

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА

ІНСТИТУТ ФІЛОЛОГІЇ

Фокін С.Б.

**КОМП'ЮТЕРНІ ІННОВАЦІЙНІ ТЕХНОЛОГІЇ
В ПЕРЕКЛАДАЦЬКІЙ ТА
ПЕРЕКЛАДОЗНАВЧІЙ ДІЯЛЬНОСТІ**

2019

Рецензенти:

проф. Голубовська І.О., д.філол.н., професор, завідувач кафедри загального мовознавства, класичної філології та неоелліністики;

доц. О.М. Калустова, к.філол.н., доцент кафедри теорії і практики перекладу з романських мов імені Миколи Зерова Інституту філології;

ас. І.М.Шиянова, к.філол.н., асистент кафедри теорії і практики перекладу з романських мов імені Миколи Зерова Інституту філології.

Комп'ютерні інноваційні технології в перекладацькій та
перекладознавчій діяльності / Навчальний посібник.
С.Б. Фокін. – К.: КНУ імені Тараса Шевченка, 2019 – 254 с.
*Затверджено Вченою радою Інституту філології
(протокол № 3 від 30 жовтня 2018 року)*

Навчальний посібник

**Батькам моїм,
Наталії Мусіївні
і Борису Яковичу,
присвячую**

ЗМІСТ

ПЕРЕДМОВА.....	7
РОЗДІЛ I. ТИПИ, СТРУКТУРА, ХАРАКТЕРИСТИКИ І ЗАСТОСУВАННЯ КОМП'ЮТЕРНИХ СЛОВНИКІВ.....	11
1.1. Визначення “слова” з погляду комп'ютерної лексикографії.....	11
1.2. Організація словника: макроструктура та мікроструктура.....	14
1.3. Принципові відмінності між електронними і паперовими словниками.....	16
1.4. Критерії класифікації електронних словників.....	20
1.4.1. Види двомовних перекладних словників.....	23
1.4.2. Види одномовних словників та їхнє значення для практики перекладу.....	30
1.4.3. Характеристики електронного конкордансу.....	37
1.5. Проблема розрізнення значень лексем у комп'ютерних словниках.....	39
1.6. Особливості структури і використання ономасіологічних словників.....	45
1.6.1. Тематичний словник.....	48
1.6.2. Атлас лексики.....	49
1.6.3. Тезаурус.....	49
1.6.4. Інверсійний словник.....	54
1.6.5. Власне ономасіологічні словники.....	56
1.7. Принципи класифікації лексики в електронних ономасіологічних словниках.....	63
Контрольні питання.....	68
РОЗДІЛ II. ТЕРМІНОГРАФІЯ ТА ЕЛЕКТРОННЕ ВПОРЯДКУВАННЯ ТЕРМІНОЛОГІЧНИХ БАЗ ДАНИХ.....	74
2.1. Проблеми термінологічної стандартизації в електронних базах даних.....	74
2.2. Характеристики термінологічних словників.....	78

2.3. Вимоги до дефініцій термінів у спеціалізованих словниках.....	80
2.4. Методи пошуку перекладу терміна за допомогою комп'ютерних технологій.....	85
2.5. Міжнародні формати термінологічних карток.....	94
2.6. Критерії вибору з-поміж термінологічних синонімів..	102
2.7. Автоматизація вибірки термінів з текстових масивів..	107
2.8. Специфіка структурування найпростіших баз даних	108
Контрольні питання.....	112
РОЗДІЛ III. КОРПУСНІ РЕСУРСИ В ДІЯЛЬНОСТІ І ПІДГОТОВЦІ ПЕРЕКЛАДАЧА.....	115
3.1. Потенціал використання корпусів у практиці перекладу.....	115
3.2. Основні принципи пошуку в корпусах текстів.....	120
3.3. Перекладацькі корпуси і корпуси паралельних текстів.....	125
3.4. Огляд українськомовних корпусів текстів.....	129
3.5. Якісні та кількісні характеристики корпусів іноземних мов.....	132
3.5.1. Багатомовні колекції корпусів.....	133
3.5.2. Англійська мова.....	134
3.5.3. Іспанська мова.....	135
3.5.4. Італійська мова.....	137
3.5.5. Німецька мова.....	137
3.5.6. Польська мова.....	138
3.5.7. Російська мова.....	138
3.5.8. Французька мова.....	139
3.6. Використання корпусів у навчанні перекладу.....	140
Контрольні питання.....	142
РОЗДІЛ IV. ПОТЕНЦІАЛ ВИКОРИСТАННЯ КОРПУСІВ У ПЕРЕКЛАДОЗНАВЧИХ ДОСЛІДЖЕННЯХ.....	145
4.1. Виявлення універсальї перекладу.....	145
4.2. Дослідження перекладної мови в зіставленні з мовою перекладу.....	147

4.3. Аналіз позамовних чинників і дискурсивних категорій.....	149
4.4. Дослідження міри індивідуального в перекладі...154	
4.5. Переваги корпусів малого обсягу.....158	
4.6. Методика укладання малого корпусу.....166	
4.7. Перевірка достовірності вибірки.....168	
4.8. Значення корпусів для розробки і налаштування роботи штучного інтелекту.....172	
Контрольні питання.....175	
РОЗДІЛ V ДОПОМІЖНІ ПРОГРАМИ В ДІЯЛЬНОСТІ ПЕРЕКЛАДАЧА І ПЕРЕКЛАДОЗНАВЦЯ.....178	
5.1. Функції текстових редакторів, корисні у практиці перекладу.....178	
5.2. Принципи роботи програм перевірки якості перекладу.....185	
5.3. Принципи роботи програм обробки природної мови у системах штучного інтелекту.....186	
5.3.1. Роль лематизаторів та програм частиномовної розмітки для зняття граматичної двозначності.....190	
5.3.2. Методи зняття лексичної двозначності шляхом автоматичного аналізу.....191	
5.3.3. Застосування комп'ютерної лінгвістики у системах штучного інтелекту.....204	
5.3.4. Комп'ютерна лексикографія і системи машинного перекладу.....206	
Контрольні питання.....209	
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....212	
Теоретичні праці.....212	
Лексикографічні, термінологічні ресурси і бази даних.....227	
ДОДАТОК 1. Параметри основних текстових корпусів.....247	

ПЕРЕДМОВА

Широке розмаїття електронних ресурсів, що буквально експоненційно зростає на наших очах, з одного боку, допомагає перекладачам-практикам і дослідникам, а з іншого – спонукає тримати руку на пульсі новітніх розробок, розвивати вміння обирати оптимальний ресурс з-поміж численної кількості електронних знарядь, відрізняти якісні професійні напрацювання від аматорських експериментальних праць. Умовно електронні ресурси, корисні для перекладача, можна поділити на декілька груп: словники, корпуси, програми аналізу й обробки тексту.

Словники. З тлумачними та перекладними словниками обізнані як філологи, так і пересічні користувачі, однак розростання кількості електронних програм і додатків, збагачення їх новими функціями вимагає детальних пояснень, спеціальних знань та ширших узагальнень. Скажімо, не всім відома функція зворотного пошуку у словниках (від дефініцій до означуваного поняття), хоча з 2017 року ця функція наявна в тому числі у тлумачному словнику української мови “СУМ”; далеко не всі під час пошуку слова або виразу вправно використовують позначки, помітки, спеціальні теги; мало хто вповні користується гіпертекстуальними функціями електронних лексикографічних джерел. Ще не всім філологам зрозуміла роль і можливості конкордансів для виконання завдань теорії і практики перекладу.

У широкому розумінні словник як упорядкована система тих чи інших мовних (або мовленнєвих) одиниць з тлумаченням їхнього значення звернена не обов'язково виключно на лексику, і цей факт потребує додаткового, глибшого осмислення з погляду ролі і завдань словникарства. Тому необхідно забезпечити ознайомлення студентів з об'єктом і завданнями комп'ютерної лексикографії, принциповими суттєвими відмінностями електронних словників і словників у паперовому форматі, сприяти закріпленню на практиці знань стосовно функцій деяких словників спеціального призначення, засвоєнню базових методів укладання двомовних електронних словників.

Корпуси текстів. Якщо ще десятиліття тому перекладачі й дослідники, що регулярно користувалися корпусами для перевірки правильності, сполучуваності, узусу чи статистики, були рідкісними винятками, сьогодні компетенція перекладача й дослідника

перекладу передбачає також знання відмінностей між різними корпусними інструментами, уміння обирати найкраще для заданої мети знаряддя. Первинний, не надто детальний пошук показує, що для таких мов як англійська, іспанська, італійська, німецька, українська, російська та французька існує більше десятків доступних корпусів текстів, і нині слово “корпус” вже не вражає, як було раніше, а ставить дослідника перед дилемою вибору між численними ресурсами. Загалом корпуси можна умовно розподілити на великі групи за такими критеріями: мова та кількість мов, наявність паралельних текстів, кількість слововживань, наявність розмітки, наявність лематизатора, можливість створення конкордансу для слів та виразів, синхронність/діахронічність зрізу мови, ареальна, соціолектна та ідіолектна розмітка текстів, розмежування усних та письмових текстів у корпусі тощо. Орієнтування у такому багатому розмаїтті вимагає підготовки і спеціальних знань.

Знаряддя автоматичного аналізу текстів. Окрім того, що з'являються нові програми, дедалі більше виникає потреб щодо необхідності їхнього вдосконалення, розширення функціональності, зміни методів опрацювання текстів. Частиномовна розмітка тексту відіграє суттєву роль у програмах комп'ютерного перекладу та інших знаряддях штучного інтелекту, оскільки є важливим етапом процесу зняття двозначності; поза тим, численні дослідження спираються на вибірку тих чи інших частин мови або окремих парадигматичних форм, що нині можна і доцільно автоматизувати; окремим викликом залишається семантична розмітка текстів, коло використання якої надзвичайно широке: від вдалого розміщення рекламної продукції, автоматичного реферування, генерування текстів новин аж до автоматичного укладання комп'ютерних словників і машинного перекладу. Формальні методи опрацювання текстів мають свої безсумнівні переваги, хоча й не позбавлені очевидних недоліків; з-поміж різних програм, що весь час видозмінюються і вдосконалюються, необхідно не тільки виокремлювати та знати наявні знаряддя (у лавині сучасних розробок знання не можуть бути навіть приблизно вичерпними), а й уміти проводити цілеспрямований пошук необхідних інструментів, оцінювати раціональні шляхи використання, зважувати “за” і “проти” одного програмного забезпечення в порівнянні з іншим. Тому підготовка сучасного перекладача й перекладознавця буде повною не лише у фактологічному, а й у методологічному планах.

Запропонований посібник має на меті забезпечити дидактичним матеріалом такі спецкурси як “Інноваційні технології в перекладацькій діяльності”, “Комп’ютерна лексикографія”, “Машинний переклад”, “Методологія перекладознавчих досліджень”. Докладніше матеріал лекцій, що стосується безпосередньо типології і укладання словників, викладено в посібнику автора “Комп’ютерна лексикографія і переклад” (іспанською мовою). Аспекти, що стосуються машинного перекладу, у посібнику висвітлюються лише в загальних рисах, оскільки ця галузь доволі детально розглянута у працях колег, зокрема, Е.М. Андрієвської “Сучасна електронна лексикографія в роботі перекладача”, О.О. Ткаченка “Машинний переклад і редагування текстів для ОКР ‘magіstr’ програми “Переклад: французька, іспанська та італійська мови”. З погляду глибинних принципів роботи, методів і алгоритмів програм машинного перекладу розглядаються у праці Н.П. Дарчук “Комп’ютерна лінгвістика (автоматичне опрацювання тексту)”.

У I розділі посібника розглядаються предмет і завдання комп’ютерної лексикографії у порівнянні з традиційним словникарством, її основні виклики і надбання, аналізується об’єкт, побудова, типи, а також деякі парадокси словників. Розділ II присвячено особливостям структурування електронних термінологічних баз даних, а також укладання термінологічних двомовних перекладних словників з точки зору стандартизації та формальних вимог до укладання термінологічних карток. У III розділі висвітлено основні моменти професійного використання корпусів, зокрема, у перекладацькій діяльності та в системах штучного інтелекту, розглядається простий і розширений пошук слів та виразів у корпусі (у тому числі, деяких регулярних виразів), принципової можливості/неможливості створення універсальної семантичної мови-посередника. Окремо описано методи пошуку еквівалентів для двомовних перекладних словників. Розділ IV охоплює методи корпусної лінгвістики, зокрема, роботу з корпусами, методику укладання дослідницького корпусу та оцінки його достовірності з перекладознавчої точки зору, методи автоматичного аналізу перекладознавчого емпіричного матеріалу, переваги і недоліки формального аналізу; окрема увага приділяється формальним маркерам теми, стилю, жанру, дискурсу оригінального і перекладного текстів та іншим позамовним чинникам. Розділ V присвячено програмам, що є допоміжними у перекладацькій і перекладознавчій діяльності: текстові редактори,

програми обробки текстів, алгоритми машинного перекладу та деякі інші знаряддя штучного інтелекту. У цьому розділі узагальнено наукові результати, що опубліковано як у наших окремих розвідках, так працях і інших учених.

У сучасному інформаційному просторі неможливо врахувати всі наявні розробки й здобутки. Поки публікується книга, виходять нові словники, захищаються дисертації, з'являється нове програмне забезпечення. Тому, на превеликий жаль, неможливо досягти вичерпності під час висвітлення питання, пов'язаного з інноваційними комп'ютерними технологіями. А з іншого боку, проведення зрізу стану речей на матеріалі сотень вітчизняних і зарубіжних праць усе ж дозволяє в певному наближенні змалювати загальну картину і спрогнозувати напрям практичного розвитку галузі й теоретичних проблем, намітити шляхи їх розв'язання. Автор щиро вдячний офіційним та неофіційним рецензентам, що завдяки уважному прочитанню та конструктивним зауваженням і порадам сприяли щонайкращому викладу узагальненого тут матеріалу: Голубовській Ірині Олександрівні, Калустовій Ользі Михайлівні, Капишону Тарасу Миколайовичу, Шияновій Ірині Михайлівні. Автор дякує також тим, хто небайдужим ставленням і щирою вірою надихав на написання цієї праці: Артемовій Людмилі Василівні, Волошину Федору Михайловичу, Дубінській Лілії Львівні, Іванову Максиму Сергійовичу, Коган Орні, Неборсіній Наталії Павлівні, Нікітіній Оксані Борисівні, а також усім близьким, родичам, приятелям, колегам.

