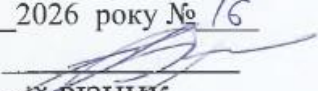


Міністерство освіти і науки України
Київський національний університет імені Тараса Шевченка
Навчально-науковий інститут філології
Кафедра української мови та прикладної лінгвістики

**УКЛАДАННЯ ТЛУМАЧНОГО ПЕРЕКЛАДНОГО СЛОВНИКА
ТЕРМІНІВ КОМП'ЮТЕРНОЇ ЛІНГВІСТИКИ**

Кваліфікаційна робота
на здобуття ОС «бакалавр»
студентки IV курсу
галузі знань 03 «Гуманітарні науки»,
спеціальності 035 «Філологія»
спеціалізації 035.10 «Прикладна
лінгвістика», ОПП «Прикладна
(комп'ютерна) лінгвістика та
англійська мова»
Анни Дмитрівни НЕТРЕБИ

Науковий керівник:
д. філософії, доц. кафедри
української мови та прикладної
лінгвістики
Юлія ЦИГВІНЦЕВА

«Допущено до захисту»
Протокол засідання
кафедри української мови та прикладної лінгвістики
від «5» 06 2026 року № 16
завідувач кафедри 
к.філол.н., доц. **Сергій РІЗНИК**

КИЇВ – 2026

ЗМІСТ

АНОТАЦІЯ.....	4
ВСТУП.....	6
РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ ТЕРМІНОГРАФІЇ ТА ЛЕКСИКОГРАФІЇ.....	9
1.1 Термінологія і термінографія.....	9
1.2 Стратегії перекладу термінів.....	11
1.3 Поняття визначення (дефініції) в термінографії.....	13
Висновки до розділу 1.....	14
РОЗДІЛ 2. ЛІНГВІСТИЧНІ ОСОБЛИВОСТІ ПЕРЕКЛАДУ ТЕРМІНІВ КОМП'ЮТЕРНОЇ ЛІНГВІСТИКИ.....	16
2.1 Актуальні проблеми в сучасних словниках та глосаріях із комп'ютерної лінгвістики.....	16
2.2 Переклад термінів комп'ютерної лінгвістики з англійської на українську мову.....	19
2.3 Проблема верифікованості новостворених тлумачних та перекладних словників.....	22
Висновки до розділу 2.....	24
РОЗДІЛ 3. СТРУКТУРА Й ПРИНЦИПИ УКЛАДАННЯ ТЛУМАЧНОГО ПЕРЕКЛАДНОГО СЛОВНИКА ТЕРМІНІВ КОМП'ЮТЕРНОЇ ЛІНГВІСТИКИ.....	25
3.1 Джерельна база словника і загальна характеристика словника.....	25
3.2 Етапи створення словника	27
3.3 Відбір термінів та особливості процесу перекладу термінів.....	28
3.4 Робота над дефініціями	30
3.5 Приклади до словникових статей.....	33
3.6 Представлення словника в “Lexonomu”.....	33
Висновки до розділу 3	37
ВИСНОВКИ.....	39

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	42
ДОДАТКИ.....	48
Додаток 1.....	48
Додаток 2.....	53
Додаток 3.....	54

АНОТАЦІЯ

Дипломна робота присвячена укладанню перекладного тлумачного словника термінів комп'ютерної лінгвістики. Таке завдання є доволі актуальним з огляду на нестійкість термінологічної бази, недостатню інформативність матеріалів англійською мовою для українських дослідників та істотне підвищення зацікавленості у представників суміжних галузей.

Об'єктом дослідження є термінологічна система англійської та української мов з галузі комп'ютерної лінгвістики. Предметом основні засади лексикографування, принципи та методи укладання тлумачних та перекладних словників. Метою роботи є моделювання та укладання тлумачно-перекладного словника. Завдання роботи полягає у вивченні основних понять лексикографії та термінографії, ознайомленні із стратегіями перекладу, а також підбір термінів, що складають реєстр словника, їх перекладі українською, добір дефініцій та прикладів використання. Були застосовані такі методи: описовий, аналіз, порівняльно-зіставний та лексикографічне моделювання.

Робота складається з трьох розділів – одного теоретичного, що висвітлює основні поняття термінографії та лексикографії, та двох практичних, що розкривають проблеми використання термінів комп'ютерної лінгвістики, їх кодифікації та перекладу, а також принципи та покроковий опис укладання словника.

Результатом дослідження став словник, що містить 124 терміни англійською мовою, до яких підібрано переклад, дефініцію українською та приклад використання. Словник розміщено на платформі "Lexonomy". Ця лексикографічна праця є інструментом для систематизації галузевої термінології, а також має перспективи для удосконалення та доповнення.

Ключові слова: *термін, комп'ютерна лінгвістика, тлумачно-перекладний словник, електронний словник, лексикографічне моделювання.*

ABSTRACT

The thesis is devoted to the compilation of a translation and explanatory dictionary of computer linguistics terms. Such a task is quite relevant because of the instability of the terminological base, the insufficient informativeness of materials in English for Ukrainian researchers and the significant increase in interest among representatives of related fields.

The object of the study is the terminological system of the English and Ukrainian languages in the field of computer linguistics. The subject of the study is the basic principles of lexicography, principles and methods of compiling explanatory and translation dictionaries. The purpose of the work is modeling, as well as compiling an explanatory and translation dictionary. The task of the thesis is to study the basic concepts of lexicography and terminography, to review translation strategies, and to select the terms, that create the dictionary register, their translation into Ukrainian, definitions and usage examples. The following methods were used: description, analysis, comparative analysis of dictionaries and lexicographic modeling.

The work consists of three sections: one theoretical, which highlights the basic concepts of terminography and lexicography, and two practical, which reveal problems in the computer linguistics terms usage, their codification and translation, as well as the principles and step-by-step description of compiling a dictionary.

The result of the research is a dictionary containing 124 terms in English, for which a translation, definition in Ukrainian and an example of use were selected. The dictionary is published on the “Lexonomy” platform. This lexicographic work serves as a tool for systematising the field's terminology, and also has prospects for improvement and completion.

Key words: *term, computational linguistics, explanatory and translation dictionary, electronic dictionary, lexicographic modeling.*

ВСТУП

Актуальність дослідження. Нестійка термінологічна база є однією з актуальних проблем у сфері української комп'ютерної лінгвістики. Щороку в міжнародному науковому середовищі з'являються нові публікації, що стосуються прикладної, комп'ютерної лінгвістики, опрацювання природної мови. Окрім фахівців-лінгвістів, інтерес до цієї тематики зростає і в представників суміжних галузей, наприклад математиків, розробників та дослідників проблем штучного інтелекту (далі – ШІ), а також і у звичайних користувачів. Однак частина закордонних розвідок та українських матеріалів, що оперують запозиченими термінами, поки що можуть бути недостатньо інформативними, особливо для початківців і тих, хто не знає англійської мови. Різниця між термінологічним апаратом, яким послуговуються різні дослідники часто занадто велика, що може викликати ускладнення в роботі та взаємодії, спричинити непорозуміння, зокрема під час вивчення курсів з комп'ютерної лінгвістики в закладах освіти. Ці аргументи виявилися достатніми, щоб зацікавитись переглядом основною джерельною базою та опрацюванням тематичних наукових, довідкових і навчальних видань. Станом на сьогодні існують деякі перекладні термінологічні словники та тезауруси з лінгвістики, зокрема “Тезаурус з лінгвістичної термінології” Н. Дарчук (Mova.info), “Словник термінів у сфері штучного інтелекту” (Чумаченко та ін., 2024), “Короткий англо-український словник з комп'ютерної лінгвістики” В. Коломієць (2022) та інші, але очевидною є потреба в доповненні наявних термінологічних словників актуальними одиницями, перекладом, тлумаченням та іншими відомостями. Тож створення та наповнення такого тлумачно-перекладного словника термінів, який міститиме інформацію про основні поняття комп'ютерної лінгвістики, вважаємо важливим завданням.

Об'єктом дослідження є термінологія комп'ютерної лінгвістики в українській та англійській фахових мовах.

Предметом дослідження є засади лексикографічного опису термінів, класичні та новітні підходи, принципи та методи укладання тлумачно-перекладного словника термінології комп'ютерної лінгвістики.

Метою кваліфікаційної роботи є моделювання та укладання тлумачного перекладного словника термінів комп'ютерної лінгвістики на основі актуальної наукової та лексикографічної літератури з теми комп'ютерної лінгвістики англійською та українською мовами.

Завдання роботи:

- з'ясувати основні поняття лексикографії, термінографії;
- ознайомитися зі стратегіями перекладу термінів;
- виконати основну роботу над словником, а саме:
 - дібрати найважливіші терміни з декількох тем підручника “Опрацювання природної мови та мовлення” Дена Джуравського та Джеймса Мартіна (2024), що є основною актуальною базою, де містяться терміни зі сфери опрацювання природного мовлення;
 - заповнити усі словникові статті відповідно до розробленої структури, а саме – переклад українською, дефініція, приклад використання та тематична належність;
- оформити готову базу даних у словник із інтерфейсом та активним доступом на платформі “Lexonomy”.

Основним джерелом термінів став підручник “Опрацювання природної мови та мовлення” Дена Джуравського та Джеймса Мартіна (2024) (дата останньої редакції підручника: 6.01.2026), де для цієї роботи обрано такі розділи:

- 4 – Логістична регресія та завдання класифікації тексту;
- 7 – Великі мовні моделі;
- 12 – Машинний переклад;
- 19 – Парсинг залежностей;
- 21 – Визначення семантичних ролей;
- А – Приховані моделі Маркова.

Теоретико-методологічною базою стали наукові праці українських учених у сфері української комп'ютерної лінгвістики, термінографії та лексикографії, зокрема Н. Дарчук, О. Зубань, Є. Карпіловської, В. Перебийніс та ін., українські словники та тезауруси з лінгвістичної термінології, а також суміжних наук.

Методи дослідження. Основними методами дослідження для написання теоретичної частини стали методи опису та аналізу, для практичної – порівняльно-зіставний аналіз словників, лексикографічне моделювання, а також методи, що застосовувались у добиранні перекладів: перефразування, порівняння, синонімічний переклад та ін.

Практичне значення. Укладений словник може слугувати інструментом систематизації та уніфікації термінів комп'ютерної лінгвістики в україномовному науковому й навчальному середовищі, а також джерелом необхідної інформації під час опрацювання англomовної спеціалізованої літератури. Словник, що містить і переклад, і тлумачення, і приклад використання із може стати основою для перспективних досліджень з лексикографії, термінографії та перекладу, стане в пригоді під час вивчення профільних дисциплін, а також відкритий до доповнення новими термінами, які можуть з'явитися пізніше й потребуватимуть кодифікації

Структура роботи. Робота складається зі вступу, трьох розділів – одного теоретичного та двох практичних, висновків до розділів, загальних висновків, списку використаної літератури (57 позицій) і додатків (список усіх реєстрових одиниць, таблична модель словника та посилання на словник). Загальний обсяг роботи – 54 сторінки, обсяг основного тексту – 40 с.

РОЗДІЛ 1.

ТЕОРЕТИЧНІ ОСНОВИ ТЕРМІНОЛОГІЇ, ТЕРМІНОГРАФІЇ ТА ЛЕКСИКОГРАФІЇ

1.1 Термінологія і термінографія

Термін – центральне поняття терміносистеми будь-якої галузі наукової діяльності. Це спеціальне слово, що використовується у науковому середовищі для представлення фактів і теоретичних спостережень, які вивчаються у галузі (Дарчук та Сорокін, 2012).

