перекладача. Калькування як спосіб перекладу може супроводжуватися й аналітичним розкладанням терміна, а надто під час перекладу на ті мови, що характеризуються високою аналітичністю, і тоді замість одного складного за структурою слова у перекладі вживатиметься словосполучення, наприклад: “одноденний” – “de un día” (ісп.), високотемпературна піч – “four à haute temperature” (фр.).

Транскрипція. Так само, як у калькування, у цього метода є свої переваги і недоліки. Слід зважати й на те, що в чистому вигляді транскрипція застосовується відносно рідко: частіше вона супроводжується мінімальними адаптаціями відповідно до фонетичних та дериваційних моделей у цільовій мові, наприклад: “diletante” (ісп.), “dilettante” (італ., фр.) – “дилетант”. Транскрипція разом з адаптацією може бути ефективним методом пошуку когнату до вихідного терміна в цільовій мові, зокрема, при перекладі численних греколатинізмів з української мови на іспанську: “бліндаж” – “blindaje”; “окуляр” – “ocular”, “дренаж” – “drenaje”; тим не менш, часто такий спосіб може й схибити: еквівалентом терміна “антиутопія” у романо-германських мовах є “dystopia” (англ.), “distopía” (ісп.), “distopia” (італ.), “dystopie” (фр.), і вкрай рідко фіксуються вживання з префіксом *anti-*; аналогічно: “міраж” – “espejismo” (ісп.); “вітраж” – “vidriera” (ісп.), “vetrata” (італ.), “vitrail” (фр.); “фактаж” – “relación de los hechos” (ісп.) тощо.

Окремим питанням залишається транскрипція слів-реалій у перекладі. Чиста транскрипція в словнику носіям мови перекладу не потрібна, адже вона не забезпечу належної семантизації, у той час як у словниках для носіїв мови оригіналу транскрипція цільовою буде цілком зрозумілою (як, скажімо, для українців не становить труднощі зрозуміти транскрипцію “oblast”, а для носіїв англійської мови це слово звучить як екзотизм і потребує пояснень).

Гіперонімічний і голонімічний пошук. Часто у перекладі допомагають багатомовні енциклопедії, адже одне й те саме явище, предмет, термін, назва тощо в них описано й названо

¹Зрозуміло, що вказані кальки супроводжуються мінімальними граматичними адаптаціями відповідно до норми та узусу.

різними мовами. Утім, коли явище, термін, предмет дуже відомий в одній культурі, а в іншій – ні, він може не потрапити до всіх мовних розділів багатомовної енциклопедії. Можна провести аналогію з назвами витворів мистецтва, які набувають популярності в одній країні з певних історичних обставин, а в іншій країні відомі лише вузьким колам. У таких випадках для пошуку відповідного терміна зручно шукати їхній переклад через голонім: якщо необхідно перекласти назву твору – необхідно знайти його автора. Аналогічно, якщо термін позначає складову частину явища або механізму – раціонально починати пошук з назви самого механізму чи явища. Позаяк більш імовірно, що голоніми з'являться в енциклопедії, існує більше ймовірності знайти їх у довідниках та енциклопедіях, а у статтях про них – відповідний термін з-поміж меронімів. Наприклад, термін “осердя котушки” може й не потрапити до багатомовної енциклопедії, однак, більш вірогідно, поняття “котушка індуктивності” в енциклопедіях зафіксований. Прочитавши статтю про котушку індуктивності, перекладач легко знайде назву осердя, оскільки це одна з її ключових деталей.

В аналогічний спосіб можна реалізувати і гіперонімічний пошук. Скажімо, для перекладу відносно рідко вживаного терміна з образною назвою “valet de nuit” (фр.), “servant of night” (англ.), “galán de noche” (ісп.), що позначає різновид меблів, доречно шукати його номінації в довідкових джерелах через узагальнюючу категорію меблів; пошук може бути звужений до субкатегорії “меблі для спальні”, серед яких доволі швидко можна знайти зазначений предмет, що в українській мові зазвичай називається “вішалка костюмна”.

Голонімічний та гіперонімічний пошук – потужні, однак не універсальні методи; їхнє застосування вимагає певного розуміння семантичних властивостей терміна, ерудиції, аналітичних здібностей; не для всіх понять легко встановити найближчий гіперонім або голонім; у довідкових описах певних явищ можуть і не зустрітися гіпоніми або мероніми, які ми прагнемо знайти.

Дистрибутивний аналіз. Дистрибуція в контексті лінгвістики – це властивість знаку поєднуватися з іншими знаками. Це означає, що для пошуку переклад дво- або багатокомпонентних термінів можливо починати пошук з ключового слова, і очікувати, що необхідні слова-визначники зустрінуться у найближчому оточенні ключового слова.

Порівнюючи списки дистрибуцій певного терміна у двох мовах, можна легко встановити еквівалентні відношення між словосполученнями за допомогою кальки та/або елементарного семантичного аналізу. Наприклад, якщо у корпусі іспанської мови CREA провести пошук слова “*pliegue*” (“складка”) у розділі “*Geología*” (109), отримаємо 15 таких дистрибуцій:

- *charnela de un mismo pliegue*;
- *elementos de un pliegue*;
- *elementos básicos de un pliegue*;
- *los flancos o lados del pliegue*;
- *charnelas del pliegue*;
- *el plano de simetría del pliegue*;
- *el eje del pliegue*;
- *pliegue simétrico*;
- *pliegue simple monoclinal*;
- *pliegue curvado*;
- *núcleo del pliegue*;
- *pliegue curvado*;
- *pliegue inclinado*.

Більшість результатів – це складні терміни. На жаль, наявні корпуси української мови наразі не дають змоги вести пошук за тематичними категоріями, але при порівнянні результатів іспанської мови зі спеціалізованими або енциклопедичними статтями про цей саме термін українською мовою, зможемо побачити його типову дистрибуцію і в українських текстах:

“У складках виділяють основні елементи:

- 1) крила – бокові частини складки;
- 2) замок – ділянка перегину чи переходу одного крила в інше;
- 3) ядро – внутрішня частина складки, що знаходиться між її крилами і замком;
- 4) осьова поверхня – поверхня рівновіддалена від крил складки;
- 5) шарнір складки ~ лінія перетину поверхні одного шару складки з її осью поверхню;
- 6) вісь складки – лінія перетину осьової поверхні з горизонтальною площиною.

Складки, в ядрі яких знаходяться молодші породи, ніж на крилах, називаються *синклінальними*. Складки, в ядрі яких залягають древніші породи, ніж на крилах називаються *антиклінальними*. У вертикальному розрізі неперекинуті

синклінальні складки вигнуті донизу, антиклінальні увігнуті наверх. Для визначення типу складок на геологічних картах треба дивитися на вік порід в ядрі та на крилах.