РОЗДІЛ І. ТИПИ, СТРУКТУРА, ХАРАКТЕРИСТИКИ І ЗАСТОСУВАННЯ КОМП'ЮТЕРНИХ СЛОВНИКІВ

1.1. Визначення “слова” з погляду комп'ютерної лексикографії

Значення іменника “слово” не є чітко окресленим з термінологічної точки зору. З міркувань точності розрізняють поняття графічного слова, фонетичного слова, вокабули, реєстрового слова та лексеми. Наприклад, дієслово “to get” (англ.), безумовно, є словом, хоча в інфінітивній формі воно займає два графічних слова, розділених пробілом, так само як у складних дієслівних часах *Present Perfect*, *Past Perfect* та інш. І якщо в графічному аспекті ми можемо говорити про два слова, з фонетичної точки зору зазначені форми переважно читатимуться разом, нероздільно. З метою автоматичної обробки тексту, укладання комп'ютерних словників та будь-якого іншого технічного опрацювання текстового матеріалу необхідно встановити чітке термінологічне розмежування між різними значеннями поняття, що його охоплює “слово”. Для уникнення непорозумінь варто розрізняти “графічне слово”, “фонетичне слово”, а також вокабулу, лексему, словоформу, слововживання. Розглянемо послідовно кожен зі згаданих термінів.

Слово. Цей термін стосується формального представлення мовної одиниці в графічному або фонетичному планах. Таким чином, графічними словами називаються сполучення літер, написаних разом¹. Аналогічно фонетичними називаються ті слова,

1

Цей термін у галузі лексикології характеризується омонімічністю. Зокрема, у французькій лексикологічній традиції “mots graphiques” — слова, що не можуть вживатися самостійно, окремо від сталого виразу [Smouchtchynska 2015, с. 577]. Натомість Ш. Мюллер графічним словом називає складову тексту, що є зручною для кількісних підрахунків, отже, це є слово, відокремлене роздільниками, з-поміж яких найтипівшим є пробіл [Ch. Muller. — Initiation à la Statistique Lexicale. Paris, Larousse. 1968]. — цит. за Р. Альваром Ескеппою [Alvar Ezquerria].

що вимовляються безперервно. У такому випадку англійське речення “I’ve got it” слід вважати одним фонетичним словом.

Вокабули – слова або вирази з іноземної мови, дібрані в певному порядку й перекладені для заучування напам’ять [СУМ]².

Лексема – це слово як частина лексичної системи певної мови. Ключовою характеристикою лексеми є наявність спільного лексичного значення для всіх форм. Так, наприклад, французькі дієслівні форми “me lève”, “te lèves”, “se lève” становлять єдину лексему, хоча йдеться про 6 різних графічних слів.

Зрештою, терміни “слово”, “лексема”, “вокабули” не мають взаємовиявляючого характеру. Один і той самий мовний знак може одночасно бути словом, вокабулою і лексемою; різниця виявляється тоді, коли виникає необхідність пов’язати певним відношенням різні мовні знаки. Так, “дромедар” і “дромедари” є однією лексемою і водночас різними графічними словами; численні випадки омонімів (“захід” – частина світу, “захід” – подія, влаштована з певної нагоди; “потяг” – прагнення, “потяг” – поїзд тощо) становлять різні лексеми, графічні слова яких між собою можуть збігатися.

Реєстрове слово – це знак, виражений сполученням графічних літер, що підлягає тлумаченню у словнику. Найчастіше реєстрові слова у словникових статтях виділені жирним шрифтом. Одне реєстрове слово не обов’язково збігається з лексемою, з огляду на можливість збігу омонімів в одній словниковій статті. Серед синонімів цього поняття також поширені назви: “гасло”, заголовне слово, “лема”. На наш погляд, “реєстрове слово” вже не вповні передає суть поняття, оскільки те, що тлумачиться у словнику, часто є словосполученням, а отже, містить декілька слів.

У професійному спілкуванні, що вимагає наукової строгості, необхідно уникати терміна “слово” без визначників, натомість слід вживати один зі згаданих термінів. Наприклад, словосполучення “словник обсягом 50 000 реєстрових слів” сформульовано точніше, ніж “словник, що містить 50 000 слів”, адже останнє

²В інших мовах у відповідні когнати, тобто, слова, що схожі за звучанням на “вокабулу”, вкладають інше значення. Зокрема, у французькій мові “vocabulary” тлумачиться як слово, термін, що характеризується набором значень, при цьому різні форми слова є однією вокабулою [Dictionnaire Larousse]. В іспанській мові “vocablo” – це слово як складова одиниця мови, графічне представлення слова [DRAE]. В англійській мові слово “vocabulary” визначають як “слово, зокрема, в його графічному, радше, ніж значеннєвому аспекті” [Oxford English Dictionary]. З огляду на такі розбіжності у подальших розділах уникаємо вживання цього терміна.

твердження можна розтлумачити по-різному: 50 000 реєстрових слів (лем), словникових статей чи навіть графічних слів.

Ще два важливих для опису лексики терміни – це **словоформа** та **слововживання**. Під час підрахунку обсягу тексту може виникнути непорозуміння. Наприклад, поданий нижче фрагмент (перші слова з “Дон Кіхота”) містить 12 графічних слів, з-поміж яких 11 унікальних: “En un lugar de la Mancha de cuyo nombre no quiero acordarme...” (ісп.). Графічне слово “de” повторюється один раз. Тому, якщо виходити з різних критеріїв, істинними будуть обидва твердження: “у фрагменті 11 слів” і “у фрагменті 12 слів”. Для уникнення подібних непорозумінь введено поняття словоформ і слововживань. Кількість “словоформ” – це кількість графічних слів у тексті без урахування повторів; “слововживання” (інколи позичається англomовний термін “токен” (англ.: “token”), що, однак, не зареєстрований у провідних словниках української мови) – усі графічні слова з урахуванням повторів. Прийменник “de” у першому реченні “Дон Кіхота” повторений і становить у даному випадку одну й ту саму словоформу, представлену двома слововживаннями. Отже, у наведеному прикладі 11 словоформ та 12 слововживань.

У численних наукових публікаціях мовознавці використовують коефіцієнт, що отримується шляхом ділення кількості словоформ на кількість слововживань у певному тексті. У наведеному прикладі цей коефіцієнт становитиме $11/12$, тобто, 0,91666. Вказаний коефіцієнт (в англійській аббревіатурі – *TTR*, “types/tokens ratio”, частка від ділення кількості словоформ на кількість слововживань у тексті) подекуди використовується для непрямого розрахунку лексичного багатства тексту. І справді, можна припустити, що чим більше у тексті словоформ, тим він більш лексично розмаїтий. Однак не без іронії слід буде визнати, що речення “To be or not to be?” через велику кількість повторів виявиться лексично відносно бідним, адже на 6 графічних слів трапляються аж дві повторені словоформи, і його *TTR* дорівнює 0,66 (4 унікальні словоформи на 6 слововживань). Насправді ж точнішу картину лексичного розмаїття твору має дати не кількість словоформ, а кількість лексем. Для приведення різних словоформ до початкової форми лексеми (приміром, “працюючи”, “працюєш”, “працюємо”, “працюймо” – до інфінітива “працювати”) використовуються спеціальні програми, що називаються лематизаторами.

Розмежування словоформ і слововживань має значення і для перекладознавчих досліджень. За даними А. Кутузова, *TTR*, безумовно відмінний в оригіналі і перекладі, показує парадоксально схожу динаміку змін у першотворі і цільовому тексті³. Схожість стає очевидною при нанесенні змін на графік: малюнок кривої оригіналу за своїми контурами дуже близький до відповідного графіка даних тексту перекладу [Kutuzov 2010]. Чи така зміна коефіцієнта є універсальною картиною в художньому тексті, чи характеризує тільки окремі жанри, окремих авторів, окремих перекладачів? Наскільки величина *TTR* може бути ознакою ідіостилу автора та індивідуального стилю перекладача? Чи можна застосовувати цей коефіцієнт для встановлення авторства перекладу? Наскільки точнішими будуть результати, якщо замість простого розрахунку словоформ брати до уваги кількість лексем? Ці та інші питання ще очікують на свого дослідника.

Отже, розмежування значень терміна “слово” шляхом уживання специфічних понять: “графічне слово”, “фонетичне слово”, “реєстрове слово”, “вокабула”, “лексема”, “словоформа” та слововживання” дозволяє уникнути непорозумінь, а також відкрити нові дослідницькі ніші у галузі автоматичного опрацювання текстів.

1.2. Організація словника: макроструктура та мікроструктура

Як уже сказано, складові частини в словнику мають бути організовані у систему за певними критеріями. А самі ці складові, у свою чергу, теж мають власну внутрішню організацію. З цією метою прийнято розрізняти макроструктуру та мікроструктуру словників.

Макроструктура словника охоплює систему зв'язків між словниковими статтями, їхню послідовність та особливості впорядкування (зазвичай – це певні знаки, що супроводжуються поясненнями їхніх значень). Мікроструктура – це будова окремої складової одиниці словника, найчастіше, словникової статті.

Найпоширеніша макроструктура в словниках передбачає сукупність словникових статей, розташованих за алфавітним порядком.

Словникова стаття складається з двох частин: лівої (що також називають гаслом, реєстровим словом, заголовним словом, лемою, яка може містити одне або декілька графічних слів), та

³Визначники “вихідний” і “цільовий” корелюють з поняттями “оригінальний” і “перекладний”. Відповідно, “вихідна мова”, “вихідний текст” – текст оригіналу; “цільова мова”, “цільовий текст” – текст перекладу.

правої частини (глоси), що містить пояснення лема і, часом, додаткові характеристики.

Оскільки ліва частина словникової статті – це лема, що підлягає тлумаченню, а права – глоса, яка власне і є тлумаченням, можна провести аналогією між словником природної мови і масивом об'єктів в інформатиці, кожен з яких складається з певного ключа та його значення (у мові програмування *Python* такий масив називається буквально терміном “dictionary”). У кожного члена масиву певне значення зазвичай знаходять за ключем, і значно рідше – навпаки, ключ знаходять за певним значенням. І справді, легше зв'язку ключів принести до дверей, щоб підібрати саме підхожий до них ключ, ніж до одного ключа принести певну кількість дверей. Так само і в словнику користувачеві зручно шукати лему за алфавітом, у той час як вкрай важко було б відшукати реєстрове слово через його дефініцію. Однак на практиці справді виникають подібного роду задачі, і в комп'ютерну еру такий зворотний пошук виявляється цілком зримим. Про певні спроби віднаходження відповідних лем за їхнім значенням докладніше йдеться у пункті 1.6.4.

За умов детальної параметризації електронного словника його технічні можливості (оптимізація пошуку, автоматизація процесів, застосування у системах штучного інтелекту) колосально зростає. Як зазначає С.В. Гриньов, “Створення типології словників передбачає виокремлення певних лексикографічних характеристик, параметрів, що слугують підставою для класифікації, опису і оцінки словників, а також їх проектування. Параметризація надає можливості і формалізації, і певного спрощення технології створення словників, і уніфікації їхніх композиційних засобів” [Гринев 1990, с. 26]. Дослідник підсумовує, що такі параметри можна узагальнити за групами: формальні ознаки, аспект мови, хронологічні характеристики, граматичні характеристики, різного роду метадані тощо [Гринев 1990, с. 27]. Отже, докладна розмітка⁴ мікроструктури сприяє значній оптимізації і підвищенню функціональності електронного словника.

В окремих випадках права частина словника дуже обмежена, вона не містить додаткових характеристик, властивостей, і в двомовних словниках інколи зводиться виключно до одного варіанта перекладу лема. У такому випадку лема і глоса мають настільки симетричний характер, що виявляється можливою їхня взаємна перестановка (якщо подається тільки один переклад

⁴Окрім терміна “розмітка”, паралельно вживаються його синоніми: “нотування”, “маркування”.

леми). Такого роду словники зі значно обмеженою глосою називаються глосаріями. Деякі глосарії можуть взагалі не містити глоси, тобто, складатися лише зі списку термінів або лексичних одиниць іншого типу. Глосарії, що містять переклад лем з однієї мови на іншу, називаються двомовними; якщо переклад подається трьома або більше мовами, глосарії називаються багатомовними.

У глосаріях мікроструктура дуже примітивна, і лема з глосою часто взаємно симетричні, у той час як словники в повному сенсі цього слова характеризуються асиметрією між лемою та глосою: глоса більша за обсягом та кількістю даних, ніж лема, тому мікроструктура словників значно багатша, ніж у глосаріїв. Зокрема, глоса словникової статті може містити дані про частинимовну належність лексичної одиниці, фонетичні особливості, граматичні характеристики, семантичні розряди, декілька лексико-семантичних варіантів (тобто, складових значення) слова, типові сполучуваності, стилістичні та соціолокальні помітки. Електронні словники, окрім мікро- та макроструктури, характеризуються також гіперструктурою, про яку детальніше сказано у пункті 1.3.

1.3. Принципові відмінності між електронними і паперовими словниками

Комп'ютерні словники відрізняються від паперових низкою кількісних та якісних характеристик.

У кількісному плані одна з найочевидніших переваг електронних словників – це їхній відносно необмежений обсяг. На відміну від традиційних словників у паперовому форматі, електронні словники можуть містити набагато масштабнішу базу даних, оскільки для їхньої публікації немає потреби заощаджувати папір, скорочувати кількість статей і пояснень тощо. Однак комп'ютеру необхідно дати чіткі вказівки для розпізнавання тих чи інших компонентів словникової статті, тому під час переведення словника у комп'ютерну версію слід змінити його формат [Перебийніс 2009, с.188-190]. Детальна розмітка структурних частин словника необхідна для забезпечення і оптимізації комп'ютерного пошуку. Так, електронні словники дозволяють проводити складний пошук, що найчастіше немислимий у словниках паперового формату: до прикладу, цілеспрямований пошук слів, що містять певний суфікс, або слів певного походження, лексем якоїсь заданої граматичної характеристики.

Такого роду автоматизований пошук напрочуд легко реалізувати в будь-якому електронному словнику. Більше того, якщо лексеми у словнику розмічено відповідними тегами (спеціальними формальними позначками, що вказують на специфічну характеристику тексту, його частини), пошук можна проводити і за семантичними розрядами, категоріями, такими як, наприклад, абстрактні іменники або акціональні дієслова.

Отже, електронний словник здатний охопити набагато більше функцій, ніж їх можна реалізувати у паперовому словнику. Приміром, В. Селегей стверджує, що словникова стаття в ідеалі має містити випадки паремії, елементи розмовника, ілюстрації узусу, типові мовні труднощі [Селегей 2003].

Окрім очевидних кількісних розбіжностей, необхідно наголосити на якісних відмінностях, які вказують власне на суть та перспективи словників нового покоління. Однією з найвиразніших рис комп'ютерних лексикографічних ресурсів є наявність у них численних гіперпосилань (зазвичай представлених у формі слів, виділених синім кольором). Розгалужена система гіперпосилань дає змогу пов'язувати реєстрове слово, а так само і будь-які інші знаки, присутні в словнику, з іншими знаками, якщо такий зв'язок корисний для користувача. Завдяки використанню гіперпосилання читання словника перестає бути лінійним. Кожен користувач проходить "власний маршрут", і кількість таких маршрутів видається такою ж безмежною та неосяжною, як і кількість можливих шахових партій. Подібна гіпертекстуальність певною мірою властива і паперовим словникам, а надто енциклопедіям, у яких прийнято виділяти курсивом ті графічні слова, для яких є окрема реєстрова позиція. Однак ступінь гіпертекстуальності комп'ютерних словників на порядки вищий, зокрема, в них кожне слово словникової статті може сприйматися як потенційне заголовне (отже, як гіперпосилання на статтю про нього), і саме за таким принципом працює тлумачний словник Королівської академії іспанської мови [DRAE].