Слово “термін” є полісемантичним, тобто таке, що має кілька лексико-семантичних варіантів. У філософії та гносеології воно визначається як закріплення результату процесу пізнання, чи як засіб формулювання нової концепції. Водночас логіка трактує “термін” як двостороннє явище – таке, що називає і позначає поняття в науці, але воно має бути частиною не лише системи термінів, а і поняттєвої системи. У семіотиці – це знак, що є моделлю певної галузі знань.

Термін у лінгвістиці можуть трактувати залежно від погляду: це може бути особлива лексема, що входить до лексичної системи мови; або слово, яке не є лексично особливим, але виконує особливу функцію.

Існують етапи, що співвідносяться з цими визначеннями – термінологізації та детермінологізації. Термінологізація – це процес переходу слова в терміносистему галузі діяльності та набуття ним відповідного значення. Детермінологізація, відповідно, це процес втрати терміном своїх характерних функцій і значень в межах термінологічного поля.

Н. Дарчук посилається на В. Лейчика, що вважає терміном “лексичну одиницю для спеціальних цілей, яка позначає загальне конкретне чи абстрактне – поняття певної галузі знань” (2012).

Терміни мають особливі ознаки, що відрізняють їх від звичайних лексем, якими послуговуються у науковому контексті та поза ним. До найголовніших

належать: однозначність, системність, мотивованість, лінгвістична правильність, точність та термінологічна упродовженість (Дарчук та Сорокін, 2012). Важливим є те, що термін має позначати лише одне конкретне явище, і не допускати двозначності, крім випадків з міжгалузевими омонімами. Термін обов'язково є членом терміносистеми певної галузі науки. Не менш важливою характеристикою є точність терміна – для найбільш коректного передання явища та чіткого розуміння поданої інформації. Одна з характеристик терміна стала основою для нашого пошукового дослідження, а саме те, що на практиці або в неформальному спілкуванні науковці однієї галузі зазвичай надають перевагу найпоширенішому терміну або такому, що має більш тривалу історію. Також термін тяжіє до простіших форм та меншої кількості слів, якщо представлений словосполученням.

Утворення термінів у будь-якій мові відбувається декількома способами: за допомогою словотвірних і синтаксичних засобів, процесу термінологізації та запозичень з інших мов, де слово уже існує як термін, і активно вживається. Саме в комп'ютерній лінгвістиці спостерігається переважання запозичень, особливо часто – з англійської мови. Цьому факту неабияк сприяє перетин галузі зі сферами інформаційних технологій, математики і логіки, де іншомовні терміни використовуються набагато активніше або вже давно затвердились спільнотою.

Термін має багатофункційний характер. Найперше, він закріплює нове знання, тобто виконує номінативну, сигніфікативну та дефінітивну функції. Термін можна замінити його визначенням або, навпаки, використати визначення замість слова чи словосполучення без спотворення розуміння позначеного явища. Додатково він виконує важливу функцію передавання інформації, тобто комунікативну. Ефективність цієї комунікації буде найвищою між дослідниками однієї наукової сфери. Тим часом для аудиторії початківців або науковців із інших сфер термін є носієм пізнавальної функції.

Щоб визначити лінгвістичні методи дослідження терміносистеми, зокрема і комп'ютерної лінгвістики, варто надати дефініцію для цього явища, таким чином описавши його визначальні особливості. Терміносистема – це система

термінів, знакова модель усіх концептів певної наукової галузі сформована лексичними одиницями конкретної мови. Для створення терміносистеми необхідно, аби обрана наукова сфера активно розвивалася, мала сформовану базу понять та теорій, що здатні описати ключові явища, що виникають чи створюються.

Загалом терміносистеми мають такі основні характеристики:

- цілісність – елементи терміносистеми галузі охоплюють усі явища, що описує та чи інша наука;
- структурованість – багато терміносистем мають ієрархічну будову;
- зв'язаність – терміни, що складають терміносистему галузі науки, можуть бути пов'язані різними типами логіко-лінгвістичних відношень: синонімії, родо-видовими, частина-ціле, антонімії тощо (Дарчук та Сорокін, 2012).

Існують два підходи у дослідженні сталих та нових терміносистем:

- 1) аналіз усієї структури, семантики окремих термінів чи усієї терміносистеми;
- 2) вивчення термінів як одної сукупності, під час якого виявляють спільні та відмінні ознаки, а також визначають формальні зв'язки всередині системи.

Опанування цих підходів допоможе не лише у вивченні обраних концептів, а й відборі термінів для словників, найбільш змістовного і доцільного перекладу, а також для підбору якомога точних визначень.

1.2 Стратегії перекладу термінів

Для терміносистеми науки, яка ще у процесі розвитку, важливо зауважити, що терміни будуть походити або від інших, дотичних наук, або з мови, якою послуговуються більшість експертів цієї галузі. З такою швидкістю розвитку комп'ютерна лінгвістика швидко поповнюється новими поняттями і науковими працями, що їх описують. Переклад термінів напряду, що розглядається в цій

роботі як один із способів появи терміну в українській мові, потребує відповідних знань та контекстів реалізації. А. Коваленко пропонує два шляхи перекладу:

- 1) з'ясування значення терміна, найкраще – у контексті;
- 2) переклад значення українською (Коваленко, 2001).

Не менш важливою проблемою у сучасній українській мові є те, що існують терміни, що вже мають усталені відповідники, але натомість науковці та дослідники частіше необгрунтовано використовують прямі запозичення з іноземної мови, і саме це створює неусталеність терміносистеми.

Бувають і випадки, коли подібного відповідника не існує. Тоді перекладач обирає певну стратегію перекладу терміну. У перекладознавстві існує багато типів міжмовних трансформацій, якими можна послуговуватися і в термінографії, їх описали Д. Фурт та Л. Дмитрук:

- лексичні:
 - транскодування (транскрибування, транслітерування, адаптація);
 - калькування;
- лексико-семантичні:
 - конкретизація;
 - генералізація;
 - модуляція;
- граматичні:
 - членування;
 - об'єднання;
 - транспозиція;
 - переставлення;
- лексико-граматичні:
 - антонімічний переклад;
 - описовий переклад;
 - додавання;

- вилучення;
- контекстуальна заміна (2020).

Ознайомлення з усіма цими базовими методами дає чітке розуміння інструментарію для перекладача, аби якісно виконати переклад термінів, які в процесі не втрачатимуть своїх суттєвих характеристик. Перекладач також зможе пристосувати термін до норм мови перекладу, у нашому випадку – української мови, щоб переклад був граматично та стилістично коректний, а також зручний для використання.

1.3 Поняття визначення (дефініції) в термінографії

Питання про значення будь-якого мовного знака є досить актуальним для різних лінгвістичних галузей. Загалом, вважається, що це будь-яка інформація про мовний знак, яку людина може отримати з нього, а також про явище, яке воно позначає. Воно ніколи не функціонує окремо, а завжди є частиною складної семантичної конструкції. Її представленням є семантичний трикутник чи багатокутник. Термін «семантичний трикутник» було запропоновано основоположником логічної семантики Г. Фреге (1892), саме він представив ідею такого мовного знака, що складається з трьох взаємопов'язаних частин: денотат, сигніфікат і десигнат. Саме останній термін і трактують як “концепт” чи “значення” слова. Різні погляди на цю модель запропонували й інші науковці: Ч. Огден та А. Річардсон (1989), С. Ульман (1951) та ін.

Лексикографія продовжує вивчати поняття “лексикографічне визначення”, а також що конкретно має входити в нього, як його правильно будувати та формулювати, а також його класифікацію. На позначення опису чи трактування певного терміну (чи будь-якої лексеми загалом) існують поняття «визначення», «дефініція», «тлумачення», «опис». Дуже часто вони вживаються як синоніми, чи є взаємозамінними. “Словник української мови” в 11 томах трактує їх так:

ДЕФІНІЦІЯ, *ї, жін., книжн.* Стисле логічне визначення, яке містить у собі найістотніші ознаки поняття, що визначається (СУМ II).

ВИЗНАЧЕННЯ, *я, сер.* 1. Дія за значенням визначити 1—4. 2. Формулювання, вислів, у якому розкривається зміст чого-небудь, його істотні ознаки (СУМ І).

ТЛУМАЧЕННЯ, *я, сер.* 1. Дія за значенням тлумачити. 2. Текст, який містить пояснення, трактування чого-небудь (СУМ Х).

ОПИС, *у, чол.* 1. Дія за значенням описувати, описати 1—3. 2. Словесне зображення, словесна передача чого-небудь (СУМ V).

Порівняльний аналіз показав, що усі вони мають схоже пояснення, адже їхніми ключовими словами, або центральними семами, є “істотні ознаки” та “пояснення”.

З цього можна зробити висновок, що дефініція в термінології не лише має фіксувати саме пояснення наукового поняття, а і перераховувати конкретні ознаки, що відрізняють це поняття від іншого в межах терміносистеми, що вивчається. Словникові дефініції як джерело інформації про термін активно досліджують (Вакуленко, 2023), тож і в межах цієї теми вирішуються такі питання, як:

- 1) дефініції як матеріал для виділення семантичних ознак терміна;
- 2) виявлення структури термінологічних систем;
- 3) проведення компонентного аналізу та виділення сем.

Наявність дефініції є однією з опорних ознак для терміна і терміносистеми загалом. Чітка дефініція терміна – це необхідна умова для того, щоб термін став компонентом певної термінологічної системи.

Висновки до розділу 1

Цей розділ присвячено основам термінології і термінографії, усталення терміносистеми української мови, перекладу термінів та їх дефініювання. Ця інформація важлива для роботи з неусталеними терміносистемами наук, що активно розвиваються.

Укладання термінологічного перекладного словника не можливе без вивчення центрального поняття – терміна. Термін є ключовим словом для

терміносистеми будь-якої сфери наукової діяльності. База понять, що описують основні явища обраної наукової галузі, та відповідна лексика на позначення цих понять забезпечують можливість фіксування доволі сталої терміносистеми, що обов'язково матиме такі характеристики, як цілісність, структурованість та зв'язність.

Термін, що був перекладений з іноземної мови, має бути точним та чітко пов'язаним з поняттям та передавати усі значущі ознаки в повній мірі. Переклад термінів потребує знань у досліджуваній галузі. Існує доволі великий список стратегій, які застосовуються при перекладі термінів.

Дефініція (визначення) – це стисла інформація про мовний знак та явище, яке він позначає, що в своїй основі містить основні диференційні ознаки, які відрізняють один термін від іншого в межах терміносистеми.

РОЗДІЛ 2.

ЛІНГВІСТИЧНІ ОСОБЛИВОСТІ ПЕРЕКЛАДУ ТЕРМІНІВ КОМП'ЮТЕРНОЇ ЛІНГВІСТИКИ

2.1 Актуальні проблеми в сучасних словниках та глосаріях із комп'ютерної лінгвістики

Із значним зростанням інтересу до теми опрацювання природної мови за останній рік, крім наукових робіт, з'явилася велика кількість інтернет-статей та блогів, що описують не лише поверхневі факти про сферу, а і більш детальні дослідження, які охоплюють і використання відповідної термінології та вимагають від авторів текстів глибших знань.