За характером нахилу осьової поверхні розрізняють складки: Пряма (симетрична) складка – її осьова площина вертикальна. Коса (похила) складка – її осьова площина нахилена, крила падають у протилежних напрямках і під різними кутами” (Геологія з основами геоморфології: Підручник для студентів екологічних і географічних спеціальностей вищих навчальних закладів / Г.І. Рудько, О.М. Адаменко, О.В. Чепіжко, М.Д. Крочак. – Чернівці: Букрек, 2010. – 400 с.).

Підкреслимо, що поєднання дистрибутивного аналізу з калькуванням та семантичним аналізом може бути дуже ефективним, хоча й не завжди; приміром, переклад терміна “flancos del pliegue” (ісп.) за допомогою кальки цього метафоричного найменування – “*сторони складки” не виводить на правильний варіант перекладу – “крила складки”, у той час як англомовний термін перекласти калькою вдається: “fold wings” – “крила складки”.

У вищезазначеному прикладі дистрибутивний аналіз проведено у корпусах співвідносних текстів. Справа в тому, що метод дистрибуції, який виник у рамках структуралізму, сам собою не є строгим, оскільки в ньому до уваги береться не властивість знака, що обумовлює його сполучуваність з іншими знаками, а формальні характеристики його оточення. Тим не менш, у поєднанні з іншими методами, зокрема, статистичним аналізом, калькуванням, семантичним аналізом, використанням паралельних корпусів, він стає значно надійнішим та результативнішим.

Використання мови-посередника. В Україні поширеною, часом невиправданою, практикою є переклад термінів за посередництва третіх мов, найчастіше англійської і російської. Це обумовлено тим, що справді велика частина термінів з російської на українську безпомилково перекладається калькою, а англо-російські, іспансько-російські, французько-російські, англо-іспанській, англо-французькі та інші словники, що містять серед цільових мов англійську, російську, масштабніші і детальніші.

Кальки через мову-посередника особливо часто виявляються влучними при перекладі термінів-інтернаціоналізмів (приміром, “двигун внутрішнього згоряння” – “internal combustion engine”

(англ.), “motor de combustión interna” (ісп.), “moteur à combustion interne” (фр.), інтернаціональних фразеологізмів (наприклад, “дамоклів меч” – “Damocles’ sword” (англ.), “espada de Damocles” (ісп.), “spada di Damocle” (італ.), “épée de Damoclès” (фр.)). І цілком можливо, що в багатьох двомовних словниках може й не бути такої фразеологічної одиниці, натомість, дуже ймовірно, що її переклад буде легко знайти через мови, поширені у міжнародному спілкуванні, а, отже, забезпечені більш детальними двомовними перекладними словниками: англійська, іспанська, французька.

Натомість під час перекладу лексем суб’єктивно-оцінної або національно маркованої семантики, що мають статус концептів або наближуються до нього, використання мови-посередника може призвести до викривлення їхнього змісту. Слід визнати, що переклад таких лексичних одиниць може ускладнюватися вже на етапі перекладу на саму мову-посередника. Тим не менш, метод у цілому ефективний для пошуку перекладу термінологічних одиниць з чітко окресленою семантикою.

Окремо слід згадати номінацію флори і фауни, під час перекладу яких трапляються казуси, адже серед народних назв тварин або рослин нараховують часом десятки синонімів. В іспанській мові, до прикладу, термінографи зареєстрували до 30 видів тунця, а назва “lenguado” позначає 72 різні види риб [Renales 1992], у той час як безпомилково і однозначно наукові назви тварин та рослин оформлюються латинською мовою, яку в цій галузі використовувати доречно не лише як мову-посередника, а й навіть як еталонне джерело термінів, так само як і в медичній галузі.

Візуальний аналіз. Додатковим і дуже ефективним способом розпізнавання значення термінів з предметним значенням, часто виражених конкретними іменниками, є візуалізація, тобто, аналіз вигляду предмета або його зображення, що в нинішні часи елементарно можна здійснити у пошукових системах мережі Інтернет. Візуальний аналіз часом ефективніший, ніж семантичний, оскільки подекуди легше сказати назву візуально знайомого предмета, ніж уявляти його образ за диференційними семами з дефініції. Наприклад, дефініція “інструмент для лінійних вимірювань яких-небудь невеликих деталей, виробів, що складається з штангової шкали з ноніусом і двох ніжок, одна з яких нерухома, а інша пересувається вздовж шкали” відповідає зображенню нижче:



Окрім вище згаданих способів перекладу, різного роду пошук у ресурсах мережі Інтернет, скажімо, у пошукачі *Google* теж становить дуже цінне вміння як для перекладача, так і для фахівців різних галузей знань. Деякі найсуттєвіші в перекладі прийоми і техніки роботи з інтернетними ресурсами описано у дидактичних матеріалах, укладених Є.В. Долинським, К.М. Скибою [Долинський 2016, с. 37-41], які тут спеціально не повторюємо.

2.5. Міжнародні формати термінологічних карток

Як пояснено у попередніх розділах, термінологічні або спеціалізовані словники призначені для фахівців відповідних галузей або галузевого перекладу. У той самий час, словник може виявитися недостатньо деталізованим з точки зору термінографів, оскільки їм для впорядкування терміносистеми часто потрібні деталі, що для перекладача зайві, такі як, наприклад, фонемна, морфологічна структура, частиномовний склад, етимологія терміна тощо або факт приналежності терміна одночасно до декількох галузей знань. Безумовно, зазначена ситуація вимагає чіткої організації і координації процесу впорядкування баз даних. І хоча спершу термінографи доволі по-різному вели реєстри термінів, з часом цю роботу почали уніфікувати.

Завдяки плідній роботі з обміну досвідом термінографи дійшли висновку щодо необхідності створення універсальних або, принаймні, універсально зрозумілих форматів опису термінологічних одиниць.

Окрім універсально визнаного формату, сам текст опису термінологічної одиниці має легко піддаватися формалізації, щоб бути зрозумілим і фахівцям, і комп'ютерній техніці. Безумовно, сучасні комп'ютери не можуть охопити знання в людському розумінні, але машину можна навчити відрізняти один тип даних від інших з метою здійснення оперативного пошуку, класифікації, порівняння, зіставного аналізу, пошуку перекладного відповідника або інших формальних операцій. Для прикладу можливостей формального опису розглянемо термін "неустойка" та його трактування в Українській юридичній енциклопедії за редакцією Ю.С. Семушченка:

НЕУСТОЙКА

(штраф, пеня) – грош. сума або майно, які боржник зобов'язаний

передати кредиторів в разі порушення ним зобов'язання. Н. є спеціальною, формою цив. відповідальності, оскільки застосовується лише в тих випадках, коли її сплата передбачена законом або договором (ст. 204 ЦивК України). Штраф і пеня на види Н. Штраф – це Н., що обчислюється у відсотках від суми невиконаного або неналежно виконаного зобов'язання; пеня – Н., що обчислюється у відсотках від суми несвоєчасно виконаного грош. зобов'язання за кожний день прострочення виконання. Право на Н. виникає незалежно від наявності у кредитора збитків, завданих невиконанням або неналеж. виконанням зобов'язання боржником. Залежно від підстав виникнення Н. розмежовуються на законні й договірні. Коли Н. передбачається певними нормат.-прав. актами щодо окр. видів зобов'язань, вона називається законною; коли ж вона узгоджена особами при укладенні договору, то називається договірною. Якщо предметом Н. є грош. сума, її розмір встановлюється договором або законом. Розмір Н., встановлений законом, може бути збільшений у договорі. Сторони мають право зменшити розмір Н., встановленої законом, крім випадків, передбачених законом. Розмір Н. може бути зменшений за рішенням суду, якщо він значно перевищує розмір збитків, та за наявності ін. обставин, які мають істотне значення. Сплата Н. не звільняє боржника від виконання зобов'язання у натурі (ст. 207). Кредитор має право вимагати від боржника виконання зобов'язання у натурі, якщо це відповідає його інтересам. Таке становище є наслідком існування у цив. праві принципу реального виконання зобов'язань.