Окрім власне великої кількості гіпертекстових зв'язків, важливою характеристикою комп'ютерних словників вважаємо те, що зв'язки між його частинами, потенційно важливі для користувача, виявляються не завжди очікуваними і передбачуваними. Цей факт уже сам собою викликає науковий інтерес. Приміром, якщо встановити зв'язки між лексемами зі схожими глосами, такі зв'язки допоможуть у розробці алгоритму автоматичного пошуку синонімів.

Кількість комп'ютерних словників щоразу більшає, однак з-поміж них зустрічаються принципово дуже відмінні ресурси. Для короткого огляду здійснимо їхню умовну класифікацію. Електронні словники можна розділити на дві великі групи: ті, що раніше існували в паперовому вигляді і згодом були переведені в електронний формат, а також ті, що укладені власне комп'ютерами автоматично [Клименко 1996, с. 12-13]. Звичайно, до цієї двочленної класифікації слід додати й ті словники, що створені людиною, однак не на основі паперової версії, а одразу в електронному форматі. Словник, переведений з паперового формату в електронний, називають комп'ютеризованим [Перебийніс 2009, с. 189]. Між комп'ютерними і комп'ютеризованими словниками існують суттєві якісні відмінності. Зупинимось на них детальніше.

Звичайно, автоматизований пошук можна здійснювати і в електронному словнику, і в комп'ютеризованому. Однак саме електронний словник, база даних якого заточена під найвибагливіші запити користувача, дозволяє проводити цей пошук на якісно новому рівні, а саме, не лише знаходити відповідний збіг символів, а розрізняти, у яких частинах структури й макроструктури трапляються збіги (у лемі, глосі, соціолокальних помітках, ілюстративних прикладах тощо), шукати у прямому порядку та зворотному і т. ін. У такий спосіб одним натисканням кнопки словник може перетворюватися на інверсійний, обернений, зведений та інші продукти лінгвістичного конструювання⁵; так само одним натисканням можна згенерувати, наприклад, словник іспанських дієслів, а, за потреби – іспанських перехідних дієслів, іспанських неперехідних дієслів. Приміром, якщо у тлумачному словнику Іспанської королівської академії DRAE обрати помітку "vulg." [DRAE], можна автоматично скомпіювати словник вульгарних слів і т. ін. Тобто, можливість сортування та вибірки за різними критеріями – властивість, якої позбавлені паперові словники – надає електронним словникам принципово нових якостей. Завдяки ним вони стають потужним допоміжним знаряддям штучного інтелекту.

Так само, як і при роботі головного мозку, ступінь розвинутості інтелекту вимірюється не кількістю нейронів і не масою мозкової

⁵Лінгвістичне конструювання – складний, багатоплановий термін, основні характеристики якого, відповідно до його автора, Ю. Караулова, – це створення принципово нового лінгвістичного об'єкта з метою експерименту. Такий експеримент дозволяє провести екстраполяцію характеристик одних явищ на інші, віднайти нові властивості явищ [Караулов 1981, с. 18].

тканини, а кількістю зв'язків, так званих синапсів, за аналогією можна стверджувати, що електронний словник тим корисніший, що більша кількість потенційно релевантних зв'язків між його частинами. Іншими словами, цінність електронного словника визначається розвиненістю його гіперструктури. Корисні зв'язки між компонентами можуть бути забезпечені шляхом встановлення явних гіпертекстових посилань з одних частин словника на інші, а також шляхом автоматичного генерування таких гіперпосилань. Звичайно, усі потенційно релевантні зв'язки не піддаються перерахуванню в одній гіперструктурі, але вважаємо, що добрим орієнтиром можуть стати зв'язки на основі повторюваних знаків (позначок, скорочень, слів). Розглянемо це на прикладі функціонування багатомовного словника фразеологічних синонімів "IdeoPhrase" [IdeoPhrase].

Під час розробки словника ми виходили з тих позицій, що будь-який повтор може виявитися функціонально важливим. Тому пропонуємо будувати принцип пошуку синонімів з-поміж фразеологічних одиниць шляхом зіставлення та пошуку однакових дескрипторів значення, тобто, дескрипторів, що повторюються. Синонімами можуть вважатися не обов'язково ті фразеологізми, дескриптори яких повністю збігаються, а й ті, в яких є більше двох-трьох спільних дескрипторів. За суттю це є спрощеним способом встановлення семантичної відстані⁶ між фразеологічними одиницями. Міра синонімічності, таким чином, пропорційна семантичній близькості. Оскільки у словнику пошук синонімів ведеться як тією ж самою, так і іншими мовами, можна автоматично вести пошук перекладу фразеологізму, а також генерувати перекладні словники усіх наявних мовних пар [IdeoPhrase]. Відповідно, чим більше спільних дескрипторів у фразеологічних одиниць, тим більше ймовірність, що вони є синонімами або взаємними перекладними еквівалентами.

Існують й інші методи визначення семантичної відстані, які теж засновані на аналізі спільних значеннєвих компонентів (напр., метод Шрейдера, метод Арапова-Ратцевої). Докладніше з ними можна ознайомитися у праці Е.Ф. Скороходька "Лінгвістичні основи автоматизації інформаційного пошуку" [Скороходько 1981, с. 181-182].

⁶ Детально методи підрахунку семантичної відстані на матеріалі термінів висвітлено у праці Е.Ф. Скороходька "Лінгвістичні основи автоматизації інформаційного пошуку" [Скороходько 1970]. Доцільно запозичити цю методику і для фразеологічних одиниць та шляхом експерименту перевірити її дієвість.

Таким чином, електронні словники відрізняються від паперових кількісними та якісними характеристиками, зокрема, більшим обсягом, ширшими можливостями пошуку не лише за заголовними словами, а за будь-якими сегментними частинами мікроструктури та макроструктури, що стає можливим за рахунок численних гіперпосилань, детальної структурної розмітки. Можливість автоматичного встановлення потенційно корисних зв'язків між структурними частинами словника шляхом електронного пошуку свідчить про потужний потенціал застосування комп'ютерної лексикографії у системах штучного інтелекту, зокрема, автоматичного пошуку синонімів, перекладу, автоматичного генерування на основі базового словника нових словників та інших об'єктів лінгвістичного конструювання.

1.4. Критерії класифікації електронних словників

У теперішній час існує практично невичерпна кількість класифікацій словників за різноманітними типами й характеристиками. Представлена в цьому підрозділі класифікація електронних словників, звичайно, охоплює не усі можливі типи, які можуть бути корисними в перекладацькій діяльності, і в цьому короткому екскурсі ми зосередимося лише на деяких рисах, що наразі мало відомі перекладачам і перекладознавцям. Підґрунтям для якісної класифікації словників мають бути також такі характеристики: кількість мов, співвідношення форми і змісту (семасіологічні, ономасіологічні), підхід до представлення мовних знаків та фрагментів мовлення (парадигматичні, синтагматичні), охоплений словником мовний або позамовний рівень (лексичний, граматичний, прагматичний, культурний тощо). Ця класифікація доречна як для паперових, так і електронних словників. Утім, для останніх деякі характеристики мають другорядне значення, а окремі риси дають підставу для утворення принципово нових типів словника.

Будь-яка класифікація ґрунтується на певному критерії. Так, приміром, словники за своїм обсягом розподіляються на малі, середні і великі (останні, згідно з В.Г. Гаком – ті, що перевищують обсяг у 100 000 реєстрових слів або словникових статей [Гак 1964, с. 73-74]). Електронні словники характеризуються надзвичайною динамічністю, інтерактивністю; їхній обсяг можна зафіксувати лише станом на певний момент часу, тому часто в їхній назві

взагалі не фігурують такі визначники як “малий”, “середній”, та “великий”; часом не зазначається і обсяг словника.

Ще один можливий, надзвичайно практичний критерій класифікації – кількість мов: у цілому словники поділяються на одномовні, двомовні, багатомовні.

Кількість можливих класифікацій видається необмеженою, і вона постійно зростає по мірі того, як з’являються потреби щодо виокремлення тих чи інших властивостей, характеристик словників. Нові характеристики виникають у відповідь на щоразу суворіші та вибагливіші запити з боку користувачів.

Функції і типи словників доречно пов’язувати з типовими проблемами, які виникають на практиці. Основні труднощі, з якими стикаються перекладачі, поділяються на два типи: 1) труднощі рецептивної стадії перекладу (розуміння лексико-семантичного варіанта, в якому вжито лексеми); 2) труднощі репродуктивної стадії перекладу (втілення змісту оригіналу у засобах іншої мови).

Для розв’язання першої проблеми необхідне широке використання семасіологічних тлумачних словників вихідної мови. Семасіологічні словники – це ті, що спрямовані від мовної форми лексеми (зазвичай, реєстрового слова) до її змісту. Вони становлять більшість як серед перекладних, так і одномовних словників; наприклад, такі словники як “СУМ”, “Мультитран”, “DRAE”, “Grande Dizionario Hoepli Italiano”, “Oxford English Dictionary” та інш. У тлумачних словниках зміст лемми тлумачиться тією ж мовою, якою сформульовано лему; у перекладних словниках – іншою мовою.

За своєю макроструктурою семасіологічні словники складаються зі словникових статей, що починаються лемами. Мікроструктура (структура словникової статті) семасіологічного словника охоплює лему, пояснення або переклад, граматичні характеристики, колокації з відповідною семантизацією.

Друга з окреслених проблем перекладу, що виникає під час репродуктивного етапу, пов’язана з пошуком відповідника цільовою мовою для певного змісту, значно делікатніша. У той час як для розуміння значення ми використовуємо аналітичні методи, під час оформлення засобів вираження для значення ми радше використовуємо синтез. Основні труднощі полягають у тому, що для здійснення синтезу немає універсальних алгоритмів, інструкцій. Відповідно, і скористатися цим методом набагато важче, ніж аналізом, подібно до того, як набагато легше розібрати на частини пристрій, ніж його зібрати з вихідних деталей. Така

закономірність охоплює багато практичних сфер людської діяльності і різних галузей наук, і лексикографія тут не є винятком. Наприклад, якщо інтуїтивно підібрати лексему для вираження такого метамовного змісту: “збіг двох запланованих у розкладі заходів” ще відносно легко (йому відповідає розмовний термін “накладка”), то не так вже елементарно знайти назву, яка відповідала б такому змісту: “слова, що пишуться або вимовляються схоже в різних мовах і мають близькі значення” або такому значенню: “те, що виглядає дуже багатим ззовні та бідним зсередини”. У паперових словниках такий пошук навряд чи можливий; в електронних словниках, натомість, можна шукати слово через його неточну, приблизну дефініцію, дескриптори значень, переклади тощо. Розв’язувати труднощі пошуку засобів вираження можуть ономасіологічні словники, які за своєю макроструктурою спрямовані від змісту до форми. Докладніше про їхню структуру, використання, особливості розробки йтиметься у підрозділі 1.6.

На окрему згадку заслуговують синтагматичні словники, які, на відміну від парадигматичних, зосереджуються на взаємодії мовних знаків у мовленні та особливо корисні для розв’язання проблем сполучуваності, колокації, керування (дієслівного, іменникового, прикметникового), аналізу значень лексем. У зв’язку з цим слід назвати “Словник сполучуваності” (І.П. Сахна) [Сахно 1999]; “Diccionario de Colocaciones del Español” [Diccionario de colocaciones]. За відсутності синтагматичного словника можна з такою самою метою скористатися конкордансом. В електронному конкордансі – словнику, який створюється автоматично на основі текстового корпусу, можна отримати дані щодо частоти вживання словоформи, лексеми або сполучення слів, усіх можливих у корпусі контекстів (ліворуч і праворуч), тобто, широкі відомості щодо різного роду синтагматичних зв’язків. Докладніше про особливості конкордансу сказано у підрозділі 1.4.3 та 3.2.

Залежно від мовного та мовленнєвого рівня, на який спрямований словник, він може бути лексичним, фразеологічним, морфологічним, синтаксичним; окремі словники охоплюють прагматичний, культурний, дискурсивний рівні. Отже, мовний і мовленнєвий рівні можуть стати основою для створення відповідних спеціалізованих типів словників.

У комп’ютерній сфері особливого значення набуває наявність і широкі можливості використання мультимедійних компонентів. Окрім численних ілюстрацій, піктографічних словників, на комп’ютері

можливо також забезпечити прослуховування вимови слова чи словосполучення носіями мови заданого регіону, а також навіть відеоілюстрацій відповідно до різних соціолінгвістичних даних мовця. Таким чином, наявність мультимедійних компонентів – теж суттєва ознака для виокремлення відповідних типів словників.

Велика кількість критеріїв, згідно з якими словники поділяються на типи, а отже, охоплюють різні рівні, спирається на їхні характеристики, структуру. Це дає змогу широко користуватися принципами лінгвістичного конструювання. Оскільки кількість характеристик словників практично невичерпна (навіть якщо взяти до уваги лише тематичний критерій), за допомогою комп'ютерних технологій можна утворювати дедалі більше нових об'єктів лінгвістичного конструювання, вивчати їхні особливості. Приміром, уже існують словники сполучуваності, однак іще немає частотних словників сполучуваності, хоча близькими до них можна вважати словники біграм⁷; для більшості мов ще не створено словників граматичної сполучуваності (для їхнього автоматичного створення потрібно мати корпус текстів з граматичною розміткою)⁸.

Отже, існує практично невичерпна кількість критеріїв, ознак і підстав для виділення тих чи інших типів електронних словників. З погляду перекладу і перекладознавства доречно класифікувати словники за такими критеріями: кількість мов, спрямованість від змісту до форми чи від форми до змісту, рівні, що охоплює словник, наявність мультимедійних компонентів словника. Докладніше типи двомовних і одномовних словників, їхні функції і специфіка розглядаються у наступних підрозділах.

1.4.1. Види двомовних перекладних словників

Електронні двомовні перекладні словники, на відміну від паперових, можна охарактеризувати низкою специфічних рис: 1) відсутність явних обмежень щодо обсягу текстового матеріалу; 2) наявність мультимедійних компонентів; 3) інтерактивність; 4) взаємодія з іншими електронними ресурсами; 5) можливість лінгвістичного конструювання на їхній основі нових лексикографічних ресурсів. Розглянемо послідовно зазначені характеристики, які обумовлюють типологію саме електронних словників.

⁷ Біграма — можлива послідовність двох словоформ у тексті.

⁸ Коли йдеться про граматичну сполучуваність, точнішим терміном слід вважати “колігація”.