Для того щоб проаналізувати, як сьогодні використовують термінологію з комп'ютерної лінгвістики не лише в наукових, а й у технічних, науково-популярних та навчальних українських текстах, було виконано пошукову роботу по декількох інтернет-платформах та блогах, зокрема “Друкарня” та “Medium”.

У наукових статтях більшою мірою автори послуговуються уже усталеними термінами, термінами, запозичені з англійської мови (транскрибовані, або ж кальковані), як-от *галюцинація (hallucination)*. Подекуди зустрічаються і публікації з використанням англійських термінів з латинським записом: *першим підходом, ... є **Back translation**, цей підхід передбачає використання **prompt-engineering*** (Абросімов, 2024).

Для науково-популярних та довідкових статей з теми опрацювання природної мови та суміжних з цією галуззю тем ситуація дещо інша. Часто автори використовують терміни непослідовно, тобто чергують запозичення і термін українською. Для порівняння, англійською зазвичай подають власні назви, наприклад назви моделі, технології, конкретної програми чи компанії: *Chat-GPT, Anthropic*. Такі випадки зустрічались і в інших науках, до прикладу *Übermensch* у філософії, чи *boni mores* у праві. Але частіше це виняток, аніж правило, що може застосовуватись до абсолютно всіх нових понять. Запозичена

лексика, що входить в ужиток, часто або поступово має змінюватися, адаптуючись до норм і принципів функціонування мови. Як показують дослідження функціонування запозичень різного типу (інтернаціоналізмів, варваризмів, неозапозичень тощо), абсолютна більшість слів проходять процес адаптації чи то в словотворі, чи то у морфології, чи то на фонетичному рівні (Клименко, Карпіловська та Кислюк, 2008, с. 9-35).

На платформі “Друкарня” за тегом “Штучний інтелект” створено понад 500 статей. Більша частина з них вживає термін AI (від англ. artificial intelligence) замість ШІ. В деяких статтях зустрічаються терміни англійською мовою, написані латинською абеткою, з поясненням або без: *Давайте приклади («few-shot prompting»)*, *open-source* рішення (kotovich, 2026), *Політика OpenAI: шифрування, ізоляція, але не end-to-end* (MaksymValin, 2026), *кращий для креативного письма та brainstormingu* (Vikt0ria, 2025), *PR для tech-брендів* (ukr_businessman, 2025), *Старі моделі демонстрували так званий recency bias (упередження новизни)*, *Сучасні моделі мають достатньо контексту з pre-training* (future.simple, 2025) тощо.

Наприклад, в тексті автора cybercalm про використання штучного інтелекту для дезінформування населень, послідовно використовуються терміни *великі мовні моделі, ШІ-системи, розпізнавання паттернів* (cybercalm, 2026). Але в той же час в статті про фатальні помилки ШІ, термін *vibe coding* записаний також латинкою (cybercalm, 2025).

Ще одним прикладом є статті технічних шкіл або курсів, які першочергово створені для подачі навчальної інформації, а не лише поверхневого ознайомлення з темою.

В глосарії від Inweb Media в одній статті використовується український термін (*галюцинація*), а в іншому – вже англійською (*LLM замість ВММ*) (Старк, 2025). У статті-глосарії від Олега Бондаренка в Medium, *NLP-ОПМ (опрацювання природної мови)* вказано як термін, але далі в тексті цей термін записується кирилицею – *НЛП* (Бондаренко, 2021).

У статті про машинний переклад та комп'ютерні засоби автоматизованого перекладу від MK-translations послідовно використовується англійська аббревіатура *CAT (computer-assisted tools)* в українському тексті. Варто зазначити, що цей термін в українських текстах спільноти перекладачів закріпився саме так і не дублюється українською (Венцковська, 2022).

В статті “Нові професії – для нових технологій. Чим займається Prompt Engineer” англійський текст використовують одночасно з українським, навіть для одного й того самого поняття. Прослідкуймо тенденцію в таблиці нижче (Гогілашвілі, 2024).

“Але <u>Prompt Engineering</u> не обмежується написанням команд для ChatGPT.”	“ <u>Промпт інжиніринг</u> – це галузь, яка передбачає наявність знань з багатьох інших галузей та сфер.”
“розуміти, як працюють <u>LLMs</u> ”	“Intel Labs навчили <u>велику мовну модель</u> ...”
“Поль зауважує, що вже бачив вакансію <u>prompt-інженера</u> на українському рекрутинговому сервісі”	“Модель непередбачувана: ми можемо на неї вплинути через <u>промпт</u> ,...”

Таблиця 1. Приклади непослідовного використання термінів комп'ютерної лінгвістики

Багато з таких проблем загалом вирішує “Словник термінів у сфері штучного інтелекту”, що був розроблений в 2024 році командою Міністерства з цифрової трансформації, і який також використовувався при укладанні словника термінів комп'ютерної лінгвістики як достовірне джерело. Але, на жаль, через швидкий розвиток галузі, недостатню кількість часу та уваги, які автори подібних розвідок приділяють вивченню теми, та відсутність узгодження в колективі науковців щодо термінології, що має використовуватись, такі лексичні неоднозначності трапляються все частіше.

2.2 Переклад термінів комп'ютерної лінгвістики з англійської на українську мову

У роботі над тлумачним перекладним словником термінів комп'ютерної лінгвістики було приділено велику увагу коректному перекладу термінів. Більшість термінів цієї сфери не мають відповідника через новизну самого явища, що позначається. Тому було запропоновано переклади, що можуть бути використані в українських текстах, з урахуванням зручності використання лексеми чи словосполучення у фаховій мові.

У цій роботі було прийнято користуватися умовно розробленою класифікацією термінів, яка вказує на сферу походження терміну. Обрана стратегія перекладу в межах таких груп майже збігається, тому що переважна більшість базових термінів комп'ютерної лінгвістики увійшли у терміносистему з однієї сфери або в один час. Це такі категорії, як:

- лінгвістичні терміни: *синтаксис, прикладка, вільний порядок слів*;
- математичні, фізичні терміни та терміни зі сфери інформаційних технологій: *корінь, вузол, лінійний список, конфігурація*;
- загальні терміни прикладної лінгвістики: *корпус, синтаксичний аналізатор (частіше вживається слово “парсер” як калька з англійської), морфологічна розмітка, штучний інтелект*;
- терміни комп'ютерної лінгвістики: *синтаксичний аналіз за принципами граматики складників, векторне представлення елемента, самонавчання, нейромережева модель* тощо.

Терміносистема традиційного мовознавства має сформовану структуру, про це свідчить велика кількість підручників, посібників та статей (Кочерган, 2006), (СУМ 1970-80), що використовують цю термінологію, нею також активно послуговуються в школах та університетах. Втім, дискусії щодо деяких термінів все ще є актуальними, лінгвісти пропонують власні назви певним поняттям та обґрунтовують вибір та потребу в заміні назви, як це, наприклад, зробив

І. Вихованець (1988). Тому терміни з цієї категорії давно мають українські відповідники.

Значна кількість понять прийшла в комп'ютерну лінгвістику з математики, фізики, математичної логіки та інших, більш вузьких галузей: теорії графів, автоматів тощо. З розвитком точних наук в Україні терміни з цих наук закріпились у дослідженнях та монографіях, і використовуються широким колом науковців і студентів. Не менш важливою сферою, з якої комп'ютерна лінгвістика запозичила чимало термінів, є кібернетика та інформаційні технології. Сучасна сфера опрацювання природної мови в своїй основі має здобутки комп'ютерних наук та інженерії програмного забезпечення. В Україні була видана велика лексикографічна праця – “Енциклопедія кібернетики” (Глушков, 1973), що змогла представити і закріпити терміносистему галузі кібернетики та інформаційних технологій. Ця енциклопедія активно використовувалась під час формування словника.

Варто згадати роботу над пошуком рішення щодо нещодавно запозичених термінів. Не можна заперечувати очевидної тенденції сучасних авторів наукових праць надавати пріоритет лексемам, що не відрізняються за вимовою або написанням від мови запозичення. Англійська є найпоширенішою в силу своєї високої популярності як мови міжнародного спілкування. Спостерігається активність в розвитку технологій та досліджень в сфері інформаційних технологій та програмування в англійськомовних країнах. З появою нових понять, проєктів, алгоритмів, мов програмування та технологій створення програмного забезпечення унормування відповідників у терміносистемі української комп'ютерної лінгвістики не встигає відбутися, тим паче на рівні словників та глосаріїв, а потреба у використанні специфічних термінів постає якнайшвидше. Про це свідчить і щорічне оновлення підручника Д. Джуравського і Дж. Мартіна, який ми використовували як основне джерело англійськомовних термінів в цій роботі, і постійне додавання нових реєстрових одиниць, перегляд їх актуальності станом на сьогодні, доопрацювання перекладів та дефініцій у словнику. Науковці та розробники тяжіють до простішої схеми використання термінів, що

потребуватиме менше мовних зусиль. Таким чином термін, хоч і не адаптувався до українського мовлення, але лексема вже входить в терміносистему українського наукового поля з моменту першого використання. Таке явище можна назвати одним поняттям “англізм” – це слово чи словосполучення, що було запозичене з англійської мови або створене за подобою (Приплюцька та Фаріон, 2019).

Загальновживані терміни комп’ютерної лінгвістики мають схожу проблему – дослідники вживають більш зручні, часом коротші англізми, хоча український відповідник також може існувати, наприклад: *parser* – *парсер* – (синтаксичний) аналізатор; *NLP* – *НЛП* (вимовляється як *ен-ел-пі*) – опрацювання природної мови; *POS-tagging* – *ПОС-тегінг* – морфологічна розмітка; *prompt* – *промпт* – *запит*, *NER* – *НЕР* – розпізнавання іменованих сутностей. Найпопулярнішими методами, що спостерігаються в цій категорії термінів, є калькування, використання відповідника, генералізація та описовий переклад.

Категорія вузькотематичних спеціалізованих лексем тут – це такі, що майже або взагалі не зустрічались в українській літературі з комп’ютерної лінгвістики. Для них не існує кодифікованого перекладу чи відповідника українською, їх немає в словниках із загальної лінгвістики та математики, в посібниках з прикладної лінгвістики. Вони зазвичай позначають нові явища, що виникають під час дослідницьких процесів. Окрім цього, деякими завданнями може займатися обмежене коло людей. Часто в процесі термінологізації використовуються такі стратегії як практичне транскрибування та описовий переклад. У такого типу перекладу існує певний недолік – певні поняття неможливо пояснити однаковою або навіть меншою кількістю слів, а довгі словосполучення не є зручними, часом не характерними для вжитку у мовленні. В нашому словнику зафіксовані такі лексеми: *dependency parser* – синтаксичний аналізатор на основі граматики залежностей; *embedding* – векторне представлення елемента (слова), частіше використовують слово *ембедінг*; *supervised finetuning (SFT)* – контрольоване / кероване дотренування

(допренування з наглядом / з учителем), CAT (computer-aided translation) – автоматизований переклад, переклад за допомогою комп'ютера та ін. Деякі з цих понять наукове коло пропонує залишити на рівні транскрипції, наприклад *ембединг* або *CAT*, аргументуючи це тим, що немає потреби в адаптації технічного терміна.