Сплата Н. не позбавляє кредитора права на відшкодування збитків, завданих невиконанням або неналеж. виконанням зобов'язання. Літ.: Зобов'язальне право. К., 1998. Я.М. Шевченко. [Юридична енциклопедія]

Очевидно, що з такого суцільного тексту комп'ютер без додаткового специфічного алгоритму не здатний виокремити дефініцію терміна, характеристики, властивості, умови описаного явища тощо. Тим не менш, можливість автоматизованої вибірки таких функцій як дефініція, причина, умова, наслідок, тип, параметр та інш. (у термінах Н.М. Леонтьєвої – так званих “енциклопедичних функцій” [Леонтьєва 1978]), що присутні в більшості текстів, мають величезне значення для розмітки тексту. Якщо здійснити розмітку цього тексту за допомогою спеціальних тегів, машина зможе добувати позначені нею дані з тексту автоматично:

Схема 6. Приклад використання тегів для розмітки тексту за семантичними функціями.

<title> НЕУСТОЙКА </title>

<synon> (штраф, пеня) </synon> – <def> грошова сума або майно, які боржник зобов'язаний передати кредиторі в разі порушення ним зобов'язання </def>. <cond> Неустойка є спеціальною формою цивільної відповідальності, оскільки застосовується лише в тих випадках, коли її сплата передбачена законом або договором (ст. 204 ЦивК України) </cond>. <typ> Штраф і пеня на види неустойки. Штраф – це неустойка, що обчислюється у відсотках від суми невиконаного або неналежно виконаного зобов'язання; пеня – неустойка, що обчислюється у відсотках від суми несвоєчасно виконаного грошового зобов'язання за кожний день прострочення виконання. </typ> <cond> Право на неустойку виникає незалежно від наявності у кредитора збитків, завданих невиконанням або неналежним виконанням зобов'язання боржником. </cond> <typ> Залежно від підстав виникнення неустойки розмежовуються на законні й договірні. Коли неустойка передбачається певними нормативно-правовими актами щодо окремих видів зобов'язань, вона називається законною; коли ж вона узгоджена особами при укладенні договору, то називається договірною. </typ> <prop> Якщо предметом неустойки є грошова сума, її розмір встановлюється договором або законом </prop>. <param> Розмір неустойки, встановлений законом, може бути збільшений у договорі </param>. <cond> Сторони мають право зменшити розмір неустойки, встановленої законом, крім випадків, передбачених законом. Розмір неустойки може бути зменшений за рішенням суду, якщо він значно перевищує розмір збитків, та за наявності інших обставин, які мають істотне значення. </cond> <consec> Сплата неустойки не звільняє боржника від виконання зобов'язання у натурі (ст. 207). Кредитор має право вимагати від боржника виконання зобов'язання у натурі, якщо це відповідає його інтересам. Таке становище є наслідком існування у цивільному праві принципу реального виконання зобов'язань. Сплата неустойки не позбавляє кредитора права на відшкодування збитків, завданих невиконанням або неналежним виконанням зобов'язання. </consec> <ref> Літ.: Зобов'язальне право. К., 1998. Я.М. Шевченко </ref>. [Юридична енциклопедія].

Окрім безпосереднього пошуку функцій, зазначених у тегах, при автоматичному опрацюванні і добуванні даних до особливо цікавих результатів може призвести інтегрування даних з різних джерел. Приміром, дані щодо умов і наслідків неустойки, а також її синонімів (“штраф”, “пеня”) можна екстраполювати і на їхній гіперонім: “штрафні санкції”. Як уже зазначено, формат енциклопедичної статті дозволяє чітко розмежовувати тип даних, до яких має стосунок та чи інша частина тексту завдяки розмітці

тегами (спеціальними позначками у трикутних дужках). До того ж, на сучасних електронних носіях немає необхідності заощаджувати кожну літеру й скорочувати часто повторювані слова.

Найпростішою мовою, оснащеною можливостями розмітки тексту, можна вважати *XML (Extensible Markup Language)*, що, на відміну від *HTML ("HyperText Markup Language")*, передбачає можливість створення своїх власних тегів, залежно від потреби користувача, які і зазначено промовисто в назві ("extensible" – букв. "розширюваний"). Таким чином, *XML* дозволяє термінографу створювати теги для опису будь-якої додаткової корисної функції, будь-якого нового типу даних, як у схемі 6 вище.

Тим не менш, оскільки формати опису термінологічних карток зрештою мають бути уніфіковані в усьому інформатизованому просторі, необхідно, щоб створювані теги були читабельними і зрозумілими для всіх. Саме з цією метою на основі мови *XML* було розроблено міжнародний формат *TBX (TermBase eXchange)*, який зрозумілий і термінографам, і комп'ютерам. Приміром, щоб дати відповідь на питання: "Що означає "неустойка"?", "Що таке "неустойка"?", комп'ютер має знайти фрагмент тексту, що міститься всередині тегу `<def></def>`. У такий самий спосіб можна отримувати відповіді щодо умов застосування неустойки, так само як і інших характеристик цього поняття, описаних у статті.

Проблема, яка виникає у зв'язку з використанням тегів – їхнє невпорядковане і суб'єктивне творення на розсуд упорядника термінологічної бази. Для одних тегом, що виокремлює дефініцію, буде `<def>`, для інших – `<df>`, `<definition>` або навіть `<definición>`. Стандарт ISO 12620 стосовно формату *TBX* розроблено саме для уніфікації опису характеристик і значень, що можуть присвоюватися тим чи іншим функціям. Наприклад, відповідно до цього стандарту, всесвітньо визнаним тегом для дефініції прийнято вважати `/definition/`, для граматичного роду – `/gender/`. Якщо значенням, що вписується в тег `/definition/` може бути будь-який текст, для тега `/gender/` можливе наповнення строго обмежене такими варіантами: "masculine", "femenine", "neuter", "otherGender".

Оскільки *XML* дозволяє створювати власні теги, можна було б додати переклади цього терміна і розмістити їх усередині тегу на позначення відповідної мови, що виглядатиме так:

<eng>penal damages</eng>
<fr>clause pénale</fr>
<ital>pena pecuniaria</ital>
<esp>cláusula penal</esp>.