1. Звичайно, комп'ютерна пам'ять має фізичні межі, однак текстові ресурси займають порівняно невеликі обсяги пам'яті на тлі інших комп'ютерних ресурсів; комп'ютерна техніка розвивається набагато динамічніше, ніж упорядники двомовних перекладних словників встигають зібрати корисного матеріалу. Відповідно, у них немає потреби заощаджувати на обсязі словникових статей. Звичайно, зростання кількості не обов'язково призведе до зростання якості (нижче розглядаються також і якісні параметри двомовних електронних словників). І хоча розлогість словникової статті за рахунок непомірної кількості варіантів перекладу може часом ставати на заваді оперативного пошуку, двомовні словники в парі з українською мовою поки що характеризуються зворотним недоліком. Разом з тим, додавання до назв електронних словників визначників “великий”, “малий” не набуло поширеної практики.

2. Електронні словники можна поділити на ті, що містять мультимедійні компоненти, і ті, що їх не мають. Супровід двомовної перекладної статті зображеннями, аудіодоріжками та відеосюжетами слугує не лише для забезпечення естетичних потреб перекладача: ознайомлення з денотативним значенням безпосередньо за допомогою органів відчуття відбувається швидше, ніж шляхом сприйняття метамовного опису. Численні імпліцитні візуальні та інші компоненти тексту оригіналу в перекладі можуть ставати більш експліцитними, зазнавати змін, тобто, без ознайомлення з позамовною дійсністю, у перекладі можуть бути викривлені. Так, у зв'язку з труднощами перекладу творів Е. Гемінґвея, Т.Є. Некряч зауважує: “коли перекладач механічно перекладає англійський іменник ‘forest’ українським “лісом”, “для читачів, що не уявляють ні Венсенського, ні Булонського лісу у Франції, які скоріше відображують наше уявлення про парк, або лісопарк – зоровий образ у перекладі викривляється” [Некряч 2014, с. 22]. Отже, коли в тексті оригіналу згадано деталі всесвітньо відомих визначних місць, витворів мистецтва, в тому числі й кіномистецтва, перегляд фотографій та відеофрагментів може виявитися не лише доречним, а й українським необхідним. Немає потреби доводити, що в технічному перекладі наявність довідкового джерела з візуальними компонентами стане тільки плюсом для перекладача. Уваги можуть потребувати і фразеологізми. Приміром, в іспанській мові поширений фразеологізм “tocar madera” (буквально – “стукати по дереву” або

“грати по дереву”, “торкатися дерева”, з огляду на багатозначність дієслова “tocar”). Так само і в нашій культурі деякі носії мови вживають цей вираз і торкаються дерева, як мовиться, “щоб не наврочити”. Проблема полягає в тому, що у нас прийнято казати “постукати по дереву”, у той час як в іспаномовній культурі при вживанні цього виразу дерева лише торкаються. Помилки при перекладі цього фразеологізму, які регулярно доводиться спостерігати, пов’язані саме з неможливістю ознайомитися на власні очі з денотативним значенням словосполучення. Під час перекладу фразеологізмів, таким чином, дуже корисним є “Словник іспанських жестів” [Diccionario de gestos españoles].

3. Електронні словники розподіляються на інтерактивні та неінтерактивні. Інтерактивність, зокрема, можливість користувачів словника обговорювати труднощі, які не вдалося розв’язати, додавати нові лексеми, варіанти перекладу робить електронний словник повнішим, динамічним, додає йому переваг для перекладацької роботи; втім, потенційною небезпекою є поширення помилок; автор, упорядник, модератор форуму словника змушений здійснювати пильний контроль дописів, перевіряти правильність доданих відомостей тощо.

4. Для електронних двомовних перекладних словників найціннішим видом взаємодії з додатковими ресурсами вважаємо ілюстрацію варіантів перекладу контекстом, взятим із баз даних двомовних паралельних текстів. За таким принципом працюють сучасні ресурси *Glosbe*, *Linguee*, *Reverso Context* та інші. Їхня специфіка докладніше розглядається у підрозділі 3.3., оскільки ці словники за суттю є перекладацькими корпусами. Якщо розглядати термін “словник” у широкому сенсі, то межі між словником і текстовим корпусом поступово стираються, оскільки електронний словник за структурою набуває дедалі більше ознак гіпертексту, а корпуси збагачуються граматичною, лексико-семантичною та іншими видами розміток. Хоча ще немає усталеної назви для таких словників, їх можна було б назвати “інтегрованими”.

5. Електронні словники – надзвичайно плідний ґрунт для лінгвістичного конструювання. Прикладів лінгвістичного конструювання нових, корисних для практики і теорії перекладу об’єктів на матеріалі електронних корпусів багато, і ми їх переважно розглядаємо у підрозділах, присвячених одномовним словникам. Однак згадаємо декілька прикладів, що стосуються безпосередньо двомовних електронних словників. Словники,

побудовані на матеріалі двомовних паралельних текстів, такі як *Glosbe*, *Linguee*, *Reverso Context* та ін., дозволяють перекладати в обох напрямках завдяки функції реверсу оригінального і перекладного текстів (такою самою функцією забезпечено програми машинного перекладу за допомогою комп'ютерної перекладацької пам'яті). Звичайно, якщо текст або його фрагмент *A* перекладено як *B*, це не означає автоматично, що *B* буде завжди перекладатися як *A*, адже переклад залежить від багатьох змінних, у тому числі позамовних чинників.

Ще одним прикладом лінгвістичного конструювання на матеріалі двомовних електронних словників може стати створення інверсійного двомовного словника. Інверсійним словником ми називаємо той, в якому користувач отримує доступ до лема через глосу. У випадку двомовних перекладних словників ідеться про пошук лема вихідною мовою через цільову. Це відкриває низку додаткових цінних можливостей для перекладача, зокрема, для перекладу багатьох понять, що в одній мові виражаються доволі регулярно лексемою, а в іншій мові – описово, іншими словами, для умовно безеквівалентних слів. За таким принципом працює іспансько-російський словник *diccionario.ru*. Приміром, для російського прикметника “чужой” в словнику можна знайти переклади “ajeno”, “extraño”, “de otro”, “enemigo”, однак в іспансько-російській частині словника також знаходимо такі відповідники: “alónimo” – “опубликованный под чужим именем”, “con nombre supuesto” – “под чужим именем”, “gorrista” – “живущий за чужой счет” [Испанско-русский]. Такі еквіваленти навряд чи потрапили б до прямого, російсько-іспанського словника. Аналогічно й поняття “успеть”, що в романо-германських мовах не має однослівного відповідника, за допомогою пошуку через зворотний словник може поповнитися деякими еквівалентами, наприклад: “perder el autobús” – “не успеть на автобус”. Цими прикладами засоби вираження ідеї не обмежуються, адже іспанська мова надзвичайно багата на лексико-граматичними засобами для вираження ідеї ‘встигання’, зокрема, у такий спосіб: “no darle tiempo a uno”, дієслово з прислівниковими виразами “apenas”, “por los pelos” тощо. Ці вирази наразі не потрапляють до глос перекладних словників. Ідея пошуку перекладних відповідників в цільовій мові для “безеквівалентних” одиниць вихідної мови зворотним шляхом, тобто, через текст перекладу вихідною мовою, розробляється і в теорії перекладу (див., до прикладу О.М. Калустова [Калустова 2002, с. 130-135]).

Інверсійність словника дозволяє суттєво розширити його функціональність, у той час як саму операцію перестановки глоси і леми можна здійснити шляхом елементарних змін у програмному коді. Це свідчить, що іще достатньо невичерпаних резервів для еволюції двомовних перекладних словників як у кількісному, так і в якісному планах, як за наповненням, так і за структурою.

Вдалим задумом, на нашу думку, є поєднання багатьох складників в одному електронному ресурсі. Згаданий словник *diccionario.ru* [Испанско-русский] можна назвати зведеним (інтегральним) двомовним перекладним словником, адже на сайті є можливість проводити пошук як в окремому словнику, так і в усій сукупності словників, як у лемах, так і у глосах. Хоча ідея зведеного словника в теоретичній лексикографії розробляється на матеріалі однієї мови (приміром, у працях Н.В. Сніжко), практичні розробки у цій сфері, як свідчить згаданий словник, відбуваються і в двомовній лексикографії.

Збагачення словника додатковими структурними частинами, параметрами, опціями передбачає наявність у перекладачів відповідних умінь та навичок їх оптимального використання. Однак з іншого боку, непомірне розростання двомовного перекладного словника теж може бути недоречним. Структура, обсяг і змістове наповнення словника можуть варіюватися залежно від його призначення. І однією з найдоречніших, хоча й не часто згадуваних класифікацій двомовних словників, залишається пропозиція В.Г. Гака, яка ґрунтується на типі користувача словника (чи є користувач носієм цільової чи вихідної мови, чи є він перекладачем, чи є він пересічним користувачем) [Гак 1964, с. 74-75]. На думку вченого, двомовні перекладні словники слід розділити на дві групи: словники загального користування, які орієнтовані переважно на розв'язання труднощів рецептивного етапу, та словники для перекладачів, що покликані допомагати у розв'язанні труднощів рецептивного й репродуктивного етапу.

На відміну від словників загального користування, ті, що призначені для перекладачів, окрім вирішення власне проблем рецептивного етапу, мають давати розв'язання типових проблем репродуктивного етапу, як, наприклад, синонімію та сполучуваність. У свою чергу словники загального користування мають бути спрямовані радше на проблеми рецептивної стадії перекладу, якщо вони призначені для носіїв цільової мови, та на продуктивні, якщо їхніми користувачами передбачаються носії мови оригіналу. На жаль, більшість двомовних словників

структуровані як словники загального користування, через що деякі словникові статті виявляються занадто деталізованими, в той час як у інших бракує часом справді суттєвих даних. Ці міркування В.Г. Гака [Гак 1964, с. 74-75], які можна коротко представити у схемі, свідчать, що для певної пари двох мов, в теорії та на практиці потрібно 8 типів різних словників:

1) перекладний словник з мови А загального користування для носіїв вихідної мови;

2) перекладний словник з мови А загального користування для носіїв цільової мови;

3) перекладний словник з мови А для перекладачів, носіїв вихідної мови;

4) перекладний словник з мови А для перекладачів, носіїв цільової мови;

5) перекладний словник з мови В загального користування для носіїв вихідної мови;

6) перекладний словник з мови В загального користування для носіїв цільової мови;

7) перекладний словник з мови В для перекладачів, носіїв вихідної мови;

8) перекладний словник з мови В для перекладачів, носіїв цільової мови.

Показана восьмичленна парадигма свідчить, що словники, які забезпечують напрямок перекладу з рідної мови на іноземну, аж ніяк не є дзеркальним відбитком зворотних словників (з іноземної мови на рідну). Прямі й зворотні словники відрізняються не лише напрямком перекладу, а й змістовим наповненням. Хоча зазвичай у двомовних перекладних словниках не сказано, чи призначені вони для носіїв вихідної чи цільової мов, за замовчуванням можна вважати, що вони орієнтовані на носіїв країни, де словник публікується. Перекладачам минулих десятиліть цілком знайома ситуація, коли англо-російський словник, придбаний у Великій Британії, суттєво відрізнявся від англо-російських словників, розроблених у СРСР. Нині, коли чимало словників розміщуються на сайтах, під час їх розробки було б доцільно враховувати потенційного користувача словника як у загальному задумі структури, так і в можливості передбачення гнучкості інтерфейсу за наявності належних опцій і налаштувань.

На сьогодні найпоширенішими є словники загального користування, еволюція яких відбувається переважно екстенсивним шляхом: за рахунок розширення словникових

статей, поповнення варіантів перекладу пропозиціями користувачів, прикладами з виконаних перекладів. Через відсутність чіткої градації до словників загального користування подекуди потрапляють зайві деталі, скажімо, синонімічні ряди, які носій цільової мови легко може надати і сам, або ж сталі вирази, які можна перекласти за допомогою елементарної кальки: “guerra civil” – “громадянська війна”, “guerra agresiva” – “агресивна війна”, “guerra fratricida” – “братовбивча війна”, “guerra fría” – “холодна війна”, “guerra injusta” – “несправедлива війна”, “guerra colonial” – “колоніальна війна”, “guerra púnica” – “пунічна війна”.

Окрім зазначеної класифікації двомовних словників, що ґрунтується на типі користувача, можна згадати й багато інших критеріїв, перерахованих у наступному розділі щодо одномовних словників. Вони чинні також для класифікації двомовних словників, тому з міркувань стислості їх тут не наводимо. Попри розроблені критерії оптимізації, сучасні двомовні перекладні словники часто поєднують у собі риси словників різних типів. На практиці перекладачі користуються різними двомовними словниками, залежно від типу перекладацької проблеми. Скажімо, для перекладу терміна, а особливо, термінологічного словосполучення пошук доцільно починати з двомовного перекладного словника: оскільки йдеться про термін, який часто складається з декількох графічних слів і супроводжується відповідною тематичною поміткою, для пошуку необхідні лічені секунди. Натомість, коли виникають проблеми з формулюванням стилістично вдалого варіанта перекладу якоїсь загальновживаної лексичної одиниці цільовою мовою, інколи справді доречно користуватися словниками, побудованими на основі паралельних текстів, таких як *Glosbe*, *Linguee*, *Reverso Context* та інш., не говорячи вже про широке застосування одномовних словників (ономасіологічних, тезаурусних та інших).

Таким чином, сучасні електронні двомовні перекладні словники характеризуються кількісною та якісною еволюцією. Кількісне зростання виявляється у збільшенні обсягу словника, деталізації і розширенні словникових статей додатковими лексико-семантичними варіантами і пропозиціями перекладу. Структурно-якісні зміни в електронних словниках спричиняються до виникнення нових об'єктів лінгвістичного конструювання, які становлять інтерес і для науковців, і для перекладачів-практиків. Найпоширенішими прикладами лінгвістичного конструювання у двомовній лексикографії є побудова словників на основі

паралельних текстів, створення інверсійних та зведених словників. З огляду на велике розмаїття словників та їхнє структурне розгалуження, від перекладачів вимагаються додаткові вміння й навички, зокрема, здатність швидко обирати необхідний словник або відповідну структурну частину залежно від типу перекладацької проблеми.

На практиці двомовні перекладні словники часто й не розв'язують труднощів перекладу. Окрім проблем оптимального укладання і правильної організації двомовних перекладних словників, існують об'єктивні передумови їхньої недосконалості (докладніше – у статті “Двомовні перекладні словники: причини недосконалості та перспективи оптимізації” [Фокін 2017]). Тому професіоналам перекладу у своїй діяльності частіше доводиться користуватися одномовними лексикографічними ресурсами, які докладніше описано у наступному підрозділі.

1.4.2. Види одномовних словників та їхнє значення для практики перекладу

З огляду на об'єктивні розбіжності між мовами, двомовні перекладні словники не можуть забезпечити перекладача ключами для вирішення всіх необхідних завдань. У той самий час потенціал одномовних словників, а надто електронних, у розв'язанні труднощів на рецептивній стадії і навіть репродуктивному етапі перекладу справді недооцінений. Тому у цьому підрозділі зосереджуємося спеціально на різноманітних типах одномовних словників та їхньому застосуванні у виконанні практичних завдань перекладу.