У своєму словнику пропонуємо переклади для більшості термінів, яких не було знайдено в словниках чи літературі українською мовою. Щоправда, декілька внесених словникових статей не мають перекладу взагалі, тому що для такого поняття немає затвердженого використання. У такому випадку пропонується внести свій варіант або ж використовувати його в зручній формі.

Отже, проблема використання англізмів, некодифікованої термінології приводить лінгвістів до проблеми нестійкості терміносистеми. Науковці та фахівці із суміжних галузей вживають різні терміни та словосполучення, хоча мають на увазі те саме явище чи процес.

Розв'язання цієї проблеми є актуальним завданням, що потребує думки та зусиль багатьох дослідників. Одну зі спроб запропоновано в цій кваліфікаційній роботі у вигляді короткого тлумачного перекладного словника. У ньому пропонуємо варіанти перекладу та дефініцій, що були виведені на основі усіх опрацьованих розвідок з різних тем комп'ютерної лінгвістики.

2.3 Проблема верифікованості новостворених тлумачних та перекладних словників

Під час створення словника термінів комп'ютерної лінгвістики був обраний підхід, що переважно спирається на порівняльний аналіз різних лексикографічних та галузевих праць, внесення тлумачення чи перекладу з посиланням на наукове джерело. Але в процесі роботи з перекладом і з тлумаченням виникли серйозні проблеми, одна з яких – це те, що термін, який фігурує в підручнику “Опрацювання природної мови та мовлення”, і, з нашого досвіду, цілком активно використовується спеціалістами, зовсім не трапляється

у наукових роботах, не є кодифікованим з різних причин: нещодавно поява поняття, вузькоспеціалізованість теми вивчення чи роботи, або відсутність надійних ресурсів для перевірки та оновлення наявних лексикографічних праць.

Власне терміни, що фігурують в тексті цього підручника, не є невідомими, адже видання наразі вважається одним з основних і найбільш ґрунтовних в цій галузі. Автори наводять приклади застосування, пояснюють явище, що описується певним терміном, що і робить цей підручник корисним матеріалом для отримання базових знань. Часто можна спостерігати використання термінології у спільнотах науковців, лінгвістів та інженерів-розробників програмного забезпечення, тобто у чатах, форумах та семінарах. Важливо додати, що спеціаліст обраного напрямку добре розуміє суть і має змогу сформулювати визначення на основі власного розуміння, навіть якщо формального, словникового визначення ще не існує. Проте наразі вибір на користь англізму чи кальки з англійської більш поширений, зокрема тому що українського відповідника може не бути, водночас і єдиної бази, на яку можна посилатися, що містила б відповідний затверджений термін, теж немає.

Усе вищезазначене порушує ще одну значну проблему – укладання сучасних словників на вузькоспеціалізовані теми стає дедалі складнішим, бо процес затвердження запропонованих у словнику перекладів та тлумачень, що були розроблені та внесені вручну, а не зі авторитетної наукової літератури, а також прийняття відповідних пропозицій науковою і дослідницькою спільнотою, є обтяжливим і повільнішим за розвиток сучасної комп'ютерної лінгвістики.

Тож можливим рішенням є перехід від прескриптивного характеру словників, який було обрано в цій роботі, до дескриптивного. Обсяг роботи для створення дескриптивного словника більший, але він і ширше представлятиме закономірну поведінку термінів у мовленні. Такий словник має містити усі найменування та відповідну інформацію для словникової статті, що може траплятися не лише у суто науковому колі під час формального спілкування, а і там, де працюють і взаємодіють усі дотичні до комп'ютерної лінгвістики

спеціалісти. В такий спосіб утвориться список, з яким можна працювати на рівні статистики, тобто обирати найвлучніші терміни, збирати дефініції і переклад, що точно є присутнім у мовленні. Дескриптивний словник забезпечує збереження реальних даних, отриманих із кіл активного вживання спеціалізованих термінів, гарантує найбільшу актуальність й може стати джерелом для стабілізації терміносистеми. Також такий словник не обов'язково має бути саме частотним, тому що першочергово покликаний на збір термінів, які вже увійшли у вжиток, та їх повномірну репрезентацію.

У роботі над нашим словником ми й надалі спираємось на попередні, уже створені словники та тезауруси з комп'ютерної лінгвістики, але також залучатимемо і метод лінгвістичного спостереження.

Висновки до розділу 2

Робота над словником термінів комп'ютерної лінгвістики не лише не втратила актуальності, а й з кожним роком набуває дедалі більшого значення. Окрім відомих галузевих праць, було опрацьовано і нові джерела, що містять актуальну термінологію. Особливу увагу було приділено проблемі нестійкого використання термінології в текстах як наукового, так і науково-популярного жанру. Невтішні тенденції спонукали до продовження роботи над словником і його оновлення більш релевантними термінами.

На конкретних прикладах було проаналізовано основні перекладацькі стратегії, застосовувані під час передачі англomовної термінології українською мовою. Встановлено, що найбільш поширеними з них є описовий переклад, калькування, практична транскрипція, а також генералізація. На основі проведеного аналізу запропоновані рішення щодо зменшення використання некодифікованої термінологічної лексики. Крім того, виявлено проблему, пов'язану з недостатньою достовірністю та коректністю унормування терміносистеми комп'ютерної лінгвістики в сучасних лексикографічних виданнях, що потребує критичного перегляду наявних словникових ресурсів.

РОЗДІЛ 3.

СТРУКТУРА Й ПРИНЦИПИ УКЛАДАННЯ ТЛУМАЧНОГО ПЕРЕКЛАДНОГО СЛОВНИКА ТЕРМІНІВ КОМП'ЮТЕРНОЇ ЛІНГВІСТИКИ

3.1 Джерельна база словника і загальна характеристика словника

Під час пошуку лексикографічних праць з комп'ютерної лінгвістики українською мовою, що могли б стати в пригоді при укладанні нашого словника, було виявлено небагато джерел. Лінгвістична термінологія з цієї галузі точно потребує оновлення та перезатвердження, тому вибрані словники будуть використані при укладанні цього словника під час перекладу, тлумачення та добору прикладів.

Найактивніше використовуються такі видання:

- “Короткий англо-український словник з комп'ютерної лінгвістики” В. Коломієць (2022);
- “Тезаурус з лінгвістичної термінології” Н. Дарчук (Mova.info);
- “Словник української термінології прикладної (комп'ютерної) лінгвістики” С. Дерби (2007);
- “Тезаурус з комп'ютерної лексикографії” О. Сірук та В. Сорокіна (Mova.info).

У 2024 році команда фахівців у сфері штучного інтелекту за сприяння Міністерства цифрової трансформації України уклала словник термінів у сфері ІІІ (Чумаченко та ін, 2024), що зміг охопити пласт найосновніших сучасних термінів, пов'язаних із використанням та розробкою систем ІІІ. Список реєстрових одиниць охоплює більш загальні терміни, які є часто вживаними. У словнику наявні як переклад, так і тлумачення українською, відповідний термін англійською, і, що доволі важливо, відсилання на тематично пов'язані терміни чи синоніми, а також похідні слова. Це дає змогу зробити висновок, що укладачі визначили тематичні межі термінів та на основі цього створили своєрідну

термінологічну мережу, встановивши їх взаємозв'язок, що в словникові представлений у вигляді гіперпосилань, розташованих у межах словникової статті. Цей словник також був використаний як актуальне джерело термінів та тлумачень.

Видання “Опрацювання природної мови та мовлення” Д. Джуравського та Дж. Мартіна (2024), а саме оновлена версія від 6 січня 2026 року, є основним джерелом для відбору англomовних термінів. Цей підручник є одним з визначних сучасних робіт з теми опрацювання природної мови. Ним рекомендують послуговуватися всім початківцям і студентам, а також тим, хто хоче детальніше дізнатися про внутрішню роботу з усним та писемним природним мовленням. У книзі представлено як основні поняття галузі, так і приклади алгоритмів та роботи програм, діаграми та таблиці, історичні довідки, описані математичні основи спеціальності та базова мовознавча інформація. Позитивною характеристикою також є широкий діапазону різноманітних тем. З кожною версією книги автори додають нові розділи, доповнюють наявні, а також продовжують виокремлювати важливі терміни, на це вказує і наявність термінопоказки в повній версії підручника. Таким чином було вирішено, що на цей підручник можна опиратися як на достовірне джерело. Звісно, при відборі релевантних термінів ми спиралися і на власний досвід опрацювання природної мови та розв'язання завдань з комп'ютерної лінгвістики.

Протягом роботи над словником було опрацьовано 7 різних тем підручника, а саме: 4 – Логістична регресія та завдання класифікації тексту (“Logistic Regression and Text Classification”), 7 – Великі мовні моделі (“Large Language Models”), 12 – Машинний переклад (“Machine Translation”), 19 – Парсинг залежностей (“Dependency parsing”), 21 – Визначення семантичних ролей (“Semantic Role Labeling”), а також додаток А – Приховані моделі Маркова (“Hidden Markov Models”).

Наш словник містить 124 реєстрові одиниці, 79 з яких стосуються теми “Парсинг залежностей”, 26 – теми “Великі мовні моделі”, 8 – “Машинний переклад”, 7 – “Логістична регресія та завдання класифікації тексту”, 2 –

“Приховані моделі Маркова” і 2 – “Визначення семантичних ролей”. До кожного англomовного терміна підібрано переклад українською, дефініцію українською, приклад англійською та позначено відповідну тематичну належність.

3.2 Етапи створення словника

Загальний робочий алгоритм створення, доповнення і оновлення словника термінів комп’ютерної лінгвістики містить такі кроки:

1. обрати розділ підручника для опрацювання і ознайомитись зі змістом розділу;
2. вибрати терміни загальної та комп’ютерної лінгвістики з кожного обраного розділу, звернути увагу як на найбільш популярні, так і на можливі незрозумілі терміни;
3. створити модель словника у вигляді простої бази даних (таблиці), де міститься структура словникової статті, наповнення статті та позначки, що вказують на джерело тої чи іншої інформації (така позначка існує лише в моделі словника і не була збережена в кінцевому варіанті словника на платформі “Lexonomy”)
4. додати реєстр усіх термінів до таблиці;
5. підібрати переклад до нових реєстрових одиниць, переглянути попередні переклади та, за потреби, оновити дані;
6. підібрати до всіх термінів влучні тлумачення, які будуть чітко пояснювати суть описуваного поняття, переглянути попередні додані тлумачення і, за потреби, оновити;
7. додати приклади вживання терміна;
8. обрати спосіб представлення словника через людино-машинний інтерфейс, реалізувати перенесення словника з табличного варіанту у підтримуваний програмою формат.

3.3 Відбір термінів та особливості процесу перекладу термінів

Процес відбирання термінів відбувався за процедурою, виробленою спеціально для цієї роботи: після опрацювання кожного розділу в таблицю були внесені лексеми та словосполучення, що виділені синім кольором та винесені вправо від основного тексту в підручнику Д. Джуравського, приклад такого запису подано на рис. 1:

pretraining What made the modern NLP revolution possible is that large language models can learn all this knowledge of language, context, and the world simply by being taught to predict the next word, again and again, based on context, in a (very) large corpus of text. In this chapter and the next we formalize this idea that we'll call **pretraining**—learning knowledge about language and the world from iteratively predicting tokens in vast amounts of text—and call the resulting pretrained models **large language models**. Large language models exhibit remarkable performance on natural language tasks because of the knowledge they learn in pretraining.