Однак створення тегів, дескрипторів, метамовних даних, описових полів тощо при внесенні термінів до бази даних не може відбуватися виключно на розсуд і смак того чи іншого фахівця. Термінологічні дані необхідно оформлювати відповідно до міжнародних вимог стосовно нормалізації термінології у стандарті ISO-1087 ("Terminological work: vocabulary" – "Терінологічна робота: словник"), визнаному 52 країнами, а з 2007 році – і Україною (держстандарт ДСТУ ISO 1087). Початково ISO-1087 розроблено комітетом ISO/TC-37. У ньому для опису термінів передбачено 132 теги; ось деякі з них: *object, subject field, domain, general concept, characteristic, subordinate concept, coordinate concept, intension, extension, hierarchical relation, generic relation, partitive relation, extensional definition, intentional definition, abbreviation designation, acronym abbreviation, synonymy, antonymy, nomenclature terminology, corpus, context, base list, grammatical label, language identifier* тощо [ISO-1087]. Формат обміну даними ФОЛИТ охоплює 168 можливих параметрів [Гринев 1990, с. 26]. Окрім цього, існує і широко використовується згаданий вище формат *TBX*. Отже, уніфікація можлива хіба що в перспективі. Приміром, поля² термінологічної картки у форматі електронного текстового документа могли б виглядати так:

Definition by intension in English:
Number of graphical words:
Parts of speech of each word:
Gender of nouns:
Etymology:
Main nomination in English:
Synonyms:
Hyperonym:
Hyponym:
Meronym:

²Полями є названі спеціальною метамовою характеристики, яким присвоюються конкретні значення; поля характеризуються певною назвою, що є ключем, та значенням, яке присвоюється ключу.

Holonym:
Ukrainian Translations:
Style register:
Typical collocations:

Якою б упорядкованою не була терміносистема, вона не позбавлена проблем і суперечностей. Так, наприклад, щодо дефініцій у стандарті зазначено, що “у виняткових випадках можна надавати два або більше визначення, якщо вони окреслюють поняття з різних точок зору; скажімо, одна й та сама картка може водночас містити спеціалізоване визначення, юридичне визначення та визначення для певної вузької галузі³” [ISO 1087-1:2000].

З точки зору комп’ютерної техніки описана ситуація означає, що для тих чи інших полів картки необхідно готувати параметри метамовного опису зі специфічними властивостями, а саме, з вузько обмеженим набором варіантів (у мовах програмування такі об’єкти часто іменуються терміном “set”, “enum”). Порухення цього питання спричинилося до створення стандарту ISO-24610-1 “Управління лінгвістичними ресурсами” (“Language Resource Management”) [ISO-24610-1] і “Загальна онтологія для опису мови” (“General Ontology for Linguistic Description”, що часто згадується під акронімом “GOLD”) [General Ontology]. У стандарті ISO-24610-1 наявні такі параметри для опису мовної одиниці: лема (реєстрові слова), лексема, морфемна структура, словотвір, морфолого-синтаксична розмітка, написання, транскрипція, синтагматичні відношення, методика вибірки тощо. Однак власне на роль еталонного претендує “Загальна онтологія для опису мови”. З-поміж інших полів вона містить такі: “стиль” або “регістр”, якому можна присвоювати лише 11 можливих значень, які тут подаються без перекладу як частини машинного коду: “bench-level”, “dialect”, “factious”, “formal”, “in-house”, “ironic”, “neutral”, “slang”, “taboo”, “technical” і “vulgar”.

Зазначені особливості формату карток необхідні для уніфікації різних термінологічних баз. На сьогодні їхня кількість разючо зростає. Тому розробники цих баз даних уже подбали про можливість їхньої взаємної сумісності і можливості автоматичної конвертації. Приміром, термінологічна база





MultiTerm може інтегруватися в системи машинного перекладу "Trados", "Déjà Vu" та інших аналогічних програм.

Найпотужніші міжнародні термінологічні бази доступні онлайн. Наприклад, ресурс *IATE* (скорочення від "InterActive Terminology for Europe") дозволяє проводити пошук двадцятьма п'ятьма мовами Євросоюзу [InterActive], так само як і Багатомовний тезаурус Європейського Союзу EuroVoc [Tesaurus Multilingüe]. В останній версії *IATE* додалися деякі фільтри пошуку, зокрема, за темами; передбачено можливість пошуку словосполучень; під час обробки запиту також залучаються результати з інших довідкових джерел ЄС: *EUR-Lex*, *EuroVoc*.

База даних термінів [EuroTermBank] охоплює всі офіційні мови Європейського Союзу, а також баскську, галісійську, ірландську, ісландську, російську, фарерську.

Ось приклад представлення термінологічного опису у Тезаурусі ЮНЕСКО:

Medicinal plants

-  Plants
-  Drugs, Ethnobotany, Traditional medicine
-  Science > Natural sciences
-  Plante médicinale (fr), Лекарственные растения (ru), Planta medicinal (es), Planta curativa (es) [THERSAURUS UNESCO].

Як видно, у першому полі міститься гіперонім, у другому – галузі, до яких має стосунок термін, у третьому – синонім терміна, у четвертому – наукова сфера терміна, у п'ятому – переклади терміна.

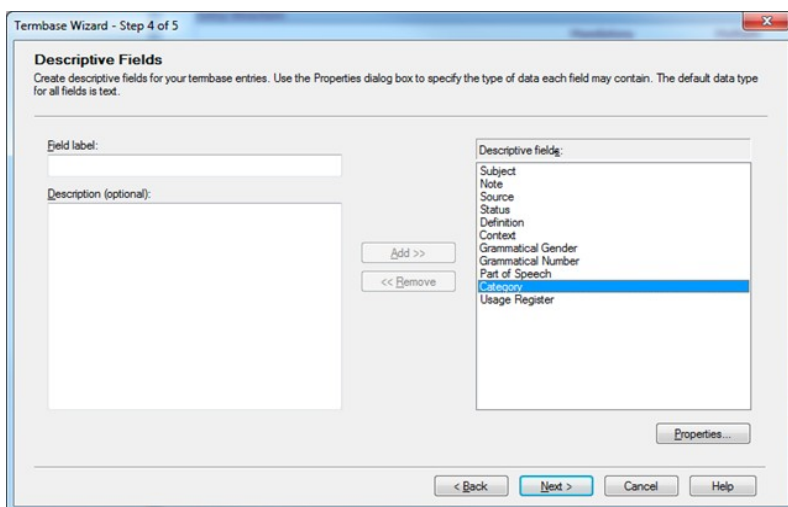
TERMDAT – термінологічна база даних Швейцарії, що передбачає можливість пошуку усіма мовами цієї країни (італійською, німецькою, ретороманською, французькою), а також англійською. Розробники пропонують спеціальні курси як для розвитку вмінь користування цією базою даних, так і для упорядкування термінологічних карток та адміністрування даних, що також детально описано в технічному керівництві, розташованому на сайті. У цьому ресурсі термінологічна картка охоплює такі поля: відповідальна державна установа (одна з восьми зазначених), тип сполучуваності, статус картки

(прийнята/ не прийнята), код достовірності (перевірена фахівцями/ не перевірена фахівцями), галузь.