Кожен перекладач, яким би скромним не був його досвід, визнає, що навіть найдетальніші двомовні перекладні словники не можуть зарадити у розв'язанні більшості лексичних труднощів, які виникають в його праці. І хоча часом вдається повноцінно передати один з лексико-семантичних варіантів певним словом іншої мови⁹, навряд чи збігатиметься об'єм семантичних ознак поняття (інтенціонал) та сукупності позначуваних денотатів (екстенціонал) вихідної та цільової лексем. До прикладу, в

⁹Ми визнаємо, що в словнику можуть міститися одиниці різних рівнів (морфеми, синтаксеми, терміни, концепти, прагмати тощо), хоча в цьому розділі ми зосереджуємось виключно на лексичному рівні з огляду на його пріоритетну значущість у словникарстві і важливість для перекладача.

романських мовах термін “juez” (ісп.), “juge” (фр.), “giudice” (італ.) залежно від контексту співвідноситься з українськими термінами “суддя” і “слідчий”. Останнє поняття в зазначених мовах можна виражати також словосполученням (“juez instructor” – іспанською, “giudice d’istruzione” – італійською, “juge d’instruction” – французькою), однак на практиці визначник часто не вживається, що очевидно показує відмінність інтенціоналу й екстенціоналу термінів “суддя” та “juez/giudice/juge”. Ще один приклад яскраво демонструє, що лексеми, які позначають однаковий денотат, не так вже й часто збігаються в усьому екстенціоналі. Іменники “chair” (англ.) і “стілець”, на перший погляд, – пара слів, що характеризується високим ступенем еквівалентності. Однак вони мають різний екстенціонал, оскільки “chair” означає, окрім поняття “стілець”, також “кафедра”, так само як іспанській іменник “mesa” – не лише “стіл”, а й “президія”. Це свідчить про те, що переклад одного слова іншим словом може бути актуальним лише у певному контексті. Якісні словники враховують імовірні контексти, проте, не вичерпно.

Зазначена проблема у двомовних словниках часто вирішується за рахунок різних перекладів для окремих лексико-семантичних варіантів. Однак навіть у такому випадку контекстуально зумовлений перенос значення, часто okazіональний, все одно залишається поза увагою словників. Приміром, в українській мові зафіксовано випадки вживання іменника “шумахер” на позначення водія, що непомірно збільшує швидкість руху, часто з ризиком для життя; епонім “сусанін” в іронічному контексті може позначати людину, яка супроводжує інших до пункту призначення хибним шляхом, відповідно до деяких випадків вживання, зареєстрованих у корпусі української мови [Corpora Collection of Leipzig University]. Якщо постать Сусаніна, цілком логічно, краще відома в Україні, ніж в романо-германському мовному ареалі, то Міхаель Шумахер – німецький гонщик, і цілком можна очікувати, що okazіональне творення такого епоніма в романо-германських мовах було б зрозумілим. Однак в іспаномовній традиції, наприклад, з еквівалентним значенням використовується епонім “Fittipaldi”. Якщо такі епонімічні значення не зафіксовано в одномовних тлумачних словниках, тим паче їх немає у двомовних перекладних словниках. Отже, у двомовних словниках неминуча величезна кількість білих плям, які ніколи не буде можливо заповнити остаточно.

Окрім okazіональних переносних значень, виникає питання, чи вжито слово в сучасному, чи в застарілому значенні. Якщо лексема

вжита в застарілому тексті, вона може прочитуватися зі значенням, сучасним для читача, значенням, сучасним для автора або навіть зі значенням, застарілим для автора. Цій проблематиці, зокрема, присвятив свою статтю С.Флорин “Чем словари не удовлетворяют переводчика?” [Флорин 1975].

Однією з найпоширеніших класифікацій словників є їхній обсяг. Однак у комп'ютерну еру поняття великих, малих і середніх словників набуває дедалі більш відносного характеру, тому детально його не розглядаємо. Однак надзвичайно багатша палітра мислимих класифікацій представлена у множині одномовних словників з огляду на їхню більшу кількість.

Так, наприклад, за своєю функцією словники розподіляються на тлумачні, орфоепічні, орфографічні, навчальні тощо. Відповідно до рівня мови, вони бувають лексичними, фразеологічними, граматичними. Деякі словники класифікуються за хронологічним критерієм (словники архаїзмів, неологізмів, етимологічний словник), за соціолінгвістичними характеристиками (словник діалектів, варіантів мови, жаргону тощо). Окрім зазначеного, додатковими суттєвими критеріями є частота вживання лексем, тип представленої лексики, галузь або сфера знань, яку охоплює словник тощо.

Одним з провідних критеріїв класифікації словників мала б бути їхня характеристика з погляду співвідношення між формою і змістом. Таким чином, словники можна розподіляти на ті, що спрямовані від форми до змісту (семасіологічні), і ті, що орієнтовані від змісту до форми (ономасіологічні). Розглянемо кожен з цих двох типів детальніше.

Семасіологічні словники становлять переважну більшість лексикографічних ресурсів, тому, відповідно, визначник “семасіологічний” навіть не фігурує у їхніх назвах. Насамперед семасіологічні словники представлені тлумачними словниками. Розглянемо далі деякі найпопулярніші тлумачні словники та прокоментуємо їхні характеристики.

Зокрема, до семасіологічних словників належить “Словник української мови” в одинадцяти томах (за редакцією І.К. Білодіда, також доступний онлайн [СУМ]); готується до друку й “Словник української мови у 20-ти томах”, наразі видано дев'ять томів, доступні також онлайн, в яких статті розташовані за алфавітним порядком до прикметника “мішурний” включно (станом на 1 травня 2019 р.).

На сайті Українського мовно-інформаційного фонду міститься ресурс “Словники України”. Словник охоплює значну кількість лексем української мови, у тому числі власних назв (також

іншомовного походження), а також усі парадигматичні форми лексем з указанням наголосу. Подібні функції є також у “Всесвітнього словника української мови” [Всесвітній словник української мови]; список синонімів, що генерує останній, надзвичайно детальний, оскільки, окрім синонімів першого порядку, у ньому користувачеві пропонуються також синоніми синонімів вихідного слова. Себто, синоніми другого порядку.

Серед новаторських і перспективних підходів в укладанні електронних семасіологічних словників відзначимо “Зведений словник української мови”, в якому інтегровано базу даних двадцяти словників української мови [Сніжко 2015]. Таке поєднання відносно легко здійснити з алгоритмічної точки зору, і внаслідок нього виникає новий лексикографічний об’єкт, що набуває нових, по відношенню до вихідних джерел, властивостей. Зокрема, зведений словник дає змогу оперативно знаходити усі можливі дані про лексему, порівнювати її тлумачення, проводити дослідження щодо історії мови, мовної картини світу, динаміки її зміни. На основі автоматизованої вибірки лексем, зафіксованих лише у найновіших словниках, створено словник новацій української мови [Сніжко 2016]. Зведені словники часом називають також інтегральними.

Серед тлумачних словників англійської мови окремо слід відзначити Оксфордський тлумачний словник “Oxford English Dictionary”. У ньому міститься 273 000 реєстрових слів, з яких 171 476 використовуються у сучасній мові, а 47 156 вийшли з ужитку; близько 9 500 реєстрових слів входять до словникових статей як складові частини. У словнику міститься 157 000 словосполучень і дериватів напівжирним шрифтом, а також 169 000 фразеологізмів і сталих виразів напівжирним курсивом, що в цілому складає понад 600 000 одиниць [Oxford English Dictionary].

Великий тлумачний словник італійської мови “Grande Dizionario Hoepli Italiano”, упорядкований Ґабріеллі Альдо, містить 500 000 лексем, лексико-семантичних варіантів і дефініцій; для кожної лєми підраховано частоту вживання. Словник охоплює як загальноновживану, так і спеціальну лексику (з природничих наук, інформатики, техніки), конотативно марковані лємами з соціолокальними позначками (регіональні, діалектні та жаргонні слова), іншомовні запозичення та неологізми. Пошук можна проводити як у тексті словника в цілому, так і за реєстровими словами [Grande Dizionario Hoepli Italiano].

Серед словників польської мови доступною і популярною є безкоштовна версія тлумачного словника “Словник польської мови PWN” на порталі *sjp.pwn.pl*, де також міститься корпус польської мови та є збірка цікавих фактів щодо значення, специфіки вживання й етимології деяких лексичних одиниць [Słownik języka polskiego PWN].

Тлумачний словник Французької академії “Dictionnaire de l’Academie Française” містить 60 000 словникових статей, лексико-семантичних варіантів і дефініцій на матеріалі лексики французької мови. На сайті міститься перше, четверте, п’яте, шосте, восьме і дев’яте видання словника. У дев’ятому виданні, укладання якого почалося ще в попередньому столітті під керівництвом М. Дрюона та Е. Каррієр д’Енкос, на момент 26 січня 2017 вже охоплено лексеми включно до слова “rimbaldien”. У цьому виданні міститься на 28 000 більше лексичних одиниць, ніж у попередньому [Dictionnaire de l’Academie Française]. Насправді заявлені обсяги тлумачних словників не можна вважати показником лексичного багатства мови, адже в англomовній традиції прийнято рахувати обсяг словника кількістю усіх реєстрових слів та словосполучень, у той час як у Тлумачному словнику Французької академії прийнято рахувати обсяг відповідно до кількості словникових статей, які, в свою чергу, можуть містити десятки складних слів, словосполучень, фразеологізмів. “Насправді, такий різнобій у кількості словникових статей відбувається внаслідок того, що англійські й американські лексикографи розподіляють в окремі статті словосполучення (...), у той час як французькі лексикографи вміщують їх в одну статтю на основі ключового слова, що їх об’єднує”¹⁰ [Villers 2014, с. 54].

Для знавців іспанської мови слід назвати, по-перше, словник Королівської академії іспанської мови. Від 2014 р. донині на сайті доступне 23-тє видання словника, що містить 93 000 реєстрових слів (у той час як у попередньому було 88 000 реєстрових слів). Остання версія словника приведена у відповідність з іншими виданнями Королівської академії іспанської мови, такими як “Нова граматика”, “Орфографія”, “Словник американізмів” [DRAE]. Словник ґрунтовний і в кількісному, і в якісному плані, оскільки охоплює як активно вживані лексеми, так і застарілі, діалектні, стилістично знижені слова. У сучасній версії словника онлайн можна проводити пошук за реєстровим словом, словоформою, за початком або кінцівкою реєстрового слова, анаграми реєстрового

¹⁰Переклад наш – С.Ф.

слова. У тих випадках, коли необхідно отримати оперативну консультацію щодо значення слова, в цілому можна рекомендувати скористатися менш деталізованими словниками. Менший обсяг словника не означає нижчу якість. Словник "Clave" містить менше лексико-семантичних варіантів, соціолокальних та ареальних характеристик, однак у ньому чітко представлено основні значення слів, фразеологізмів та сталих словосполучень. На сьогодні у словнику "Clave" міститься понад 80 000 словникових статей [Clave].

Ономасіологічні словники, попри незначну поширеність, надзвичайно цінне знаряддя для професійного перекладача, тому що вони призначені для розв'язання труднощів, пов'язаних з вираженням певного змісту на репродуктивному етапі перекладу. Вони упорядковані саме за змістовим критерієм. Словники цього типу, менш відомі широкому загалу і навіть філологічній спільноті, детально розглядаються у пунктах 1.6., 1.7. До ономасіологічних словників часто відносять також тезауруси та словники "споріднених ідей", адже вони справді побудовані за принципом "від змісту до форми". Докладніше їх розглянуто у пункті 1.6.3.

Словники труднощів. Незамінним довідковим джерелом для перекладачів слід вважати словники, що допомагають розв'язувати найтипівіші сумніви щодо норми та узусу тих чи інших одиниць у мові та мовленні. Часом їх називають словниками сумнівів, інколи – словниками помилок; в українській лексикографічній традиції їх прийнято називати словниками культури. З-поміж них слід згадати "Словник сумнівів іспанської мови Королівської академії іспанської мови" [DPD], "Словник сумнівів і труднощів іспанської мови" М. Секо [Seco], словник Фундації "Швидка допомога з іспанської мови" FUNDEU [Buscador], довідник Б.Д. Антоненко-Давидовича "Як ми говоримо?" [Антоненко-Давидович 1997], "Словник-довідник з культури української мови" [Гринчишин 1996], упорядкований Д. Гринчишиним, А. Капелюшним, О. Сербенською, З. Терлак. Подібним застосуванням характеризуються також орфографічні та орфоепічні словники. Безумовно, кожен словник має містити орфографічно правильні словформи, що супроводжуються коментарями орфоепічних своєрідностей, але найчастіше це стосується початкової словникової форми, зазначеної у лемі, тоді як спеціалізований орфоепічний словник містить усі словформи різних відмінків, відмін та інших граматичних категорій та їхніх форм. Тут слід згадати ресурс "Словники України" онлайн, в якому вичерпно представлено

граматичні парадигми усіх наявних лексем з зазначенням наголосу для кожної словоформи.

Словники відомих цитат, висловів, афоризмів можуть полегшити працю перекладача, завдяки яким їм не доведеться вдаватися до довгого пошуку, адже при перекладі відомих висловлювань прийнято використовувати прецедентний переклад, який не завжди доступний у відкритих джерелах.

У цілому з'являється щоразу більше нелексичних словників. У 1981 р. Ю.О. Караулов узагальнив тенденції і параметри, відповідно до яких нові наукові результати призводять до прагнення представляти різні мовні плани і рівні у вигляді словників [Караулов 1981, с.5]. Саме тому цілком логічним є виникнення словників культури, словників прагматичного рівня (до яких, наприклад, слід віднести і словники мовленнєвого етикету, розмовники, словники ситуативних виразів тощо), і, в перспективі, дискурсивних словників.

Крім того, існує ще велика кількість одномовних словників, які, можливо, рідше потрібні перекладачам, як, наприклад, етимологічні та деякі інші, які тут не аналізуємо. В українській лексикографічній школі є суттєві надбання у галузі укладання словників словотвірної та морфемної будови слова, зокрема, “Словник афіксальних морфем української мови” (упорядники – Н.Ф. Клименко, Є.А. Карпіловська, В.С. Карпіловський, Т.І. Недозим).

Для наукових та дидактичних цілей особливого значення набувають **частотні словники**. Серед них можна виокремити ті, що упорядковано людиною (до прикладу, “Частотний словник сучасної української художньої прози”, “Обернений частотний словник української художньої прози” за загальною редакцією В.С. Перебийніс, а також за участі Т.О. Грязнухіної, Н.П. Дарчук, Є.А. Карпіловської, Н.Ф. Клименко, Л.І. Комарової, В.І. Критської, Л.В. Орлової). Сучасні частотні словники генеруються переважно автоматично за результатом обробки корпусу (приміром, у корпусі Королівської академії іспанської мови можна переглянути список найуживаніших 1 000, 5 000, 10 000 словоформ, а також повний список словоформ у порядку зменшення частоти з усього корпусу [Listado de frecuencias]); на лінгвістичному порталі *mova.info* є можливість згенерувати частотний словник словоформ як усього дослідницького корпусу української мови, так і його частини: законодавчих текстів, публіцистичних текстів, наукових текстів, поетичних текстів, текстів художньої прози, фольклорних текстів, а також і для окремих їхніх складових [Частотні].