*Рис. 1 Приклад організації винесення термінів у підручнику
“Опрацювання природної мови і мовлення”*

Автори наголошують, що спеціально виділяють нові для відповідного розділу терміни, на які варто звернути увагу. Деякі слова, що виділені жирним шрифтом, також були внесені у таблицю. Це обґрунтовується бажанням авторів звернути особливу увагу на цих поняттях, навіть якщо вони вже зустрічались раніше в підручнику. Зазвичай такі терміни є більш незрозумілими, або вузькоспеціалізованими, бо не належать до ключових саме для опрацьованого розділу, але для розуміння матеріалу вони необхідні.

Увесь список реєстрових одиниць за алфавітним порядком, який можна використовувати як машиночитний варіант словника, продемонстровано в додатку 1.

Підбір перекладу для термінів виявився загалом доволі складним завданням. Як описано в попередньому розділі, українські спеціалісти часто послуговуються переважно англійськомовною термінологією, не перекладаючи її українською або використовуючи калькування. Тому надійних джерел із нормативним перекладом цих термінів зараз досить мало. В основі вибору

перекладу лежить метод аналізу різних лексикографічних джерел, їхніх словникових статей, а також наукових робіт та статей з комп'ютерної лінгвістики. У ході роботи ми здійснювали пошук терміна англійською у словниках та виконували порівняльний аналіз. Пошук ґрунтувався на кількох словниках, щоб порівняти результати та обрати оптимальний варіант. Остаточним варіантом ставав найдоцільніший переклад, або, як у більшості випадків, єдиний наявний варіант. Для деяких термінів не було знайдено кодифікованого відповідника українською взагалі. Тоді процес перекладу відбувався за допомогою наукових праць, що пояснювали певне явище детальніше. Таке порівняння описів дає змогу чіткіше представити суть поняття і шукати найбільш підхожий переклад із загальної або спеціалізованої лексики суміжних галузей. Якщо термін використовують вкрай рідко, і наявного матеріалу недостатньо для його коректного перекладу, тоді позиція “українська” у словнику, що позначає переклад англійського терміна українською, залишається порожньою, а саме терміни *teacher forcing*, *in-context learning*, *zero-shot*, *few-shot*. Добір перекладу для однослівних термінів не викликав труднощів, адже більшість мали свої відповідники українською, проте дво- та багатослівні терміни майже не зафіксовані в аналізованих джерелах. У таких випадках вдавалися до перекладу окремих компонентів терміносполучення та стратегій перекладу, що були описані у розділі 2, до прикладу: *text summarization* – *реферування тексту*, *автоматичне реферування тексту* (описовий переклад, синонімічний переклад), *continued pre-training (CPT)* – *систематичне / тривале попереднє тренування* (описовий переклад), *data augmentation* – *аугментація / збагачення даних* (синонімічний переклад).

На етапі наповнення словника у табличному вигляді був створений список джерел для пошуку українського відповідника для певного терміна. Здебільшого це словники, тезауруси, рідше – наукові статті. Повний список ресурсів, з яких добирали переклади, наведений нижче:

- “Короткий англо-український словник з комп'ютерної лінгвістики”
В. Коломієць (2022);

- “Англійсько-український словник з математики та інформатики” Є. Мейнаровича, М. Кратка (2010);
- “Українсько-англійський словник з радіоелектроніки” Б. Рицара, Л. Сніцарука, Р. Мисака (2018);
- “Англійсько-українсько-англійський словник наукової мови (фізика та споріднені науки). Частина II українсько-англійська” О. Кочерги, Є. Мейнаровича (2019);
- Англійсько-український словник “Кембридж” (Cambridge Dictionary);
- “Українсько-англійський словник лінгвістичної термінології” Л. Коломієць, О. Паламарчук, Г. Стрельчук, М. Шевченко (2013);
- словник термінів у сфері ШІ (2024);
- стаття “Локалізація у перекладі: адаптація для нових аудиторій” Л. Лікарчук (2025);
- стаття “Дослідження сучасних методів аугментації текстових даних” Є. Абросімова (2024).

Також в словнику є терміни, що були перекладені самостійно за допомогою декількох ресурсів, такі як *null hypothesis*, *thematic grid*, *period disambiguation* тощо.

3.4 Робота над дефініціями

У цій роботі значна кількість часу була присвячена тлумаченню – опису та поясненню отриманих з опрацьованого підручника термінів. Для цього був застосований порівняльний аналіз лексикографічних та інших наукових джерел. У процесі ми уклали список доступних тлумачних словників для українських та англійських термінів комп’ютерної лінгвістики, що зазначені у попередньому підрозділі, а також загальної лінгвістики, математики та програмування. Деколи інформації, отриманої зі сформованого списку літератури, недостатньо, або вона недостовірна чи неактуальна, тоді тлумачення чи пояснення взяті із підручника

Д. Джуравського та Дж. Мартіна, наприклад: *розмічування семантичних ролей, визначення семантичних ролей – завдання автоматичного присвоєння семантичних ролей кожному елементу в реченні*. Такий спосіб дефінування термінів є можливий через те, що в розділі підручника автори словесно подають опис поняття, яке з'являється у тексті, наводять приклади функціонування. А оскільки мова підручника – англійська, у нашому словнику подано цей опис як дефініцію, але в українському перекладі. Якщо з різних причин підручник подавав невідповідне пояснення, відбувався пошук подібного тлумачення в англійських наукових статтях за темою описуваного або дотичного до нього поняття. Під час проведення розвідок щодо тлумачень українських джерел було виявлено небагато, тож більша частина тлумачень становить переклад з англійської, але з авторитетної наукової літератури.

Варто звернути увагу на таку групу термінів, як загальні лінгвістичні. Вона охоплює одиниці “прямий додаток”, “непрямий додаток”, “прикладка”, “сурядність”, “детермінатив” тощо. Для визначення дефініцій були вибрані пояснення, подані проєктом “Універсальні залежності” (2021), адже саме на нього посилається Д. Джуравський та Дж. Мартін у своєму підручнику у частині, де описується про опис цих явищ у прикладному аспекті, а саме частиномовна розмітка, що використовується проєктом. Ще однією причиною є факт відсутності в українських словниках загальних мовознавчих понять дефініцій для термінів, що не притаманні українській мові.

З українських видань найбільш корисними виявилися посібник Н. Дарчук “Ідеографічні словники та тезаурусотворення” (2023), підручник Є. Карпіловської “Вступ до комп'ютерної лінгвістики” (2006). У кінці цих праць є тлумачні словники термінів, серед яких трапляються і терміни комп'ютерної лінгвістики, що значно полегшило нашу роботу над дефініціями. У пригоді також став і “Сучасний лінгвістичний словник” А. Загнітка (2020), де чітко витлумачено багато основних мовознавчих понять.

Для заповнення словникової зони “тлумачення” було створено список ресурсів:

- “Тлумачний словник основних понять і термінів програмування” І. Шевчука (2013);
- тлумачний словник англійської мови “Меріам-Вебстер” (Meriam-Webster.com);
- “Англо-український словник термінів з корпусної лінгвістики” І. Чупири (2024);
- “Словник синтаксичних термінів і понять” Д. Решетнікова;
- словник термінів у сфері ІІІ (Чумаченко та ін., 2024);
- “Сучасний лінгвістичний словник” А. Загнітка (2020);
- словник англійської мови “Кембридж” (Cambridge Dictionary);
- підручник “Сучасна українська мова: морфологія” (Алексієнко, Зубань, Козленко, 2013);
- посібник “Теорія паралельних обчислень. Частина ІІ” В. Коцовського (2019);
- підручник “Загальне мовознавство” (Кочерган, 2006);
- посібник “Ідеографічні словники та тезаурусотворення” Н. Дарчук (2023);
- проєкт “Універсальні залежності” (2021);
- посібник “Практикум з машинного перекладу: навчальний посібник” Ю. Сидоренка (2017);
- низка статей [Абросімов, 2024; Бар, 2024; Бондаренко 2021; Лікарчук, 2025; Singh та ін., 2025; Pinaya та ін., 2023; Varughese, 2026];
- підручник “Опрацювання природної мови та мовлення” Д. Джуравського та Дж. Мартіна (2024).

У словнику наявні також статті з незаповненим полем “тлумачення”, всього їх 5. Це зумовлено тим, що вдалого тлумачення не знайшлося, а спроба сформулювати його самостійно була невдалою. Тому цю зону словникової статті залишено порожньою з можливістю подальшого пошуку відповідного тлумачення і доповнення словника.

3.5 Приклади до словникових статей

У словниковій статті також є ілюстративна зона – з прикладами вживання кожної лексеми в контексті. Попередньо ми планували подати приклади вживання і англійською, і українською мовами для кращої репрезентативності функціонування терміна. Утім, сучасних матеріалів українською наразі недостатньо, щоб повною мірою представити термін. Через це у кінцевому варіанті словника наявний лише приклад англійською.

Ілюстрації були взяті із двох джерел – підручника Д. Джуравського і корпусу текстів Кембриджського словника (Cambridge Dictionary). Вибір цих джерел обґрунтовуємо їх активним оновленням, широким використанням і схваленням спільнотою дослідників, що користуються цими матеріалами.

Важливою умовою в підборі прикладів для нас було те, щоб речення не повторювали дефініції, але також були зрозумілими для користувача, репрезентували термін чи терміносполуку в достатньому контексті і стосувалися теми, в якій фігурує такий термін. Часто це ставало перешкодою, адже деякі непопулярні терміни траплялись у тексті лише кілька разів, де одне з уживань було власне тлумаченням чи поясненням терміна. Тому виникла потреба у залученні інших праць. Наприклад, з корпусу текстів Кембриджського словника ми вибирали приклади до загальнолінгвістичних термінів та найбільш популярних термінів у сфері ІІІ.

3.6 Представлення словника в “Lexopomy”

Модель словника у вигляді таблиці є практичною у створенні розмітки, швидкого оновлення структури словника та словникової статті. Але таке представлення не є зручним у користуванні. Саме тому було вирішено перенести базу даних у інший формат з людино-машинним інтерфейсом. Серед декількох

варіантів найоптимальнішим і швидким було розміщення словника на платформі “Lexonomy” (портал Lexonomy.eu).

“Lexonomy” (“Лексономі”) – це система для створення та представлення онлайн-словників, глосаріїв, списків та інших подібних баз даних. Вона працює на базі корпусного алгоритму Sketch Engine. Цей інструмент є мовно незалежним і підходить як для великих за обсягом, так і малих лексикографічних проєктів. Функціонал дозволяє авторам створювати одно-, дво- та багатомовні словники, користуватися готовими шаблонами словників, але також і створювати власну макро- та мікроструктуру, що не обмежується лише текстовою інформацією про будь-яку характеристику слова (етимологія, частина мови, діалект, фонетична транскрипція тощо), а й дає змогу додати посилання на інші ресурси мережі Інтернет та будь-який тип мультимедіа (фото, відео, звук, анімації). Не менш важливою особливістю “Lexonomy” є редагування стилю інтерфейсу словника. На сайті є модуль для редагування розташування елементів всередині словникової статті, кольору, розміру та шрифту тексту.

