

## 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, частининомовний 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 всевітньо 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 досягнути 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. Семущенка:

### НЕУСТОЙКА

(штраф, пеня) – грош. сума або майно, які боржник зобов'язаний передати кредиторі 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. Леонтєвої – так званих "енциклопедичних функцій" [Леонтєва 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 кожен літер і 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, поля<sup>1</sup> 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:  
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 галузі<sup>2</sup>” [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”). Порушення 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

<sup>1</sup>Полями є названі спеціальною метамовою характеристики, яким присвоюються конкретні значення; поля характеризуються певною назвою, що є ключем, та значенням, яке присвоюється ключу.

<sup>2</sup> Переклад наш — С.Ф.

(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", "dialtect", "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 різко 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].

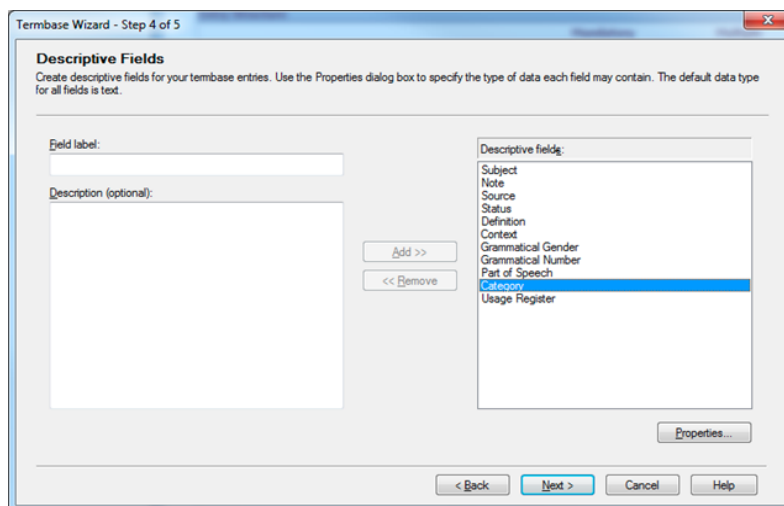
Як видно, у першому полі міститься гіперонім, у другому – галузі, до яких має стосунок термін, у третьому – синонім терміна, у четвертому – наукова сфера терміна, у п'ятому – переклади терміна.

*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, онемасіологічні 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*.

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 – [Електронний документ]. Режим доступу: <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. InterActive Terminology for Europe (IATE). – [Електронний документ]. Режим доступу: <http://iate.europa.eu/SearchByQuery.do>
12. ISO-1087-1:2000. Terminology work – [Електронний документ]. Режим доступу: Vocabulary [https://edisciplinas.usp.br/pluginfile.php/312608/mod\\_resource/content/1/ISO\\_1087-1\\_2000\\_PDF\\_version\\_%28en\\_fr%29\\_CPDF.pdf](https://edisciplinas.usp.br/pluginfile.php/312608/mod_resource/content/1/ISO_1087-1_2000_PDF_version_%28en_fr%29_CPDF.pdf)
13. Melby A., Budin G. & Wright S.-E. Terminology Interchange Format (TIF) // Sendebare: Revista de la Facultad de Traducción e Interpretación de Granada. – Granada, 1993. – nº4 – P. 69-110.
14. Termdat, Banco de Datos de la Administración Federal de Suiza [Електронний документ]. Режим доступу: <https://www.termdat.bk.admin.ch>