Такі словники знайшли своє застосування у методиці викладання іноземних мов та у стенографії [Перебийніс 2014, с. 48]. Поза тим, “частотні словники дозволяють вирішити чимало мовознавчих проблем загальнотеоретичного й філософського характеру, наприклад: семіотичного закону простоти (прості за будовою одиниці вживаються найчастіше), зв'язок лексичної та граматичної семантики: за найвищою частотою певного відмінка іменника чи певної часової форми дієслова можна визначити його семантику: найвища частота називного відмінка свідчить про те, що іменник позначає живу істоту, а найвища частота *ing*-форми англійського дієслова свідчить про те, що він виражає супутню дію”¹¹ [Perebeynoss, Khidekheh 2004].

У цілому детальна й систематична класифікація словників різних типів за десятками можливих критеріїв зібрано у праці В.І. Перебийніс, В.М. Сорокіна “Традиційна та комп'ютерна лексикографія”.

1.4.3. Характеристики електронного конкордансу

Конкорданс початково виник як список лексем, вжитих у певному тексті або сукупності текстів. Один з найдавніших конкордансів знайдено ще у Сувойських рукописах, що датуються періодом 408 р. до н.е. – 318 р. н.е. Можливо, найпопулярнішими в історії людства були конкорданси Біблії, наприклад, упорядкований Джеймсом Стронґом (грецькою мовою та івритом). Конкорданс зазвичай складався зі списку словоформ, наявних у певного автора, що супроводжувалися посиланням на контекст вживання цього слова. У комп'ютерну еру обсяги опрацьовуваних текстів збільшилися, тому створення списку всіх словоформ, словосполучень було б непомірно великим, та й у цьому зазвичай немає потреби: лише для однієї лексеми може бути знайдено сотні та тисячі контекстів. Тому поняття конкордансу змінилися, і його дефініцію слід уточнити: конкорданс – це список контекстів для наданих у запиті користувача текстових фрагментів (словоформ, лексем, словосполучень, речень) відібраних з корпусу текстів.

Раніше конкорданси упорядковувалися вручну, а тепер майже завжди генеруються автоматично комп'ютером і складаються з переліку відповідних фрагментів з певної сукупності текстів. Кожне

¹¹Переклад наш – С.Ф.

слово у них представлено з мінімальним попереднім та подальшим контекстом.

Табл. 1. Приклад оформлення конкордансу, згенерованого комп'ютером

como si tocasse una guitarra . De vez en cuando, el hombre se sabe Luego tomó una guitarra de manos de otro de sus siquiera una mala guitarra , se ha convenido en que los 8 acompaña además con la guitarra . – A las nueve en punto en la con una guitarra : hay a su lado un mozo chupando lero que tocaba la guitarra como con uñas de pedernal y altura en forma de guitarra , previa su filiación regular y está trasteando la guitarra ?... Quiebrahuesitos, en el
--

[CREA].

Недолік конкордансу, у порівнянні зі словниками сполучуваності, в тому, що на запит користувача в ньому може з'являтися забагато зайвих даних, які користувачеві доводиться аналітично опрацьовувати та відкидати.

Сучасні конкорданси генеруються автоматично і в такий спосіб, щоб одразу було видно і правий, і лівий контексти. З огляду на колосальні обсяги нинішніх корпусів текстів, для оформлення конкордансу відповідно до запиту користувача потрібні десятки сторінок, і в Табл. 1. показано, що на видимій частині екрану вміщується заledве одна словоформа з контекстом. При натисненні на гіперпосилання відкривається ширший контекст вживання заданого слова.

Найчастіше електронний конкорданс створюється комп'ютером автоматично. Він є результатом опрацювання корпусу текстів відповідно до запиту користувача, отже, електронний конкорданс характеризується високою динамічністю та інтерактивністю. Його наповнення й характеристики суттєво залежать від програмного забезпечення, його інтерфейсу й типу корпусу текстів. Тому докладніше про специфіку налаштування фільтрів пошуку й укладання конкордансу на основі корпусу текстів сказано у підрозділі 3.2 у зв'язку з використанням корпусних ресурсів.

1.5. Проблема розрізнення значень лексем у комп'ютерних словниках

Проблема виокремлення лексико-семантичних варіантів лексеми та ступеня їхньої деталізації залишається актуальною, тому що саме собою значення дуже динамічне. Оскільки саме електронні словники характеризуються більшою динамічністю у порівнянні з паперовими, для них відкриваються додаткові можливості вирішення цієї проблеми.

Залежно від потреби мовця можливі різні механізми творення значень слів: класифікаційний, симілятивний (метафоричний) та імплікаційний (метонімічний) [Семантична деривація, 2006]. З іншого боку, творення нових значень також можна класифікувати за такими характеристиками: звуження значення, розширення значення, зсув і переміщення значення, посилення значення, послаблення значення, погіршення значення і покращення значення [Smouchtchynska 2015, с. 173]. Оскільки перенесення значень, а також механізм їхнього розпізнавання пов'язаний з суб'єктивною оцінкою, наразі важко застосувати його як критерій виокремлення значення у сфері комп'ютерних технологій. Якщо значення зафіксоване в комп'ютерному словнику, його ідентифікація не становить нездоланну проблему, однак проблематичним є розпізнавання саме okazionalно створюваних значень.

Детальне розрізнення значень у словниках – суттєве питання ще й тому, що в еквівалентних різномовних лексем набір значень може не збігатися. Для перекладачів існує ризик мимовільного перенесення такого значення на іноземну мову. Це помітно, зокрема, у назвах ігор, коли в різних мовах на позначення ролей та уявних предметів використовуються різні образи, зокрема, в дитячих іграх: "зіпсований телефон" – "Chinese whispers" (англ.), "teléfono descompuesto, teléfono escarchado" (ісп.); "грати у дочку-матері" – "jugar a las familias" (ісп.), "giocare alla famiglia felice" (італ.); "грати у козаки-розбійники" – "to play cops and robbers" (англ.), "jugar a policías y ladrones" (ісп.), "jouer aux gendarms et aux voleurs" (фр.).

Розглянемо ще один приклад. У романських мовах іменник на позначення заліза може метонімічно позначати і будь-який предмет з заліза:

*"Un motorista que circulaba por la calle Princesa también resultó herido leve al ser golpeado por **un hierro** que arrastró el viento, según señaló el Samur. El herido fue trasladado al Clínico" (El Mundo, 07/02/1996, icn.).*

*Giacomo veniva "tenagliato" (una tortura che consiste nello strappare con **un ferro** arroventato dei brani di carne da petto e schiena) prima della morte, mentre le due donne facevano testamento e indicavano il luogo per la loro sepoltura: (<http://www.fantasmitalia.it/bea.htm>, ital.).*

*Des petits tas de pâte compacte sont posés sur **un fer** et l' autre est rabattu dessus. Lorsque la gaufre est cuite on ouvre les fers et la gaufre gonfle. Elle est alors coupée en deux (à l' endroit de la bulle) et fourrée chaude (<http://fredkitchen.canalblog.com/fpp>.)*

Прикметно й те, що цей лексико-семантичний варіант не фігурує ані в іспансько-українському словнику за ред. В. Бусела [Іспансько-український, 2013], ані навіть у тлумачних академічних словниках іспанської та французької мов, що свідчить про живий характер метонімії. Втім, цей відтінок значення важливий, адже для нас іменник "залізо" позначає речовину, а не предмет з заліза, у той час як існує розмовний іменник "залізяка". Аналогічно важливою виявляється розрізнення сем "предмет" – "вміст предмета" для іменника "ложка", що залежно від наявності першого чи другого семантичного компонента в англійській, іспанській, французькій мовах позначатиметься термінами "spoon/spoonful", "cuchara/cucharada" і "cuillère/cuillerée" відповідно, хоча в італійській мові "cuchiaio" номінує і ложку як столовий прибор, і її вміст, так само, як і в українській.

Розрізнення відповідних сем в електронних словниках можливе різними шляхами: спеціальною розміткою, а також за допомогою формального та статистичного аналізу.

Розмітка передбачає явну вказівку на значення або його компонент шляхом системи спеціальних тегів. Окрім розрізнення значення в словнику, розмітка відіграє суттєву роль під час обробки текстів у мережі Інтернет з метою вивчення їхніх соціологічних, політичних, комерційних та інших характеристик; у корпусах текстів – для оптимізації проведення досліджень певних рис тексту. Докладніше про семантичну розмітку сказано у підрозділі 5.3.2.

Формальний аналіз графічного слова та його контекстуального оточення дає можливість аналізувати його семантику. Відповідно до найближчого графічного слова, що міститься ліворуч або

праворуч, можна розуміти, в якому лексико-семантичному значенні вжито саму лексему. Виокремлення лексико-семантичного варіанта засновано на аналізі біграм. Іменник "машина" в поєднанні з дієсловом "керувати" вірогідніше означатиме "автомобіль", а в поєднанні з виразом "встановлювати програмне забезпечення" означатиме "комп'ютер" або "предмет побутової техніки". Однак задля встановлення кореляції між компонентом біграми і лексико-семантичним варіантом необхідна спеціальна база даних. Звичайно, біграмою вважається більш-менш типове поєднання слів; з метою виключення випадкових сполучень необхідно проводити статистичний аналіз. Докладніше про аналіз значення на основі біграм можна прочитати у підрозділах 5.3.1, 5.3.2.

Важко заздалегідь перебачити, які взагалі лексико-семантичні варіанти можуть виявитися за певних обставин релевантними. Відповідно, в іменнику "книга", можливо, при перекладі на якісь мови варто розрізняти такі значення: "книга як предмет", "книга як джерело знань", "книга друкована", "книга, в якій передбачені записи" і т. інш. Для українсько-англійського перекладу розрізнення цих смислових відтінків переважно не актуально (хоча книга з записами може називатися "register", а не "book"), тому для виокремлення значень слова потрібен певний спільний знаменник та розумна межа. Іспанський учений, лексикограф світової величини Х. Касарес зазначає: "розподіл на галузі, підгалузі і гілочки, доведений до крайнощів, серйозно погіршує загальну картину, хоча, з іншого боку, ліпше висвітлює генеалогію кожного лексико-семантичного варіанта. Натомість, недоліком надмірного узагальнення, зокрема, є те, що воно змушує відмовитися від багатьох ілюстративних цитат, часто блискучо сформульованих, і значення недоладно змішуються між собою, поза тим, втрачається можливість спостерігати нещодавні процеси виокремлення нових значень слова. Тому недоречно встановлювати якісь правила на більшість можливих випадків, отже, маємо задовольнятися хіба що деякими емпіричними принципами, що слугуватимуть загальним орієнтиром"¹² [Casares 1950, с. 57-58].

Тим не менш, лексикограф виокремлює деякі емпіричні критерії, за якими можна розрізняти ті чи інші значення лексеми: "1) коли в вихідній мові, з якої потрапило слово, у нього було декілька значень, і ці значення більш-менш закріпилися в іншій мові; 2) коли поняття, що має широкі логічні рамки, може бути

¹²Переклад наш – С.Ф.

розтлумачено в вузких рамках, воно має розглядатися як окреме значення (нині це описують у термінах відношень гіперо- гіпонімії); 3) коли okazіональне вживання стає регулярним; 4) переносні значення не слід розглядати як окремі лексико-семантичні варіанти; 5) коли означення починає використовуватися без означуваного, необхідно переконатися, що цей семантичний процес завершено (в іспанській мові субстантивований прикметник “acorazado” – “броньований” вживається замість терміна “buque acorazado” – “броньоване судно, броненосець”) [Casares 1950, с. 64; Stock 2008, с. 156-158].

П.Ф. Сток зазначає, що з-поміж лексико-семантичних варіантів слід розрізняти чітко виокремлювані та нечітко виокремлювані значення. До чітко виокремлюваних дослідниця відносить ті значення, різниця між якими виявляється у відмінних синтаксичних властивостях, колокаціях, а також у прикладах вживання, в яких допускаються різні прочитання [Stock 2008, с. 156-157]; до нечітко виокремлюваних значень слід віднести переносні вживання, а також випадки, в яких можливе вживання слова на межі різних значень [Stock 2008, с. 158-159]. Авторка підкреслює, що у класичних тлумачних словниках розрізнення значень слова зазвичай велося за історичним принципом, що не виключається як можливий метод, однак з появою словників, що представляють лексику у синхронії, поняття лексико-семантичного варіанта, натомість, можна вводити незалежно від діахронічних зв'язків, а ґрунтуватися слід на колокації, синтаксичних зв'язках та колігації¹³ [Stock 2008, с. 160]. Цей підхід видається продуктивним і в світлі цифрових технологій, адже дані про колокацію, синтаксичні зв'язки і колігацію піддаються формалізації, і їх можна добувати з корпусів текстів шляхом автоматичного аналізу.

Е.Ф. Скороходько пропонує алгоритм виокремлення семантичних складників зі словникових статей тлумачного словника; до них дослідник відносить не всі номінативні (повнозначні) слова, що містяться в глосі, а ті, які, не можуть бути розтлумачені без кругових дефініцій. Цей алгоритм теж може бути формалізований у вигляді програмного коду. Однак

¹³Колігація – типова сполучуваність двох одиниць за граматичними показниками: перехідне дієслово зі знахідним відмінком іменника, узгодження в роді, відмінку й числі прикметника з іменником тощо. На відміну від колігації, колокація перебдачає типову лексико-семантичну сполучуваність, наприклад, різні іменники поєднуються з різними дієсловами, які за суттю є синонімами дієслова “робити”: робити замовлення, виконувати доручення, проводити семінар.

сам метод далекий від ідеального, як зазначає сам Е.Ф. Скороходько, через недосконалість словників і логічні помилки в них [Скороходько 1983, с. 151-186].

Окремим перспективним методом видається “лексичне портретування” слова, розроблений у Московському державному педагогічному інституті ім. Мориса Тореза у 60-ті роки минулого століття. Він полягає у тому, що семантичні складники виокремлюються спочатку інтуїтивно, а пізніше виділяються стрижневі категорії, відповідно до яких вносяться правки до всіх описів [Дарчук 2008, с. 243]. Цей процес можна порівняти з нанесенням множини крапок, за якими пізніше наводяться чіткі контури малюнку, де явно виокремлювані лінії посилюються, а поодинокі малоінформативні точки вилучаються як зайві.

Окрім розрізнення лексико-семантичних варіантів актуальна і проблема їхньої кількості, адже при детальному аналізі лексичної семантики кількість значеннєвих відтінків може виявитися справді невичерпною. Наприклад, тлумачний словник іспанської мови “Clave” містить 9 значень іменника “maestro” [Clave], у той час як тлумачний словник Королівської академії іспанської мови приписує йому 24 окремих значення (термінологічно коректно їх слід називати лексико-семантичними варіантами) [DRAE]. Це в жодному разі не означає, що словник “Clave” не повний. Справа в тому, що ступені деталізації значення слова можуть бути різними. І поки що не визначено, наскільки детальними мають бути тлумачні словники. Хоча здійснюються спроби знайти якусь золоту середину. П.Н. Денисов стверджує, що “природно” слову мати 7 лексико-семантичних варіантів, оскільки оперативна пам'ять людини здатна опрацьовувати 7 ± 2 одиниці одночасно [Денисов 1988, 16].