Сам словник має широку систему пошукових запитів, які можна здійснювати як за заголовковим словом статті, так і за будь-якою частиною статті. Це дозволяє розширити межі взаємодії зі словником. Всередині словник можна розмістити в будь-якому порядку – хаотичному, алфавітному чи зворотному, така функція також доступна і для користувача. Під час роботи зі словником через інтерфейс користувача програма буде пропонувати випадкові статті для перегляду, а також відображає пошуковий рядок. Статті можна переглянути як окремо від інших, так і разом, включно із повним списком реєстрових одиниць. Наш словник розташований в алфавітному порядку й автоматично пронумерований програмою від 1 до 124.

Окрім того, що словник можна створити та редагувати у ручному керуванні всередині програми, словник можна імпортувати через формати XML та NVH. Останній, Name-Value Hierarchy, що перекладається як “ієрархія ім’я-значення”, – це формат даних, що використовує “Lexonomy” як основний. Це проста мова розмітки, спеціально створена для кодування словникових статей в

лексикографії. NVH є менш деталізованою, коротшою альтернативою XML. Водночас там не використовується багато спеціальних символів, тому його легко читати і людині, і машині. Файл NVH має невеликий розмір. Структура NVH є вбудованою в редактор програми, але інша опція створення словника – імпорт словника у файлі NVH.

У цій роботі було вирішено перетворити словник з табличного варіанту (файлу .xlsx) (розміщено в додатку 2) у файл .nvh, що у своїй суті є простим текстовим файлом, де словникові статті були розміщені у вигляді певної ієрархії. Потім цей файл вирішено занести у “Lexopomy”, щоб програма на основі отриманої розмітки побудувала список реєстрових слів, структуру словникової статті та внесла дані усередину словника.

Приклад вигляду словникової статті слова *парсер*, записаної розміткою NVH наведено нижче:

```
entry: parser
  translation: (синтаксичний) аналізатор, парсер
  definition: алгоритм, що виконує автоматичний синтаксичний аналіз
  example: The parser walks through the sentence left-to-right, successively shifting items from the buffer onto the stack.
  label: 19 - Парсинг залежностей
```

Як можна побачити із прикладу, центральним елементом завжди є заголовкове слово, йому підпорядковуються такі елементи статті, як переклад, дефініція та тема, а дефініції підпорядковується приклад використання слова в цьому значенні. Також на Рис. 2 продемонстровано ту ж структуру, але у внутрішньому редакторі програми:

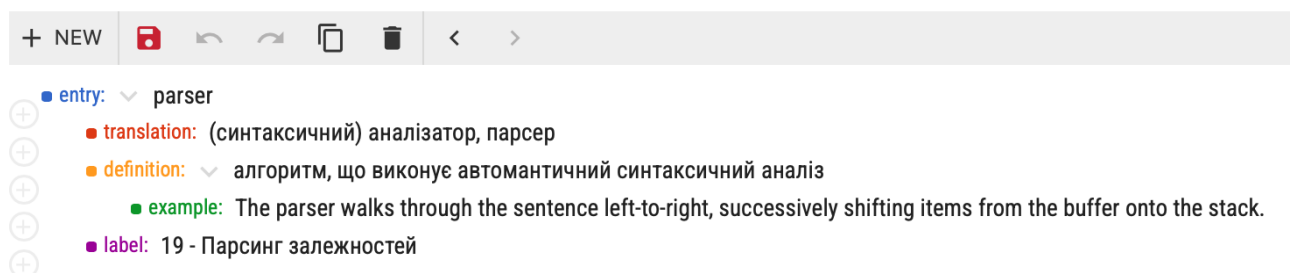


Рис. 2 Словникова стаття слова “парсер” і її структура мовою розмітки NVH у програмі “Lexopomy”.

Після імпорту файлу у програму, програма автоматично розташувала терміни в алфавітному порядку і пронумерувала усі статті.

Наступним етапом роботи над словником було створення стилю інтерфейсу. Було обрано простий, базовий шрифт, синій колір для виділення слова англійською та відступи для виділення перекладу й прикладу. Зрозумілий для роботи і пошуку стиль інтерфейсу має забезпечити зручне користування словником для будь-кого. Також важливою частиною словника є його опис, розташований справа від пошукової частини, який містить інформацію про словник. Вигляд головної сторінки словника разом зі словниковою статтею показано на Рис. 3.

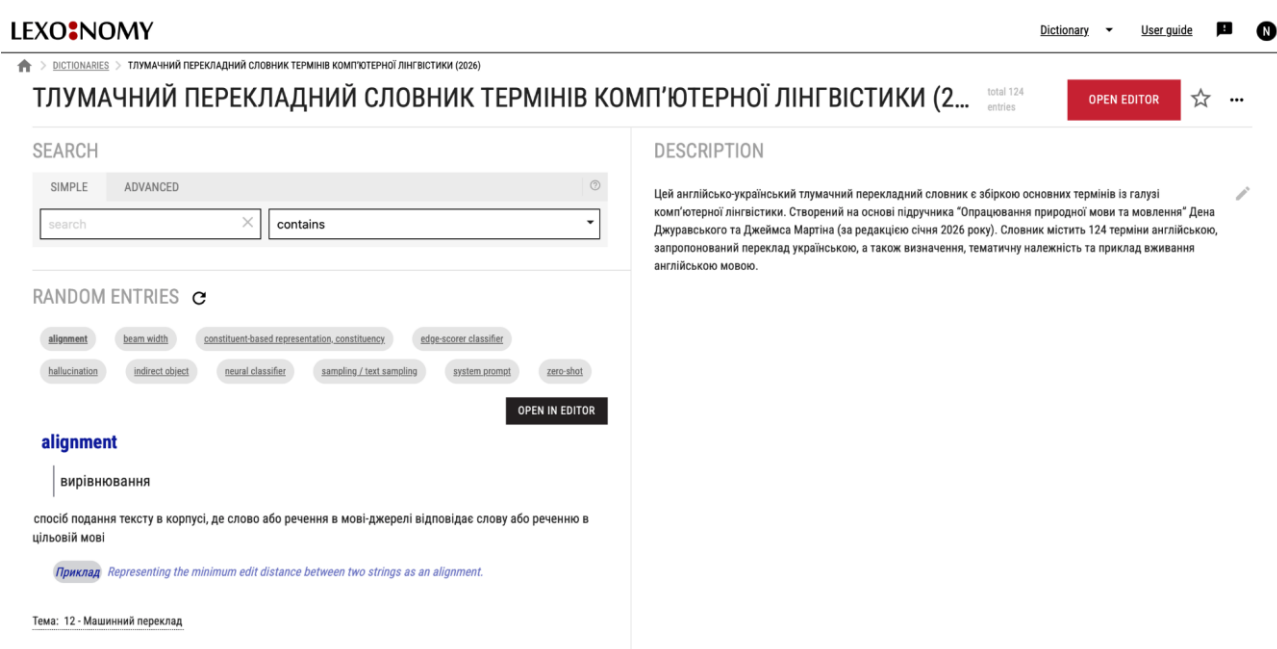


Рис. 3 Інтерфейс словника термінів комп'ютерної лінгвістики на платформі "Lexonomy"

Основна частина роботи над словником завершена, але його особливість полягає в тому, що він завжди відкритий до доповнень, оновлень та удосконалень. Кілька статей досі не мають прикладу чи перекладу, а також ми не могли охопити абсолютно всі терміни комп'ютерної лінгвістики. Редагування словника доволі просте через легке і зрозуміле представлення розміткою NVH, тож його оновлення буде і надалі актуальним завданням.

Робоче посилання на словник термінів комп'ютерної лінгвістики знаходиться в Додатку 3.

Висновки до розділу 3

Джерелом для доповнення та оновлення словника стали 6 різних розділів підручника “Опрацювання природної мови і мовлення”, з яких було вибрано, на наш погляд, найактуальнішу термінологію, яка може бути корисна для розуміння тем, або ж складна для перекладу чи тлумачення українською. Лексику з цих розділів було повністю опрацьовано та перевірено. Наразі загальна кількість реєстрових одиниць у словнику складає 124.

Станом на зараз майже для кожного слова підібраний оптимальний переклад на основі порівняння реєстру словників, наукових праць і власних пропозицій.

Найбільше ресурсів було витрачено на підбір тлумачення, що достатньо зрозуміло і точно пояснює суть поняття. У процесі ми зверталися до 20 словників та інших праць, розробляли власні тлумачення на основі описів із підручника Д. Джуравського. Також було підібрано по одному прикладу вживання терміна чи терміносполуки в контексті англійською мовою із підручника Д. Джуравського та корпусу Кембриджського словника.

Перенесення словника з табличної моделі у вигляд із робочим інтерфейсом став ключовим на останніх етапах роботи. Було обрано платформу “Lexonomy”, де можна представити свою лексикографічну працю, розробивши власну структуру та стиль словника. Програма містить усі потрібні засоби для редагування словника вручну, але ми обрали шлях автоматичного генерування структури та заповнення інформацією на основі імпорту словника у форматі NVN, який отримали перетворивши таблицю xlsx за допомогою алгоритму конвертації.

Словник внесено повністю, відповідно до розробленого плану. Розташований в алфавітному порядку, він має простий стиль та легку для редагування ієрархію елементів. Макроструктура словника охоплює сам реєстр

та опис, мікроструктура репрезентована зонами реєстрового (заголовкового) слова, українськомовним відповідником, тлумаченням, ілюстрацією та відомостями про тему підручника, яку представляє термін. Наразі роботу над словником завершено, але проєкт буде відкритий для доповнення та оновлення відповідно до розвитку сфери комп'ютерної лінгвістики.

ВИСНОВКИ

Кваліфікаційна робота присвячена докладній роботі над теоретичними та практичними завданнями створення електронного словника термінів комп'ютерної лінгвістики. Було акцентовано на основах термінології, термінографії та лексикографії, особливостях будови терміносистеми комп'ютерної лінгвістики української мови, перекладі термінів з англійської на українську, їх тлумаченні, специфіці кодифікації такої лексики та засобах створення електронних словників.

Термін є центральною одиницею терміносистеми будь-якої наукової галузі. Йому властива низка характерних ознак: однозначність, системність та термінологічна усталеність, тобто ступінь активності вживання слова або словосполучення у відповідній сфері.

Український термін-відповідник іншомовній одиниці має бути точним, пов'язаним із поняттям та передавати його головні ознаки. Способи перекладу термінів потребують знань та загального розуміння досліджуваної сфери, у цій роботі – комп'ютерної лінгвістики та суміжних сфер. У процесі дослідження було розглянуто більшість відомих стратегій перекладу термінів з англійської мови на українську. Аналіз досліджуваної терміносистеми засвідчив переважання таких перекладацьких прийомів, як калькування, описовий метод, транскрибування та генералізація.

Також приділено увага особливостям дефініювання реєстрових одиниць, з'ясовано, що дефініція (визначення, тлумачення) – це інформація про мовний знак та поняття, що описується, неодмінний компонент мікроструктури такого типу словників. Вона має бути точною, достатньою та зрозумілою. Ці ознаки слід враховувати при підборі або створенні визначення. Забезпечення якісного та зрозумілого тлумачення для терміна не лише полегшує вивчення нового в обраній науковій галузі, а й пояснює основні ознаки названого терміном явища, водночас поєднуючи його з іншими складниками терміносистеми.

Практична частина роботи передбачала два етапи – лінгвістичне дослідження функціонування термінів комп'ютерної лінгвістики в українських

текстах, з'ясування та опис актуальних проблем під час перекладу і тлумачення термінів, а також моделювання та укладання словника термінів комп'ютерної лінгвістики і представлення його електронному вигляді з користувацьким інтерфейсом.