Звичайно, у цифрову епоху поняття “картка” на позначення відповідного фрагмента бази даних стає радше фігуральним. Термінологічна картка *TERMDAT* включає такі дані: термінологічне бюро, ідентифікаційний код, відношення, код предметної галузі, автор, код достовірності, лінгвістичний код, термін, скорочення терміна, ключові слова, дефініція, контекст, помітки, країна, джерело, коментар. У свою чергу, кожне з цих полів характеризується набором параметрів, які необхідно брати до уваги [Termdat].

Окрім загальних термінологічних баз даних, для перекладача надзвичайну вартість становлять і спеціалізовані ресурси. До прикладу, база даних назв видів риб *Fish Base*, яка утворена двадцятьма п'ятьма мовами світу, сьогодні охоплює 325 000 загальних назв риб [Fish Base].

У системах машинного перекладу користувачі часто послуговуються форматом *SDL Multiterm*. Інтеграція термінологічної бази даних у програми комп'ютерного перекладу забезпечує уніфікацію перекладу термінів в одному або декількох перекладацьких проектах. Ось приклад описових полів терміна під час інсталяції термінологічної бази даних (поля можуть адаптуватися залежно від специфіки бази даних):



Наявність різних принципів і схем оформлення термінологічних баз даних свідчить про те, що попри міжнародну стандартизацію далеко не всі установи її дотримуються при укладанні термінологічних карток. Попри численні зусилля з уніфікації формату, оформлення карток суттєво відрізняється як в окремих країнах, так і міжнародних організаціях, і часто це пов'язано з фінансовим питанням: для багатьох процесів не потрібна суттєва деталізація полів у картці терміна, а спрощений формат вимагає менше людських і фінансових ресурсів. Натомість для специфічних потреб розробляються дуже детальні бази даних. А. Мелбі, Г. Будін, С. Райт у свою чергу вказують, що формат обміну термінологічними даними має включати, як мінімум, такі обов'язкові поля: сам термін, мова оригіналу, еквіваленти іншими мовами, предметна галузь, джерело, а також іноді – посилання на інші пов'язані з ними терміни [Melby 1993, с. 71-88]. Тим не менш, зазначені автори пропонують гнучке рішення для подолання розбіжності форматів шляхом розподілу системи обміну термінологічними даними на варіанти з різним ступенем деталізації [Melby 1993, с. 105]. Видається, що саме регулюючим установам у царині термінологічної роботи належить сказати останнє слово у цих дискусіях і, зокрема, визначити універсальний необхідний мінімум інформації, що має зазначатися у термінологічній картці.

Для комп'ютерної лінгвістики такі формати мають суттєве прикладне значення. Справа в тому, що в електронній базі даних можна здійснювати сортування термінів за будь-яким наявним параметром, а це дає змогу автоматично або напівавтоматично упорядковувати не лише спеціалізовані перекладні словники, а й також одномовні й багатомовні тезауруси, ономазіологічні словники та інші ресурси, причому величезна частина процесу упорядкування може виконуватися автоматично.

Таким чином, більшість термінологічних форматів регулюється стандартами Міжнародної організації зі стандартизації (*ISO*), що охоплюють як формальні, так і змістові аспекти впорядкування термінології. Однак уніфікації немає: в окремих базах даних використовують відмінні формати, виходячи зі специфічних потреб виробництва або дослідження. Поширеним комп'ютерним форматом обміну термінологічними даними є *TBX*. Також перспективним для розмітки і позначення полів термінологічної картки є використання мови *XML*.

2.6. Критерії вибору з-поміж термінологічних синонімів

Певне наукове поняття має виражатися в ідеалі одним терміном, позбавленим синонімії. Але, з огляду на розмаїтість регулюючих установ та відсутність регуляції в окремих науково-технічних царинах, зрештою з'являються, часом численні термінологічні синоніми. Яскравою ілюстрацією цього може бути широка синонімія у назвах біологічних видів. Так, в іспанській мові, до прикладу, термінографи зареєстрували до 30 видів тунця [Kirchner 1992]. Не без іронії сказано про це устами Сервантеса в “Дон Кіхоті”: “Діло, як на те, діялося в п'ятницю, і в усій корчмі не знайшлось ніякого їства, окрім риби (та й тієї було обмаль), що в Кастілії називають тріска, в Андалузії дорш, в інших краях кабліон, а в інших – знов форелька. Дон Кіхота спитали, чи не спожил б його добродійство форельки, бо іншої риби все одно нема”⁴ [Сервантес 1985, с. 31].

Завдання лексикографа в такій ситуації – включати до словників тільки такі синоніми, наявність яких виправдана. І хоча це питання потребує спеціальних глибоких досліджень, в загальних рисах, виходячи зі спостережень сучасних словників, у словникових статтях часто з'являються синоніми термінів. Зокрема, це трапляється, наприклад, тоді, коли термін є реалією, що може бути відтворена цільовою мовою за допомогою різних перекладацьких прийомів, які не обов'язково забезпечують денотативну еквівалентність. Переклад аналогом не повною мірою передає значення вихідного терміна, а транскрипція не є достатньо зрозумілою; регулярно варіантність термінології виникає в тих випадках, коли відмінні номінації термінів пропонують різні міжнародні організації та національні термінологічні служби, як видно з записів міжнародної бази даних “Fishbase”:

Таблиця 7. Приклад оформлення назв риб у базі даних “Fishbase”

Загальна назва	Мова	Країна	Наукова назва	Установа
-----------------------	-------------	---------------	----------------------	-----------------

Abadejo	Spanish	Mexico	<i>Mycteroperca microlepis</i>	AFS
Abadejo	Spanish	Spain	<i>Pollachius pollachius</i>	FAO

[Fishbase]

Науковими назвами вважаються латинські номінації біологічних видів; назви, сформульовані іншими мовами – народні, комерційні, кулінарні, деякі з них можуть бути емоційно забарвлені, як у вище процитованому прикладі з “Дон Кіхота”. Однак навіть якщо приймаються ті чи інші назви як еталонні на міжнародному рівні, усе одно залишаються документи, що містять старі назви, які треба читати, розуміти, перекладати. Отже, процес нормалізації остаточно не завершиться ніколи, тим паче, якщо взяти до уваги ідеологічно марковані терміни, номінація яких викликає численні палкі дискусії і призводить до виникнення синонімів, що відповідають тій чи іншій ідеології. Тому перекладачеві необхідно виходити з чітких позицій під час вибору з ряду термінологічних синонімів, обґрунтовуючи свій вибір відповідним завданням перекладу та комунікативною ситуацією.