Ці міркування потрібно продовжувати й поглиблювати для розробок словників, а надто в галузі комп'ютерної лексикографії, і значний поступ простежується вже в тому, що теоретичні розробки лексикографів та здобутки інноваційних технологій органічно поєднуються і спричиняються до появи якісно нового продукту. З метою створення детального словника для комп'ютера, штучного інтелекту, електронної комунікації справді потрібен ретельний семантико-статистичний аналіз великих корпусів текстів. Докладніше проблема семантичного аналізу для розрізнення значення лексеми в комп'ютерній техніці розглядається у п. 5.3.2.

Якщо ж виходити з потреб практики перекладу, то широка деталізація, заснована на ретельному семантичному аналізі і

статистичних обґрунтуваннях, може виявитися часом зайвою і обтяжливою: якщо велика кількість значень у лексемі вихідної і цільової мов збігається, увага перекладача розпорошується на зайві деталі. Приміром, словник “Мультитран” [Словарь Мультитран] надзвичайно корисний під час перекладу сталих виразів, фразеологізмів, паремій, термінологічних словосполучень, однак при з’ясуванні значення однослівної лексеми користувачеві доводиться справді витратити значний час на читання словникової статті. В.Г. Гак підкреслює, що словники для перекладачів принципово відрізняються за структурою і за змістом від словників загального користування [Гак 1964, с. 76-78]. Утім, у сучасних словниках підхід змішаний: вони орієнтовані і на пересічних користувачів, і на перекладачів-професіоналів; для перших словники виявляються надто детальними і громіздкими, а для останніх – неповними. І все ж маємо визнати, що перекладач – це людина, що пройшла ретельну філологічну підготовку, яка знає основи лексикології, стилістики, термінології, а також, завдяки досвіду, начитаності, інтенсивній практиці має розвинену філологічну інтуїцію, достатню, щоб цілком бути здатною прогнозувати нове значення лексеми на основі власних емпіричних спостережень. А навіть якби й існував дуже деталізований словник, перекладачам усе одно періодично доводилося б самотійно розв’язувати труднощі розуміння різного роду okazіоналізмів. Навички розуміння нових, свіже створених значень, таким чином, перекладачеві дуже потрібні, однак вони цілком можуть спрацювати й під час розуміння значення лексем, які давно існують і закріпилися в мові, і які він пізнає буквально під час виконання перекладу: у такий спосіб значно заощаджується час на пошук перекладного відповідника. Однак інтуїтивне розуміння тексту може схибити, тому й трапляються помилки, пов’язані з різного роду інтерференціями навичок.

Отже, розрізнення лексико-семантичних варіантів лексеми в електронних словниках актуально як з погляду практики перекладу, так і з точки зору розробки лексикону штучного інтелекту. Виокремлення значення може відбуватися як парадигматичним (за допомогою аналізу семантики лексеми), так і синтагматичним шляхом (за допомогою аналізу оточення). Кількість семантичних відтінків може виявитися практично невичерпною, тому, з огляду на потреби перекладу, у двомовних перекладних словниках необхідно виокремлювати насамперед ті

лексико-семантичні варіанти, при відтворенні яких існує ризик помилок внаслідок інтерференції.

1.6. Особливості структури і використання ономасіологічних словників

Переважна більшість сучасних словників має семасіологічний характер, хоча у їхніх назвах не обов'язково міститься цей визначник. Семасіологічним словником називається той, який спрямований від форми вираження до трактування значень лексичних одиниць (від давн.-гр. “σημασις” – “значення”). Завдяки організації, що ґрунтується на формі, семасіологічні словники опрацьовувати порівняно легко, оскільки словникові одиниці в них групуються за формальними ознаками, наприклад, за алфавітним порядком. І хоча такі словники відносно легко впорядковувати і використовувати, їхнім мінусом є те, що вони мало можуть зарадити перекладачеві на репродуктивному етапі перекладу. Якщо на рецептивному етапі дуже легко віднайти якусь незнайому лексему і прочитати пояснення щодо її значення, під час репродуктивного етапу перекладач переважно самостійно веде пошук засобів вираження в цільовій мові для певного змісту, ідеї, яку має на думці¹⁴. Оскільки немає “алфавітних списків” ідей, що можуть спасти на думку перекладачеві, сам процес пошуку перекладного відповідника, що характеризується евристичною (творчою, нестрогою) стратегією і непередбаченим результатом, може заводити перекладача у глухий кут, чи то за браком натхнення, чи то з інших причин. А переглядати всі слова в словнику у пошуку потрібного слова теж безперспективно.

Проблеми пошуку відповідних засобів вираження трапляються не лише в перекладі: носії певної мови час від часу відчувають неспроможними підібрати необхідну, достатньо точну для висловлення думки лексему, хоча з професійної точки зору найчастіше і найсвідоміше зазнають таких труднощів, напевно, перекладачі. Коли пересічний мовець зіштовхується з проблемою пошуку вдалого засобу вираження, він може відмовитися від вираження думки, дещо видозмінити її, навіть не усвідомивши цього. Натомість перекладач змушений відтворити смисл повідомлення, а за окремих обставин – і його форму, і творчий

¹⁴Перекладачі-практики добре усвідомлюють, що двомовні перекладні словники зaledве допомагають у вирішенні хіба що невеликої частини доволі типових проблем практики перекладу.

пошук вдалого рішення може стати справді виснажливим. У процесі такого пошуку перекладачі вдаються до численних словників цільової мови (тлумачних, словників синонімів і антонімів, словників сполучуваності, енциклопедій), що частково допомагають вирішити проблему. Рідше сучасні користувачі обирають тематичні словники, тезауруси, інверсійні словники, можливості яких значно ширші, ніж пошук синонімів і антонімів.

Історія ономасіологічних словників своїми витокami сягає доволі давніх часів, оскільки мислителі вже у глибокому минулому збагнули нагальну необхідність створити механізми пошуку назв для тих чи інших речей, явищ. На терені греко-латинської цивілізації одним з перших якщо не в сучасному розумінні ономасіологічних, однак, принаймні, тематичних словників стала праця Юлія Поллукса “Ономастикон”, написана приблизно у 170 р. н.е. У Середньовіччі Абан ібн Тагліб уклав галузевий тематичний словник “Кітаб” арабською мовою. Тим не менш, першим справді широко відомим ономасіологічним словником, який відносно повно охопив загальну лексику мови, у 1852 став “Тезаурус слів і словосполучень англійської мови” П.М. Роже, що складався з 15 000 одиниць, поділених на шість класів: “абстрактні відношення”, “простір”, “матерія”, “інтелект”, “воля/бажання”, “почуття”, які, у свою чергу, поділяються на підкласи, категорії і рубрики (у цілому їх приблизно тисяча). У ХХ столітті на цій ниві відзначився своєю працею “Ідеологічний”¹⁵ словник іспанської мови Х. Касареса в 1942 р., що містить 80 000 лексем. Тим не менш, ономасіологічні словники настільки мало відомі, що до їхньої допомоги перекладачі вдаються вже в зовсім поодиноких випадках.

Раніше існувала традиція називати ономасіологічні словники ідеологічними, хоча на сьогодні ця назва радше асоціюється з ідеологією, ніж з ідеєю, яка насправді лежить в основі задуму таких словників.

Найбільшим викликом для упорядників подібних словників є класифікація ідей, від яких користувач має переходити до лексичних засобів вираження. Оскільки ідеї, на відміну від графічних слів, належать до нематеріальних сутностей, існує безліч різноманітних способів їхнього вираження. Саме це і

¹⁵“Ідеологічний” тут походить від поняття “ідея”, а не “ідеологія”. Задля уникнення двозначності у нинішній час провідним визначником для словників такого типу має стати прикметник “ономасіологічний”.

зумовлює відсутність єдиної універсальної схеми для їхнього впорядкування, а також складність їхнього пошуку у схемі.

Х. Касарес, який зробив величезний внесок як до розвитку іспанської мови, так і лексикографії в цілому, згадує життєвий випадок, що нині став хрестоматійним: його приятель марно намагався пригадати, як іспанською мовою назвати людину, що народилася під час морської подорожі (в іспанській мові це називається іменником “naonato”). З подібними випадками ми зустрічаємось повсюдно і щоденно, як і сталося з персонажем М. Булгакова в оповіданні “Рушник з півнем”:

“Стоял я в таком виде: ноги окостенели, и настолько, что я смутно тут же, во дворе, мысленно перелистывал страницы учебников, тупо стараясь припомнить — существует ли действительно, или мне это померещилось во вчерашнем сне в деревне Грабиловке, болезнь, при которой у человека окостеневают мышцы? Как ее, проклятую, зовут по-латыни? Каждая из мышц этих болела нестерпимой болью, напоминающей зубную боль. О пальцах на ногах говорить не приходится — они уже не шевелились в сапогах, лежали смирно, были похожи на деревянные культяпки. Сознаю, что в порыве малодушия я проклинал шепотом медицину и свое заявление, поданное пять лет назад ректору университета. Сверху в это время сеяло, как сквозь сито. Пальто мое набухло, как губка. Пальцами правой руки я тщетно пытался ухватиться за ручку чемодана и, наконец, плюнул на мокрую траву. Пальцы мои ничего не могли хватать, и опять мне, начиненному всякими знаниями из интересных медицинских книжек, вспомнилась болезнь — паралич... “Парализис”, — отчаянно, мысленно и черт знает зачем сказал я себе” [Булгаков].

Подібні примхи пам'яті, які знайомі всім мовцям, у свій час підштовхнули лексикографів до укладання ономасіологічних словників, і не лише для загальної лексики, а й для термінології, фразеологічних одиниць.

Розуміння поняття ономасіологічності ускладнюється тим, що спостерігається значний різнобій у термінах на позначення цього виду словників: тематичний словник, ідеологічний словник, тезаурус, ідеографічний словник, лінгвістичний атлас — це ще не вичерпний список синонімів ономасіологічного словника. У зв'язку з цим вважаємо за необхідне пояснити наявні між ними розбіжності у наступних підпунктах.

1.6.1. Тематичний словник

Найпростішим і найяскравішим прикладом словника, що за суттю можна вважати ономасіологічним, є тематичний словник, наприклад: “Тематичний словник популярних українських прислів’їв та приказок з коментарями” (упорядник – В.М. Северинюк). Двомовні перекладні словники теж часто структуруються за тематичним критерієм, і тоді вони особливо корисні для туристів, для початківців, що прагнуть опанувати певний лексичний мінімум, для викладачів, які розробляють дидактичні матеріали з певної теми.

До прикладу, дуже плідним на цій ниві є Видавництво Вадима Карпенка, у якому, зокрема, видано українсько-болгарський, українсько-польський, українсько-німецький, українсько-іспанський, українсько-французький тематичні словники обсягом порядку 10 000 слів, що охоплюють такі теми: людина, родина, дім, страви, напої, офіс, комп’ютер, місто, будівництво, транспорт, мистецтво, астрономія, економіка, банк, митниця, право, держава, ЗМІ, військо та ін. За подібним принципом побудована класифікація лексики і в розмовниках. Їх легко укласти, адже створена основа для однієї пари мов може слугувати й для інших. Однак практика показує, що професіонали такими словниками не користуються, а початківцям, які хочуть ними поповнюють свої книжкові колекції через багатообіцяючу назву. І все ж ними користуватися важко й незвично: тематичний пошук довший за алфавітний, а подекуди безуспішний довгий пошук, спричинений обмеженістю таких словників, залишає відчуття безпорадності, у той час як вивчення сотень слів для опанування теми теж не викликає особливого ентузіазму.

Вельми популярними є й тематичні словники крилатих виразів і цитат, що класифіковані за тематичними категоріями. І хоча ними можуть користуватися допитливі читачі, журналісти, а також усі ті, хто жадає блиснути ерудицією, перекладачам корисніші загальні словники цитат і афоризмів, зокрема, в електронному вигляді.

Перекладачеві тематичні словники могли б бути корисними, якби за своїм обсягом вони наближалися до майже повного охоплення тематичної лексики й містили б порядку сотні тисяч статей. Щоправда, пошук за тематичним критерієм не є чітко регламентованим: окремі терміни, такі як “термоядерний синтез”, можуть входити одночасно до різних тем: “енергетика”, “фізика”,

“хімія”, “астрономія”, “екологія”, і користувачеві доведеться здогадуватися, в якій рубриці шукати необхідну лексему¹⁶.

У комп'ютерних словниках тематична класифікація виявляється більш раціональною і ефективною, оскільки в них вона може поєднуватися з алфавітною, одну й ту саму одиницю можна класифікувати за різними темами, а також формальний пошук можливо проводити не лише за початковими літерами лексеми, не лише за алфавітним порядком, а наскрізно, як за повним графічним словом, так і за його частиною або за сукупністю слів. Докладніше про особливості й переваги комп'ютерного пошуку можна прочитати у підрозділі 3.2.

1.6.2. Атлас лексики

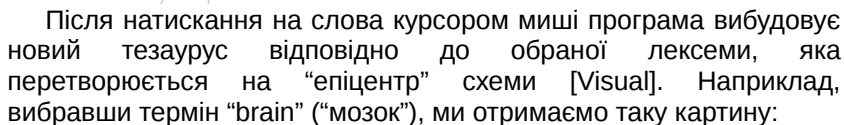
Останнім часом набувають популярності так звані “атласи лексики” або “лексичні атласи”, часто вузькотематичні, деякі з яких за суттю є ономасіологічними словниками. Відзначимо, що початково “атлас лексики” сприймався як словник, спеціалізований на ареальних характеристиках слів певної мови, як, наприклад, “Лінгвістичний атлас лексики народного побуту українських східнословобожанських говірок” К.Д. Глуховцевої [Глуховцева 2003]. У той самий час “Ідеологічний словник іспанської мови” Р. дель Мораля, альтернативною назвою якого є “Лексичний атлас іспанської мови” [Moral 2009], спирається не на географічні, а на семантичні критерії, отже, фактично є тезаурусом.