Актуальною проблемою у використанні термінів комп'ютерної лінгвістики є як надмірне калькування, так і непослідовне використання українських та англійських варіантів. Така тенденція спонукала нас до пошуку авторитетних розвідок з теми, аналізу галузевих словників та намагання унормувати цю термінологію, створивши тлумачно-перекладний словник.

Окрім базових праць із комп'ютерної лінгвістики, були опрацьовані і загальнолінгвістичні видання. Ми не оминули увагою і сучасних лексикографічних праць, активно використовували їх як надійне джерело для свого словника. Аналіз джерел вплинув і на вибір актуальних термінів, і на переклад та тлумачення.

Головним джерелом термінів нашого словника є 6 розділів підручника Д. Джуравського та Дж. Мартіна. Було опрацьовано 124 терміни з тем “Парсинг залежностей”, “Машинний переклад”, “Великі мовні моделі”, “Визначення семантичних ролей”, “Логістична регресія і завдання класифікації тексту”, “Приховані моделі Маркова”.

Для усіх термінів був підібраний переклад, дефініція та приклад. На всіх етапах робота опиралась як на порівняння словникових статей лексикографічних праць, так і на власний досвід. Найбільшими практичними проблемами стали відсутність можливих перекладів термінів українською мовою через активне вживання англізму та малої кількості української літератури на відповідну тему.

Спочатку словник було змодельовано у вигляді таблиці, де колонки відображали структуру словникової статті, а рядки – наповнення словника. Після того словник було опубліковано на платформі “Lexonomy”, що дає можливість легко імпортувати таблиці у формат Name-Value Hierarchy. Такий формат використовується для створення будь-яких публікацій на платформі. Його особливості – простота, машиночитність, розташування статей у вигляді ієрархії

елементів. До словника також створено власний графічний інтерфейс у віконному редакторі сайту та подано короткий опис. Зараз словник доступний для публічного перегляду.

Укладений словник термінів комп'ютерної лінгвістики є важливим інструментом для систематизації та кодифікації галузевої термінології української мови. Його практична цінність полягає у забезпеченні дослідників, перекладачів і студентів надійною лексикографічною базою в умовах стрімкого розвитку галузі.

Словник має багато перспектив для удосконалення – поповнення новими термінами, додавання тлумачення та перекладу, українських прикладів уживання, а також є можливість встановити семантичні зв'язки між термінами та показати, як в межах теми терміни пов'язані між собою. З огляду на невідомий розвиток галузі комп'ютерної лінгвістики, створений словник зберігатиме свою практичну цінність, а його подальше оновлення та розширення залишається перспективним і необхідним завданням.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. @cybercalm. Слосквоттинг: як помилки ШІ створюють нові можливості для кіберзлочинців. *Друкарня від WE.UA. Cybercalm.* 22.12.2025. URL: <https://drukarnia.com.ua/articles/slopskvotting-yak-pomilki-shi-stvoryuyut-novi-mozhlivosti-dlya-kiberzlochinciv-nyVxp> (дата звернення: 26.05.2026).
2. @cybercalm. Як ШІ перетворився на машину дезінформації – дослідження НАТО. *Друкарня від WE.UA. Cybercalm.* 16.04.2026. URL: <https://drukarnia.com.ua/articles/yak-shi-peretvorivsia-na-mashinu-dezinformaciyi-doslidzhennya-nato-JZ8cs> (дата звернення: 26.05.2026).
3. @future.simple. «Act as an Expert»: чому твій промптинг – це карго-культ та як це виправити?. *Друкарня від WE.UA.* 06.08.2025. URL: <https://drukarnia.com.ua/articles/act-as-an-expert-chomu-tvii-prompting-ce-kargo-kult-ta-yak-ce-vipraviti-FMb5y> (дата звернення: 26.05.2026).
4. @kotovich. Як перестати гратися з ШІ та почати отримувати від нього реальну користь у 2026 році. *Друкарня від WE.UA. Юра. ШІ для людей і бізнесу..* 21.05.2026. URL: <https://drukarnia.com.ua/articles/yak-perestati-gratisya-z-shi-ta-pochati-otrimuvati-vid-nogo-realnu-korist-u-2026-roci-EuvLH> (дата звернення: 26.05.2026).
5. @MaksymValin. ChatGPT Health, уразливості безпеки та медичні дані: чому це має нас хвилювати. *Друкарня від WE.UA.* 08.01.2026. URL: <https://drukarnia.com.ua/articles/chatgpt-health-urazlivosti-bezpeki-ta-medichni-dani-chomu-ce-maye-nas-khvilyuvati-Y5RvQ> (дата звернення: 26.05.2026).
6. @ukr_businessman. Як бізнесу отримувати органічний трафік від ChatGPT. *Друкарня від WE.UA.* 13.08.2025. URL: <https://drukarnia.com.ua/articles/yak-biznesu-otrimuvati-organichniy-trafik-vid-chatgpt-M7TMw> (дата звернення: 26.05.2026).
7. @Vikt0ria. AI Studio - ChatGPT 🏛️ Переваги та недоліки 💰 Ціноутворення та доступність. *Друкарня від WE.UA.* 02.11.2025. URL:

<https://drukarnia.com.ua/articles/ai-studio-chatgpt-perevagi-ta-nedoliki-cinoutvorenniya-ta-dostupnist-OE0Fr> (дата звернення: 26.05.2026).

8. Гогілашвілі Є. Нові професії – для нових технологій. Чим займається Prompt Engineer. *High Bar Journal*. 14.06.2024. URL: <https://journal.gen.tech/post/chym-zaymayetsya-prompt-engineer> (дата звернення: 26.05.2026).
9. Абросімов Є. О., Дейнеко А. О. Дослідження сучасних методів аугментації текстових даних. *Радіoeлектроніка та молодь у ХХІ столітті: матеріали 28-го Міжнар. Молодіж. Форуму*, Харків, Україна, 16 квіт. 2024. ХНУРЕ, 2024. С. 35–36.
10. Алексієнко Л., Зубань О., Козленко І. Сучасна українська мова: Морфологія / ред. А. Мойсієнко; Київ. нац. ун-т ім. Т. Шевченка. Київ: Знання, 2013.
11. Бар Б. Синтаксичні структури - Вступ до лінгвістики. *Fiveable*. 2024., URL: <https://library.fiveable.me/key-terms/introduction-linguistics/syntactic-structure>. (дата звернення: 26.05.2026)
12. Белова В. Дефініція як основний засіб семантизації терміна. *Збірник наукових праць студентів та молодих учених (За матеріалами II Всеукраїнської студентської науково-практичної Інтернет-конференції «Сучасні філологічні дослідження: традиції та інновації»*. Умань: Візаві, 2020. С. 57–60.
13. Білодід І. К. та ін. (ред.) Словник української мови. Київ : Наукова думка., 1970-1980.
14. Бондаренко О. Матриця невідповідностей. *Medium.com*. 31.10.2021. URL: <https://medium.com/stinopys/матриця-невідповідностей-329e7e4bf05e> (дата звернення: 26.05.2026).
15. Бондаренко О. Словник NLP. *Medium.com*. 10.10.2021. URL: <https://medium.com/stinopys/словник-nlp-b0fab1027551> (дата звернення: 26.05.2026).

- 16.Вакуленко М. О. Сучасна українська термінологія: методологія, кодифікація, лексикографічна практика : дис. ... д-ра філол. наук : 10.02.01 / Київський національний університет імені Тараса Шевченка. Київ, 2023. 445 с.
- 17.Венцковська Ю. 7 CAT-програм, які полегшать життя перекладачу. *MK-translations*. 27.07.2022. URL: <https://mk-translations.ua/blog/7-cat-program-yaki-polegshat-zhittya-perekladachu/> (дата звернення: 26.05.2026).
- 18.Вихованець І. Частини мови в семантико-граматичному аспекті. Київ: АН УРСР, Ін-т мовознавства ім. О. О. Потебні., 1988. 255 с.
- 19.Дарчук Н. П., Сорокін В. М. Термін в лінгвістичній інформації. Київ: Видавничо-полігр. центр "Київ. ун-т", 2012. 143 с.
- 20.Дарчук Н.П. Ідеографічні словники та тезаурусотворення: навчальний посібник (електронне видання). К., 2023. 136 с.
- 21.Дерба С. Словник української термінології прикладної (комп'ютерної) лінгвістики. Київ, 2007. 350 с.
- 22.Енциклопедія кібернетики: [у 2 т.] / редкол.: В. М. Глушков (відп. ред) [та ін.]; АН Української РСР. Київ: Голов. ред. Укр, рад. енцикл., 1973.
- 23.Желязкова В. В. Семантика: теорія і практика: навч.-метод. посіб. для студ. спец. 035 Філологія («Прикладна лінгвістика»). Миколаїв: Іліон, 2018. С. 76-84.
- 24.Загнітко А. П. Теорії сучасних лінгвістичних вчень: навчальний посібник. Вінниця: ДонНУ імені Василя Стуса, 2019. 528 с.
- 25.Загнітко А. Сучасний лінгвістичний словник. Вінниця: -ТВОРИ, 2020.
- 26.Карпіловська Є. Вступ до прикладної лінгвістики: комп'ютерна лінгвістика. Донецьк: Юго-Восток, 2006. 188 с.
- 27.Клименко Н. Ф., Карпіловська Є. А., Кислюк Л. П. Динамічні процеси в сучасному українському лексиконі. Київ: Видавничий дім Дмитра Бураго, 2008. 335 с.
- 28.Коваленко А. Загальний курс науково-технічного перекладу: Навчальний посібник. К.: Інкос, 2001. 290 с.

29. Коломієць В. О. Короткий англо-український словник з комп'ютерної лінгвістики. Київ: Вид. центр КНЛУ, 2022. 101 с.
30. Коцовський В. М. Теорія паралельних обчислень. Частина II: Методичний посібник. Ужгород: Видавництво УжНУ "Говерла", 2019. 52 с.
31. Кочерга О., Менайрович Є. Англійсько-українсько-англійський словник наукової мови (фізика та споріднені науки). Частина II. URL: <https://e2u.org.ua>.
32. Кочерган М.П. Загальне мовознавство. Вид-во: Академія, 2006. 456 с.
33. Лікарчук Л. І. Локалізація у перекладі: адаптація для нових аудиторій. *Вчені записки ТНУ імені В. І. Вернадського. Серія: Філологія. Журналістика*. 2025. т. 36. с. 310–315. URL: <https://doi.org/10.32782/2710-4656/2025.1.1/50>.
34. Мейнарович Є., Кратко М. Англійсько-український словник з математики та інформатики. 2010. URL: <https://e2u.org.ua/dicts/meinarovych> (дата звернення: 26.05.2026).
35. Приплюцька Ю., Фаріон І. Д. Англізми в сучасній українській мові. *Українська мова в просторі й часі: матеріали Всеукраїнської студентської науково-практичної конференції*, Львів, 17–19 квіт. 2019. Львів: Видавництво Львівської політехніки, 2019. 169 с.
36. Решетніков Д. Словник синтаксичних термінів і понять. *Lexonomy.eu*. URL: <https://www.lexonomy.eu/#syntaksys> (дата звернення: 26.05.2026).
37. Річардсон А., Огден Ч. The Meaning of Meaning. Harvest/HBJ, 1989.
38. Сидоренко Ю. Практикум з машинного перекладу: навчальний посібник. Миколаїв, 2017.
39. Сірук О., Сорокін В. Тезаурус з комп'ютерної лексикографії. *Лінгвістичний портал MOVA.info*. URL: <http://www.mova.info/Page3.aspx?l1=188&vocid=1> (дата звернення: 26.05.2026).
40. Словник Меріам-Вебстер. *Merriam-Webster.com*. URL: <https://www.merriam-webster.com> (дата звернення: 26.05.2026).