Т.Р. Кияк у цьому зв'язку пропонує проводити семантичний аналіз, заснований на ідеях Е.Ф. Скороходька, а саме, шляхом структурування значення слова у вигляді бінарного дерева, в якому кожен вузол відповідає певній семі. У такий спосіб проводиться бінарне розкладання компонентів значення терміна. Той синонім, що містить найбільшу кількість назв вузлів або ключових сем, найточніше відбиває значення терміна, і це забезпечує прозорість його мотивації та сприяє його легшому розумінню [Скороходько 1983, с. 78-81]. Поза тим, слід звертати увагу і на словотвірні особливості: споріднені терміни, безумовно, доречно номінувати однокореновими словами; ці споріднені терміни, у свою чергу, мають утворювати мережу споріднених понять, тобто, зовнішня зв'язність має віддзеркалювати змістові відношення між термінологічними одиницями [Скороходько 1983, с. 84-85].

Г.Г. Верба та З.О. Гетьман у свою чергу зазначають, що при перекладі іще одним суттєвим критерієм вибору з-поміж

термінологічних синонімів є тип і жанр тексту. Окрім того, перекладач має розрізняти суто стилістичні синоніми та квазісиноніми; останні характеризуються лише частковим збігом семантичних характеристик [Verba 2013, с. 216] і синонімами можуть виступати лише у певному контексті.

У пастку квазісинонімії можна потрапити, якщо безапеляційно використовувати відповідники, що пропонуються у двомовних перекладних словниках. Якщо одні перекладні словники вкрай лаконічні й неповні, інші вражають багатющою синонімією відповідників. Наприклад, у російсько-англійській частині словника "Мультитран" для перекладу юридичного терміна "истец" користувачеві пропонуються 30 синонімів: *complainant, prosecutor, libellant, litigant, claimant, suitor, party presenting a claim, complainant party, querens, appellant, claimer, litigant plaintiff, litigant in person, demandant, pretendant, actor, libelant, orator, petitioner; pursuer; querent; applicant; the petitioner; exactor, pretendent, plaintful, litigant in person, pro se defendant, counterpart, suppliant* [Словарь Мультитран], а 17 з них позначені як такі, що використовуються специфічно в юридичних текстах; в російсько-іспанській частині цього ж словника подаються 32 відповідники: *litigador, litigante, demandador, demandante, peticionario, querellador, querellante, reclamador, actor, accionante, actor civil, actor del proceso, acusador, deprecante, parte actora, parte acusadora, parte demandante, peticionante, pretendiente, promovente, promoviente, quejoso, reclamante, recurrente, solicitante, suplicante, titular de la denuncia, parte civil, parte querellante, querelloso, petente* [Словарь Мультитран]. Безумовно, деякі з них є квазісинонімами, оскільки окремі з них пов'язані гіпо-гіперонімічним або голо-меронімічним відношенням з вихідним терміном; деякі з них обмежені у своєму вживанні певним ареалом, а інші – юридичним контекстом (типом інстанції, типом суду тощо). Резонним, на перший погляд, кроком мав би стати семантичний аналіз, тобто, порівняння лексичного значення вихідного та цільового терміна у тлумачних словниках, щоб оцінити можливість їхньої взаємної еквівалентності в перекладі. Однак цей метод, інколи дуже цінний, не обов'язково допомагає розв'язати проблему вибору з-поміж низки синонімів. Порівняймо, наприклад, значення еквівалентних термінів "lawsuit" і "позов" англійською та українською мовами:

Lawsuit: *a claim or dispute brought to a law court for adjudication [Oxford English Dictionary].*

Позов: *звертання до суду з вимогою захисту цивільних прав [Словник Української мови].*

Одразу впадає в око, що за змістом дефініції суттєво різняться. Дослідниця з Автономного університету м. Мехіко Е. Лосанова Богомілова у зв'язку з цим наголошує: "Міра поняттєвої еквівалентності між мовами різниться і залежить від спеціальної підмови, в яких вони використовуються (...). Поняття можна зрозуміти тільки в контексті системи, до якої воно належить; тобто, перед тим, як порівнювати дві мови, необхідно в першу чергу охарактеризувати незалежні системи понять, що наявні в кожній мові окремо (...). Аналіз джерел дозволяє виявляти відношення синонімії, антонімії, гіперонімії, гіпонімії, співгіпонімії, стильові й регіональні риси, а отже – в такий спосіб встановлювати ієрархічні відношення категорій вищого рівня з категоріями нижчого рівня: субординації і координації. Випадки повної еквівалентності понять двох різних мов, тобто, ідентичність змісту порівнюваних термінів мають впливати з того факту, що вони займають однакові позиції в системі"⁵ [Lozanova Bogomilova 2006, с. 681]. У зв'язку з цим згадується думка Х. Ортеги-і-Гассета про те, що зміст поняття визначається радше завдяки тій ніші, яку йому відводять інші поняття, ніж його внутрішнім наповненням [Ortega y Gasset 1983, с. 490-491]. Аналогічні зв'язки можна спостерігати і в семантичних відношеннях термінів. Тим не менш, щоб дійти висновку стосовно однаковості позицій різномовних термінів у системах понять, необхідно провести глибоке, майже вичерпне дослідження характеристик терміна. Таке контрастивне дослідження може мати різні ступені складності, навіть коли йдеться про однакові об'єкти, оскільки різні стандарти в різних мовах спираються на відмінні описові структури та точки зору, що використовуються для опису поняття [Lozanova Bogomilova 2006, с. 681]. Зрештою, аналіз можна будувати на таких основних семантичних функціях як гіпо-гіперонімія, голо-меронімія, співгіпонімія. Що стосується пари "lawsuit" (англ.) – "позов", в англійській мові спостерігається значно більша кількість гіпонімів, тому для встановлення часткової еквівалентності необхідно

переконалися, що синоніми підходять для текстів заданого типу, реєстру та тематичної галузі.

Окрім зазначеного критерію, необхідно, щоб обраний термінологічний синонім легко й органічно інтегрувався в систему термінів відповідної сфери науки як за своєю формою, так і за змістом, а надто коли йдеться про формулювання споріднених термінів. Приміром, терміни "biathlon", "duathlon", "pentathlon", "deathlon" в англійській мові утворюють впорядковану номенклатуру назв видів спорту за змістом і за формою; в українській мові, натомість, у формальному аспекті зазначені терміни не уніфіковані, адже паралельно вживаються дублети: "біатлон", "тріатлон", "пентатлон", "декатлон", але: "п'ятиборство", "десятиборство". Хоча й в інтернетних джерелах починають вживатися однорідні назви: "двоборство", "триборство", "семиборство" тощо, проблеми впорядкування все одно залишаються: 1) останні номінації ще не потрапили до провідних тлумачних словників української мови; 2) необхідно забезпечити впорядкованість і відповідних дериватів: "біатлоніст/двоборець", "тріатлоніст/триборець", творення відносних прикметників від цих термінів тощо. Отже, в цій номенклатурі змішано і слов'янські, і греколатинські корені, що взагалі притаманно нашій термінології.

Під час перекладу терміна завдання термінографа й лексикографа відрізняється від завдання перекладача: коли останній має обрати один оптимальний термін для цільового тексту, термінограф має зібрати всі або більшість прецедентних перекладів. Вони слугуватимуть перекладачеві орієнтиром для вибору найдоречнішого в ситуації перекладу синоніма. Синоніми відрізняються між собою характером мотивації терміна, етимологією і внутрішньою формою, конотативною, часовою, локальною маркованістю. У зв'язку з розрізненням сфер вживань синонімів особливої значущості в термінологічних словниках набувають тематичні, стилістичні, соціолінгвістичні помітки.

2.7. Автоматизація вибірки термінів з текстових масивів

Науково-технічна комунікація, обмін досвідом у цій сфері, впровадження винаходів постійно поживляються. Відповідно, і обсяг текстів, які необхідно опрацьовувати, настільки зростає, що

доводиться широко використовувати знаряддя комп'ютерного перекладу та автоматичного реферування для оперативного ознайомлення зі змістом документів у загальних рисах. Для термінографів ця ситуація теж додає викликів: необхідно відстежувати, реєструвати, виносити на обговорення, звіряти і перекладати колосальну кількість нових термінів або паратермінологічних лексичних одиниць. На допомогу у цій праці приходять програми автоматизованої вибірки термінів. Переважна більшість таких програм працює з одномовними масивами текстів.

Ґрунтовний і водночас компактний огляд програм вибірки термінології з текстів однією мовою здійснює М. Чанталь Перес Ернандес. Програма *MATE (Machine Assisted Terminology Elicitation)* може виокремлювати з текстів терміни на підставі статистичних критеріїв або за даними колокацій. Згаданий автор описує й інші програми, а також можливі алгоритми вибірки з паралельних текстів одиниць, що є кандидатами у терміни. Програма *System Quirk* здійснює вибірку термінів на основі спеціальних поняттєвих схем ("conceptual graphs", англ.), розроблених Дж.Ф. Совою. Типовою є також вибірка термінів на основі аналізу структури; приміром, кандидатами у терміни вважаються словосполучення з такою структурою: іменник+прикметник та прикметник+іменник [Chantal, Pérez Hernández 2002].

Вибірку термінів з паралельних та квазіпаралельних текстів може здійснювати і програма *ParsEqv*, яка, однак, націлена переважно на вибірку будь-яких словосполучень і речень з метою подрібнення комп'ютерної перекладацької пам'яті на дрібні корисні еквіваленти [ParsEqv].

У програмах автоматизованого перекладу, заснованих на використанні комп'ютерної перекладацької пам'яті, також існує можливість автоматично вибирати терміни, термінологічні словосполучення з паралельних текстів. Така вибірка теж заснована на структурі речення, зокрема, частина речення, що стоїть перед дієсловом-зв'язкою або модальним дієсловом часто справді визначає або характеризує якесь ключове поняття. Докладніше про цю вибірку сказано у пункті 3.2.

Безумовно, наразі можна вести мову лише про допоміжну роль програм автоматичної вибірки термінів; словники, глосарії, списки термінів, що генеруються автоматично на основі паралельних текстів, потребують вичитування і звірки людиною.

2.8. Специфіка структурування найпростіших баз даних

Більшість сучасних електронних словників складаються з двох основних компонентів: інтерфейсу та бази даних, у якій зберігаються словникові статті. І якщо форма інтерфейсу суттєво модифікується з огляду на розмаїтість дизайну та вподобань різних користувачів, власне база даних, якщо і варіюється, то значно рідше, адже зміна структури бази даних неминуче заторкує архітектуру і принцип дії програмного забезпечення. Поповнення, коригування бази даних не впливає на функціональність програми.

Розглянемо найтиповіші властивості баз даних, а також її два найпоширеніші формати. Взагалі база даних являє собою сукупність відомостей, що впорядковані за певною системою так, аби забезпечити раціональний і безпомилковий доступ до компонентів бази. Коли в базі даних одні відомості пов'язані з іншими (а словник саме так і влаштований), вона називається реляційною. Двомовний словник, у якому еквіваленти цільової мови пов'язані з лемою вихідної мови, є прикладом такої реляційної бази даних. Натомість текст оповідання, що входить до складу звичайного текстового корпусу, можна розглядати як нереляційну базу даних. Оскільки все ж більшість словників містять список слів та пов'язаних з ними перекладів, значень, тлумачень або іншого роду коментарів, розглянемо найпоширеніші формати реляційних баз даних.

У більшості випадків реляційні бази даних найлегше представляти у вигляді двовимірних таблиць, де горизонтальні лінії (назви рядків) відповідають тим чи іншим об'єктам, а вертикальні лінії (колонки) містять значення, властивості, характеристики полів цих об'єктів, як ілюструється у таблиці 8. Приміром, для поля "Lemma" об'єкта 1 значенням є "Monday"; для поля "Grammar Class" об'єкта "Tuesday" значенням є "noun".

Таблиця 8. Приклад бази даних SQL

ID	Lemma	Grammar_Class	Translation
1	Monday	noun	понеділок
2	Tuesday	noun	вівторок
3	Wednesday	noun	середа
4	Thursday	noun	четвер

5	Friday	noun	п'ятниця
6	Saturday	noun	субота
7	Sunday	noun	неділя
8	week	noun	тиждень

Для доступу до якоїсь комірки цього формату, а саме, *SQL* ("*Structured Query Language*"), потрібні, щонайменше, дві координати: назва терміна (або порядковий номер, так званий ідентифікатор "ID" об'єкта) і назва характеристики. Назвою умовно можна вважати порядковий номер об'єкта, однак при укладанні словників за назву доречніше приймати саме форму леми.

Фрагмент у схемі 8а, на перший погляд, позбавлений таких координат, однак його обробка теж може бути формалізована за допомогою мови програмування. Якщо уважно поглянути на рядки схемі 8а, вони теж легко піддаються формалізації за допомогою комп'ютерної мови:

Схема 8а. Приклад бази даних

Monday n. понеділок
Tuesday n. вівторок
Wednesday n. середа
Thursday n. четвер
Friday n. п'ятниця
Saturday n. субота
Sunday n. неділя
week n. тиждень

Іншими словами, за допомогою машини можна дати інструкції шляхом певних команд, щоб дані, розташовані в одному рядку, вважалися дотичними лише до одного дня тижня. Проблема в тому, що, з огляду на широкую розмаїтість форматів та операційних систем, не можна бути впевненим, що одна й та сама послідовність символів завжди прочитуватиметься однаково. Таким чином, база даних у вигляді суцільного тексту, як у схемі 8а, виявляється доволі ненадійним джерелом інформації. Для зведення до мінімуму проблеми неоднозначності прочитання бази даних формуються з чіткими знаками-роздільниками, хоча й останні не є уніфікованими. Якщо у таблиці роздільниками слугують перегородки комірок, звичайно, при переведенні таблиці в машинний код перегородки зникають, і на їхньому місці використовуються спеціальні позначки кінця комірок. Найтипівішим (хоча не єдиним) роздільником програмісти світу вважають знак коми “,”. Це створює певні проблеми, адже відомості, що містяться в таблиці, теж можуть містити коми. Однак такий збіг може трапитися з будь-яким іншим символом, і лише за умов правильного адміністрування бази даних небажаної двозначності роздільників і знаків змістових одиниць можна уникнути.