- 41.Словник термінів у сфері штучного інтелекту / упорядники: Чумаченко Д., Мішкін Д., Андрієнко О., Краковецький О., Турута О., Дубно О., Хрущова Д., Кобрін А., Авдєєва Т., Кравець І., Герасимyak В., Шабанов О., Бистрицька А. Київ: Міністерство цифрової трансформації України, 2024. 37 с.
- 42.Старк С. Словник термінів ШІ: що таке AGI, LLM та ШІ-агенти простими словами. *Inweb Media*. 27.05.2025. URL: <https://theinweb.media/slovnyk-ai-terminiv-inweb/> (дата звернення: 26.05.2026).
- 43.Тезаурус з лінгвістичної термінології. *Лінгвістичний портал MOVA.info*. URL: http://www.mova.info/mov_thes.aspx?l1=68 (дата звернення: 26.05.2026).
- 44.Українсько-англійський словник з радіоелектроніки. Ukrainian-English Dictionary of Electronics [Електронний ресурс] / Богдан Рицар, Леонід Сніцарук, Роман Мисак; під ред. Богдана Рицара, Василя Старка. 2-ге вид., оновл. 2018. (Термінографічна серія СловоСвіт ; № 16). URL: <https://e2u.org.ua>.
- 45.Українсько-англійський словник лінгвістичної термінології / Л. Коломієць та ін. Київ: Освіта України, 2013. 455 с. URL: <https://e2u.org.ua/dicts/linguistic> (дата звернення: 26.05.2026).
- 46.Фурт Д. В., Дмитрук Л. А. Термінологія. Навч. посібник. Кривий Ріг: Донец. нац. ун-т економіки і торгівлі ім. М. Туган-Баран., каф. інозем. філології, українознавства та соц.-прав. дисциплін, 2020. 172 с.
- 47.Чупира І. Англо-український словник термінів з корпусної лінгвістики. *Lexonomy.eu*. URL: <https://www.lexonomy.eu/#w3w9btn4> (дата звернення: 26.05.2026).
- 48.Шевчук І.Б. Тлумачний словник основних понять і термінів програмування. ЛДФА, Львів: Видавництво ВТЗНВ, 2013. 45 с.
- 49.Cambridge English–Ukrainian dictionary: translate from English to Ukrainian. *Cambridge Dictionary | English Dictionary, Translations & Thesaurus*.

- URL:<https://dictionary.cambridge.org/dictionary/english-ukrainian/> (дата звернення: 26.05.2026).
- 50.De Marneffe M.-K., Manning C., Nivre J. Zeman D. Universal Dependencies. *Computational Linguistics* 47(2), 2021. с. 255–308.
- 51.Frege G. On Sense and Reference. *Duke University Press*. 1892. с. 209–230.
- 52.Jurafsky D., Martin J. Speech and Language Processing. 2024. URL: <https://web.stanford.edu/~jurafsky/slp3/> (дата звернення: 26.05.2026)
- 53.Лехonomy: онлайн-платформа для написання та публікації словників. URL: <https://lexonomy.eu/>
- 54.Pinaya W. H. L. et al. Generative AI for Medical Imaging: extending the MONAI Framework. London: King's College London et. al., 2023.
- 55.Singh P. A Comprehensive Guide to Pre-training LLMs. *Analytics Vidhya. LLMs*. 12.02.2025. URL: <https://www.analyticsvidhya.com/blog/2025/02/llm-pre-training/#h-what-is-the-llm-pre-training> (дата звернення: 26.05.2026).
- 56.Ullmann S. The principles of semantics. Glasgow: Glasgow University Publications, 1951.
- 57.Varughese J. What is in-context learning?. *IBM. Prompt Engineering Guide*. URL: <https://www.ibm.com/think/topics/in-context-learning> (дата звернення: 26.05.2026).

ДОДАТКИ

Додаток 1. Список усіх реєстрових одиниць словника у алфавітному порядку

1. adjectival modifier-прикметникове означення
2. adverb-прислівник
3. alignment-вирівнювання
4. appositional modifier-прикладка
5. arc standard parsing-синтаксичний аналіз за принципом стандартних дуг
6. arc-eager transition system-дуго-спрямована система переходів
7. argument-аргумент
8. authorship attribution-авторизація
9. autoregressive generation-авторегресивне генерування
- 10.backtranslation-зворотний переклад
- 11.beam search-променевий пошук
- 12.beam width-ширина променя
- 13.biaffine algorithm-біафінний алгоритм
- 14.CAT (computer-aided translation)-автоматизований переклад, переклад за допомогою комп'ютера (часто залишають CAT)
- 15.classifier-класифікатор
- 16.clausal component-компонент елементарного речення
- 17.clausal relation-синтаксичні відношення
- 18.concatenation-нанизування, конкатенація
- 19.conditional generation-умовне генерування
- 20.configuration-конфігурація
- 21.confusion matrix-матриця невідповідностей
- 22.conjunction-сполучник, сполучне слово, паратакис, сурядність
- 23.constituent-based representation, constituency-репрезентація у вигляді складників
- 24.context-free grammar-контекстно-вільна граматика

- 25.continued pre-training (CPT)-систематичне / тривале попереднє тренування
- 26.coordinating conjunction-сурядний сполучник
- 27.corpus-корпус
- 28.cross-entropy loss / log loss-втрата, похибка перехресної ентропії;
логарифмічний декремент
- 29.data augmentation-аугментація даних, збагачення даних
- 30.decoding-декодування
- 31.dependency grammar-граматика залежностей
- 32.dependency parse-синтаксичний розбір за принципом граматики
залежностей
- 33.dependency parser-синтаксичний аналізатор на основі граматики
залежностей
- 34.dependency tree-дерево залежностей
- 35.dependent element of a dependency parsed pair-залежний елемент, тут:
залежний елемент розібраної за принципом граматики залежностей пари
слів
- 36.determiner-визначальне слово, детермінатив
- 37.direct object-прямий додаток
- 38.directed binary grammatical relation-орієнтований бінарний/двійковий
граматичний зв'язок
- 39.edge-factored-факторизований за ребром графа
- 40.edge-scorer classifier-класифікатор з лічильником ребер графа
- 41.embedding-вбудовування, векторне представлення елементів, ембединг
- 42.feature template-характеристичний шаблон, шаблон характеристик
- 43.few-shot-?
- 44.finetuning-дотренування
- 45.foundation model-модель-основа
- 46.free word order-вільний порядок слів
- 47.generative artificial intelligence (AI)-генеративний штучний інтелект
- 48.gold label, gold standard-золотий стандарт, еталон

- 49.grammar formalism-формалізм граматики
- 50.grammatical function-граматична функція
- 51.grammatical relation-граматичний зв'язок
- 52.greedy algorithm-жадібний алгоритм
- 53.hallucination-галюцинування
- 54.head of a dependency parsed pair-вершина, тут: вершина розібраної за принципом граматики залежностей пари слів
- 55.hidden Markov model-прихована модель Маркова
- 56.in-context learning-?
- 57.indirect object-непрямий додаток
- 58.inference-based learning-навчання на основі міркувань, умовиводів (інференцій)
- 59.label-scorer classifier-класифікатор з лічильником тегів, міток
- 60.large language model (LLM)-велика мовна модель
- 61.lemma-лема
- 62.localization-локалізація, адаптація
- 63.machine translation-машинний переклад
- 64.Markov chain-марковський ланцюг, ланцюг Маркова
- 65.maximum spanning tree-дерево максимального обсягу, охоплення
- 66.modifier-означення
- 67.modifier relation-означальні відношення
- 68.neural classifier-нейронний класифікатор
- 69.neural language models-нейромережева модель
- 70.neural network-нейронна мережа
- 71.NLP-[natural language processing] опрацювання природної мови
- 72.node-вузол
- 73.nominal modifier-іменникове означення
- 74.nominal subject-іменний (субстантивний) підмет
- 75.normalization-нормалізація
- 76.null hypothesis -нульова гіпотеза

- 77.numeric modifier-числове (числівникове) означення
- 78.object-об'єкт, додаток
- 79.oracle-оракул
- 80.parallel corpus-паралельний корпус
- 81.parse tree-дерево синтаксичного розбору, синтаксичне дерево
- 82.parser-(синтаксичний) аналізатор, парсер
- 83.parsing-аналіз, (автоматичний) синтаксичний аналіз, (автоматичний) синтаксичний розбір; (автоматичний) граматичний аналіз, парсинг
- 84.partial parsing-частковий синтаксичний аналіз
- 85.period disambiguation-усунення неоднозначності ролі крапки
- 86.perplexity-перплексія
- 87.phrase-structure analysis-синтаксичний аналіз за принципами граматики складників
- 88.POS-частина мови
- 89.POS-tagging-морфологічна розмітка, морфологічне розмічування, частиномовна розмітка, частиномовне розмічування
- 90.post-editing-постредагування
- 91.precision-точність
- 92.pretraining -попереднє тренування
- 93.projective arc-проєктивна дуга
- 94.prompt-запит
- 95.prompt engineering-промпт інженерія , інженерія запитів
- 96.recall-повнота
- 97.root of a structure-корінь структури, корінь дерева
- 98.sampling / text sampling-вибирання, відбір
- 99.scaling laws-закони подібності
- 100. self-training-самонавчання
- 101. semantic relation-семантичне відношення
- 102. semantic role labeling-розмічування семантичних ролей, визначення семантичних ролей

- 103. shift-reduce parsing-синтаксичний аналіз за принципом "зсув-зменшення"
- 104. stack-стек, лінійний список
- 105. subject-підмет
- 106. supervised finetuning (SFT)-контрольоване / кероване дотренування (з наглядом, з учителем)
- 107. syntactic structure-синтаксична структура
- 108. syntax-синтаксис
- 109. system prompt-системний запит/промпт, початковий запит/промпт
- 110. teacher forcing-?
- 111. text categorization-класифікація текстів
- 112. text summarization-реферування тексту, автоматичне реферування тексту
- 113. the Universal Dependencies project-проект "Універсальні залежності"
- 114. thematic grid, θ -grid, case frame-тематична сітка, θ -сітка
- 115. to freeze the parameters-фіксувати параметри
- 116. token-токен, слововживання
- 117. token buffer-буфер токенів
- 118. training oracle-оракул навчання
- 119. transformer-трансформер
- 120. transition-based parsing-синтаксичний аналіз на основі системи переходів
- 121. treebank-банк (синтаксичних) дерев
- 122. typed dependency structure -типізована структура залежностей
- 123. wordform-словоформа
- 124. zero-shot-?

*Додаток 2. Посилання на табличну модель словника у форматі *xlsx*.*

dictionary_main.xlsx

*Додаток 3. Посилання на тлумачний перекладний словник термінів
комп'ютерної лінгвістики*

<https://www.lexonomy.eu/syxt6d4e>