Ще одним дуже поширеним форматом збереження даних, зокрема, для використання в мережі Інтернет, є *JSON* (“JavaScript Object Notation”):

Схема 9. Приклад оформлення бази даних у форматі JSON:

```
[  
{  
  "Lema" : "Monday",  
  "Grammar_Class" : "noun",
```

```

"Translation":      "понеділок",
},
{
"Lema": "Tuesday",
"Grammar_Class" :      "noun",
"Translation":      "вівторок",
},
{
"Lema": "Wednesday",
"Clase léxico-gramatical" :      "noun",
"Translation":      "середа",
},
{
"Lema": "Thursday",
"Grammar_Class" :      "noun",
"Translation":      "четвер",
},
{
"Lema": "Friday",
"Grammar_Class" :      "noun",
"Translation":      "п'ятниця",
},
{
"Lema": "Saturday",
"Grammar_Class" :      "noun",
"Translation":      "субота",
},
{
"Lema": "Sunday",
"Grammar_Class" :      "noun",
"Translation":      "неділя",
},
{
"Lema": "semana",
"Grammar_Class" :      "noun.",
"Translation":      "тиждень",
},
}
]

```

Як бачимо, у цьому фрагменті база даних складається з масиву об'єктів, доступ до яких теж відбувається за параметрами. Для отримання доступу до певної характеристики об'єкта теж необхідні параметри: назва лема, назва граматичного класу або порядковий номер та назва характеристики (останній варіант менш бажаний з огляду на

можливе розширення бази і зміни порядкового номера). Кількість параметрів, що міститься ліворуч у лапках, так само як і кількість колонок у форматі SQL теоретично не обмежена і залежить від ступеня деталізації бази даних, відповідно до потреб словника.

Отже, для впорядкування термінології найліпше підходять реляційні бази даних, які характеризуються наявністю взаємозв'язків між назвою об'єкта та його полів. Назвою об'єкта може виступати його порядковий номер; у лексикографічній практиці за назву об'єкта краще приймати лему.

Контрольні питання

1. Чим викликана необхідність співпраці термінографів світу?
2. Які труднощі можуть виникати під час електронного обміну термінографічними базами даних?
3. У зв'язку з чим перші термінологічні бази, укладені термінографами, не були уніфікованими?
4. Які міжнародні організації регулюють упорядкування термінології у світі?
5. Якими документами послуговуються термінографи для уніфікації описів термінологічних одиниць?
6. Які існують вимоги до упорядкування електронних термінологічних одиниць?
7. Чи можуть бути синоніми у термінологічній одиниці? Яким чином розв'язується проблема синонімії термінів у спеціалізованих словниках та термінологічних картках?
8. Чому деякі формати укладання термінологічних карток передбачають меншу, а деякі – більшу деталізацію.
9. Назвіть щонайменше десять параметрів термінологічної одиниці, що можуть міститись у термінологічній картці.
10. Назвіть один-два поширені формати укладання термінологічних карток для термінологічних баз даних.
11. Які існують вимоги до формулювання дефініцій термінів?
12. Які формальні параметри дефініцій можна використовувати для автоматичного отримання даних зі словників?
13. У чому полягає відмінність між дефініцією за інтенціоналом та дефініцією за екстенціоналом?
14. Якими рисами характеризуються реляційні бази даних? Яке їхнє значення для укладання термінологічних баз даних?
15. Чим відрізняється представлення даних у форматі SQL від

формату *JSON*?

16. Які складові частини термінологічної статті можуть підлягати комп'ютерній розмітці?
17. Які енциклопедичні функції можуть використовуватися у тегах для розмітки термінологічної енциклопедичної статті?
19. Назвіть способи пошуку перекладу терміна (за умов його відсутності у двомовних перекладних словниках).
20. Як доречно проводити пошук термінів, виражених конкретним іменником?
21. Яким методом варто скористатися для пошуку перекладу терміна, якщо відомо ймовірний переклад одного з його компонентів?
22. Чи може одна й та сама термінологічна сфера регламентуватися різними установами? Відповідь аргументуйте.

РЕКОМЕНДОВАНІ ДЖЕРЕЛА

1. Долинський Є.В. Інформаційні технології у професійній діяльності перекладача: навчальний посібник з інформаційних технологій для студентів 3 першого (бакалаврського) рівня, а також другого (магістерського) рівня вищої освіти, галузі знань “Гуманітарні науки”, спеціальності “Філологія”, “Переклад” / Є.В. Долинський, К.М. Скиба – Хмельницький: ХНУ, 2016. – 75 с.
2. Кияк Т.Р. Лінгвістика спеціальних мов і науково-технічний переклад // Українська конференція перекладацької галузі. – 2013. – Відео доповіді. – [Електронний документ]. Режим доступу: <http://2016.utic.eu/video/videodoklady>
3. Леонтьева Н.Н., Волковысская Е.В., Копылова О.В., Молчанова Т.В., Штернова О.А. Словарь энциклопедических функций и его роль в автоматическом индексировании // НТИ, Сер. 2., 1978, – № 7. – С. 23-29.
4. Гринев С.В. Основы лексикографического описания терминосистем: автореф. дис. докт. филол. наук. М., 1990. – 43 с.
5. Фокін С.Б. Терміни з іхтіології: перекладознавчий та лексикографічний аспекти // Структурно-семантичні і когнітивно-дискурсивні парадигми сучасного романського мовознавства: Матеріали Третьої Всеукраїнської наукової конференції романістів. – Одеса: Вид-во КП ОМД, 2010. – С. 85-86.
6. Aguilar L. Lexicografía y terminología aplicadas a la traducción. Curso práctico de introducción. – Bellaterra: Servicio de Publicaciones de la Universidad Autónoma de Barcelona, 2001. – 86 p.

7. Cabré M.T. La terminología y la traducción especializada // Manual de documentación y terminología para la traducción especializada. – Madrid: Arco/Libros. – P. 89-122.
8. Dancette J. L'intégration des relations sémantiques dans les dictionnaires spécialisés multilingues : du corpus ciblé à l'organisation des connaissances. – 2011 – [Электронный документ]. Режим доступа: <http://id.erudit.org/iderudit/1006177ar>
9. Fraile Vicente E. El tratamiento terminográfico de las expresiones idiomáticas en los diccionarios y glosarios de economía en línea en inglés y español. Su utilidad para el traductor // SENDEBAR, 2007. – nº 18. – P. 223 – 261.
10. Gonzalo García C.; García Yebra V. Manual de documentación y terminología para la traducción especializada. – Madrid: Arco/Libros; 2004. – 557 p.
11. ISO-1087-1:2000. Terminology work – [Электронный документ]. Режим доступа: Vocabulary https://edisciplinas.usp.br/pluginfile.php/312608/mod_resource/content/1/ISO_1087-1_2000_PDF_version_%28en_fr%29_CPDF.pdf