1.6.3. Тезаурус

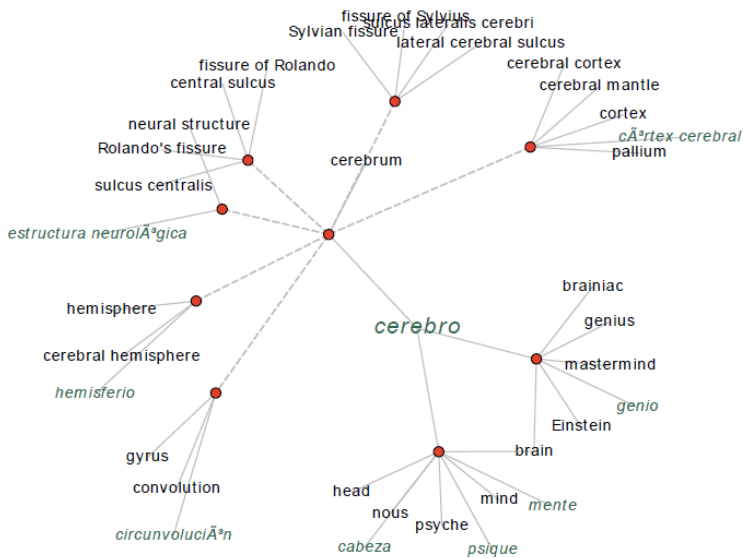
Хоча й тезауруси близькі за суттю до ономасіологічних словників, останні мають складнішу структуру; тезауруси укладати простіше, ніж власне ономасіологічний словник, тому з-поміж різних ономасіологічних ресурсів вони доволі поширені. Тезаурус містить списки лексем, об'єднаних за асоціативно-семантичними критеріями. Розглянемо деякі приклади.

¹⁶ Пошук за рубриками можна суттєво полегшити, якщо тематичні рубрики ділитимуться на підрубрики, тобто, якщо лексика буде ієрархізована за розділами та підрозділами, і власне так організовані повноцінні ономасіологічні словники, однак це пов'язано з колосальними труднощами розробки, пошуком оптимальної класифікації лексики за розділами та підрозділами; натомість тематичні словники передбачають однорівневий розподіл лексики за рубриками, які можна обирати довільно, керуючись власним життєвим досвідом і інтуїцією.

cabeza



cerebro

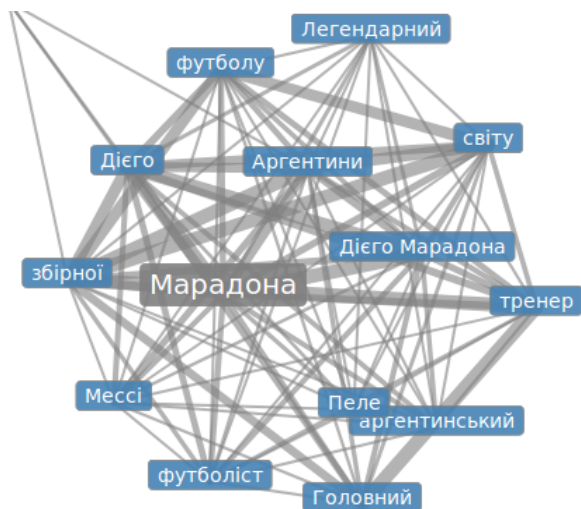


Інколи тезауруси також іменують словниками споріднених ідей. Однією з наймасштабніших у цій галузі праць є "Diccionario de ideas afines" Ф. Коррініо Переса, що охоплює більше 400 000 позицій, багато з яких вимушено повторюються, адже в ньому поєднується ономазіологічна систематизація з алфавітною, і одна й та сама лема може стосуватися різних груп споріднених ідей водночас. На матеріалі англійської мови масштабним є "Мерріам-Уебстер Тезаурус", онлайн-версія якого містить 275 000 лексем [Merriam-Webster].

Тезаурус української мови став спеціальним предметом кандидатського дисертаційного дослідження О.Б. Сірук, виконаного в лабораторії комп'ютерної лінгвістики Інституту філології КНУ імені Тараса Шевченка під керівництвом Л.А. Алексієнко [Сірук 2006].

Методику укладання тезаурусів можна і необхідно вдосконалювати, хоча й наявні сьогодні здобутки вражають своїм рівнем. Одним з методів автоматичного генерування тезаурусу на основі корпусу є аналіз взаємно близького розташування графічних слів, коли частота такого розташування перевищує випадково ймовірну величину¹⁷. До прикладу, саме так укладається тезаурус у Лейпцизькій колекції корпусів:

Схема 2. Приклад візуального тезауруса



[Corpora Collection of Leipzig University]

Як бачимо, складові тезауруса підбираються на основі надзвичайно різнопланових відношень між лексемами, і важко встановити між ними якусь чітку межу. Переважно асоціативно-семантичні зв'язки ґрунтуються на основі відношень часткового-

¹⁷ Для цього терміна іще немає усталеного відповідника в українській мові (англ., фр. — *cooccurrence*, ісп. — *coocurrencia*). Часом пропонується відповідник “сполучуваність” не передає значення терміна, адже сполучуваність обов'язково передбачає врахування семантичних властивостей мовних знаків; дистрибуція теж не є відповідником, оскільки включає в себе як випадкові, так і не випадкові величини.

цілого, видового-родового, причини-наслідку та інш. Однак цим списком відношення не вичерпуються. Зв'язати відношеннями можна можна практично будь-яку пару понять, як у відомій керролівській грі, навіть “чорний ворон” і “письмовий стіл”, “білу троянду” і “чорну жабу”. Тому розробники тезаурусу на практиці все ж встановлюють певні рамки для кількості семантичних відношень. Наприклад, “в інформаційній мові системи Перрі-Кента (Клівлендському семантичному коді) передається більша кількість іманентних відношень [...]: 1) бути підкласом; 2) бути виготовленим; 3) бути частиною; 4) складатися з; 5) бути призначеним; 6) використовувати для виконання своєї функції; 7) зазнавати дії; 8) не мати; 9) мати; 10) бути подібним” [Скороходько 1970, с. 88]. У свою чергу Н.М. Леонтьєва виокремлює так звані “енциклопедичні функції” для опису поняття, які зв'язують його відношеннями з іншими поняттями. Вчена доходить висновку, що основними функціями є “тип” та “параметр” [Леонтьєва 1978, с. 29].

Методичні проблеми укладання тезаурусу, які належить вирішити в найближчому майбутньому, пов'язані з такими аспектами:

– можливість / неможливість обмеження тезаурусних функцій (оскільки теоретично будь-які мислимі семантичні відношення можна включати до тезаурусу); деякі функції можуть взаємно перетинатися: скажімо, квазісиноніми “arm”, “hand” (обидва іменники позначають англійською мовою різні частини руки) є водночас співгіпонімами; у той самий час “присвоєння терміну тегу семантичного відношення – операція вкрай делікатна; окремі типи відношень (квазісинонімія, на відміну від синонімії та антонімії) виявляти важко¹⁸” [Jeanne Dancette, 295];

– уточнення методики встановлення семантичних відношень між компонентами тезаурусу;

– варіантність реалізацій певних тезаурусних функцій; приміром, антонімом “смерті” є поняття “життя”, однак також “народження”; вже навіть поверхневий аналіз розкриває, що ці антоніми підходять до різних лексико-семантичних варіантів поняття “смерть”: перший містить сему стану, другий – процесуальності.

¹⁸Переклад наш – С.Ф.

1.6.4. Інверсійний словник

Словник, у якому змінено місцями глосу і лему, називається інверсійним. При цьому в паперовому варіанті така перестановка, якщо й мислима, виглядала б абсурдно; в електронній версії подібна зміна цілком раціональна й уможливорює пошук не лише за дефініцією, а й за її складовими, а це дозволяє віднаходити відповідний засіб вираження певної ідеї за інтуїтивно підібраними ключовими словами. До таких словників належить “Інверсійний словник Королівської академії іспанської мови”¹⁹ [DIRAE]. Для англійської мови доступні, щонайменше, два інверсійні словники: 1) “Onelook Reverse Dictionary” [Onelook Reverse]; як заявлено на сайті, він об’єднує 19 мільйонів словникових статей з приблизно тисячі доступних онлайн словників, індексованих за допомогою пошукової системи *OneLook®*; на сайті містяться також *API* ресурси цього словника (*API* – набір широко використовуваного програмного забезпечення з відкритим кодом); 2) “Reverse Dictionary” характеризується надзвичайно простим інтерфейсом без додаткових опцій, однак він є доволі популярним серед користувачів [Reverse].

Словник української мови [СУМ] з 2016 також оснащений функцією інверсійного пошуку (на сайті вона названа “пошук в означеннях”).

Такі словники теж слід вважати ономазіологічними, адже вони спрямовані від змісту до форми. Знаючи дефініцію терміна або лексеми, ми можемо віднайти відповідну їй лему. Якщо у словнику СУМ обрати опцію “шукати в означеннях” і ввести сполучення “виоране поле”, програма знайде 6 результатів, другим з них буде “рілля”. Пошук ускладнюється відсутністю чіткого алгоритму формулювання й пошуку ідеї, а вгадати дослівно точну дефініцію шансів небагато. Слід враховувати, що одне й те саме поняття може визначатися по-різному, з різних точок зору. Саме тому видається доцільним вести такий пошук не за точними визначеннями, а за деякими ключовими словами, які потенційно

¹⁹Прикметник “інверсійний” може сприйматися неоднозначно, з огляду на неусталеність відповідної термінології: “інверсійним словником” у певних контекстах також називають той, в якому пошук лем відбувається за кінцівками графічного слова; в українській лексикографії також зустрічаються назви “зворотний словник”, “обернений словник”.

пов'язані з дефініціями. Наприклад, для пошуку терміна “дверна петля” іспанською мовою ми можемо ввести в рядок пошуку словника “DIRAE” слова “sujetar”, “puerta” (“тримати”, “двері”). Пошук дає такі результати:

postigo

/pos.ˈti.go/ - Frec. 36544

... . Cada una de las **puertas** pequeñas que hay en las ventanas o puertaventanas. 5. m. Tablero **sujeto** con bisagras o goznes ... postigo. (Del lat. posticum). 1. m. **Puerta** falsa que ordinariamente está colocada en sitio ... excusado de la casa. 2. m. **Puerta** que está fabricada en una pieza sin tener división ni más de una hoja ...

clavetear

/kla.be.ˈte.ˈar/ - Frec. 1171

... clavetear. (De clavete). 1. tr. **Sujetar** con clavos. 2. tr. Guarnecer o adornar con clavos de oro ... , plata u otro metal algo, como una caja, una **puerta**, un coche, etc. 3. tr. herretear (|| poner herretes ...

alzapaño

/al.sa.ˈpa.ɲo/ - Frec. 101

... clavadas en la pared, sirven para tener recogida la cortina hacia los lados del balcón o la **puerta**. 2. m Cada una de las tiras de tela o cordonería que, **sujetas** a los alzapauos, abrazan y tienen recogida la cortina. ...

forzar^{dpd}

/for.ˈsar/ - Frec. 318188

... para conseguir algo que habitualmente no debe ser conseguido por la fuerza. Forzar una **puerta**. 2. tr. Entrar, **sujetar** y rendir a fuerza ...

anclaje

/an.ˈkla.xe/ - Frec. 103276

... que se paga por fondear en un **puerto**. 4. m. Acción y efecto de anclar (|| **sujetar** al suelo o a otro lugar). 5 ...

alcayata

/al.ka.ˈja.ˈta/ - Frec. 2479

... de **puertas** o ventanas, sirven para **sujetar** los postigos una vez abiertos. ...

bisagra

/bi.ˈsa.ɡra/ - Frec. 39556

... y **sujetas** una a un sostén fijo y otra a la **puerta** o tapa, permiten el giro de estas. 2. f. Palo de boj ...

telar

/te.ˈlar/ - Frec. 124896

... . En automovilismo, disco de chapa embutida que **sujeta** la llanta al cubo, en las ruedas desprovistas de radios. 6. m Arq. Parte del espesor del vano de una **puerta** o ventana, más próxima al paramento exterior de la pared ...

Як бачимо, сьомий результат, "bisagra", є саме перекладом терміна “дверна петля”, який ми прагнули знайти. Отже, такий інверсійний словник слід теж віднести до ономасіологічних.

Розробники сайтів, програмного забезпечення можуть інтуїтивно знаходити шляхи розширення можливостей пошуку у словнику і втілювати їх в алгоритм програми задля привернення

уваги ширшого кола користувачів. Унаслідок цього парадоксом інверсійних словників, а також деяких інших об'єктів лінгвістичного конструювання стає той факт, що розробники цих об'єктів інколи не знають, як термінологічно точно їх назвати, у той час як теоретикам, обізнаним з термінологією, не відомо про існування цих об'єктів, оскільки їх не названо коректно.

1.6.5. Власне ономасіологічні словники

Найпотужнішим, хоча й найважчим у використанні, знаряддям, власне призначеним для підбору найліпшого способу вираження певної ідеї, є ономасіологічний словник. В основу макроструктури такого словника покладено списки дескрипторів, що згруповано відповідно до тих чи інших категорій, поділених за дрібнішими рубриками і групами. Така структура називається синоптичною схемою. Кожній групі з синоптичної схеми відповідає список лексем, що виражають заданий зміст. Загальні відмінності ономасіологічних словників, у порівнянні з семасіологічними та інверсійними, показано у таблиці 3.

Найбільший парадокс ономасіологічних словників – це система його дескрипторів, яка теоретично невичерпна. Наприклад, термін “каюта” мав би відноситися до категорії лексем, пов'язаних з дескриптором “транспорт”. Однак у такий самий спосіб він мав би бути зв'язаним і з дескриптором “море”, “подорож”, “житло”, “приміщення”. Тобто, виявляється практично неможливим створити загальну таксономію лексики, адже вона спирається радше на інтуїцію, ніж на науково обґрунтовані методи. Якщо для укладання семасіологічних словників ідеально слугує алфавітний порядок, для ономасіологічних словників немає ще такої однозначної і безпомилкової системи впорядкування ідей.

Табл. 3. Відмінності між семасіологічними, інверсійним і ономасіологічними словниками.

СЕМАСІОЛОГІЧНИЙ СЛОВНИК	
Форма	Зміст
Лема	Глоса
Палітурка	Обкладинка, оправа для зброшурованих листків книги, зошита і т. ін. з картону, матерії, шкіри тощо [СУМ]

ІНВЕРСІЙНИЙ СЛОВНИК	
Зміст	Форма
Глоса	Лема
Обкладинка, оправа для зброшурованих листків книги, зошита і т. ін. з картону, матерії, шкіри тощо [СУМ]	Палітурка

ОНОМАСІОЛОГІЧНИЙ СЛОВНИК		
Зміст (метамовні дескриптори)		Форма вираження
Загальна категорія	Субкатегорія	
Книги	Видавнича справа	Видання Палітурка Примірник Редактор Тираж Формат

З-поміж ономасіологічних словників, окрім згаданих словників П.-М. Роже, Х. Касареса, О.О.Баранова, слід згадати ще деякі, на наш погляд, недооцінені як користувачами, так і філологічною спільнотою в цілому. Важливим знаряддям для перекладача є багатомовний словник паремій, розроблений Інститутом Сервантеса для двадцяти двох мов і доступний на його сторінці ("Refranero Multilingüe del Centro Virtual Cervantes" [Refranero Multilingüe]). Назвемо також десятимовний фразеологічний словник "AUTOFRAS", в якому задіяна і українська мова [Pamies 1998], упорядкований А. Памієсом, Е.-М. Інєстою Меною, А. Бальмач, О.М. Калустовою; "Словник фразеологічних синонімів української мови" (упорядники – Н.Ф. Коломієць, Е.Ф. Регушеський), онлайн словник фразеологічних синонімів на інтернет-порталі "Rozum" [Словник фразеологічних синонімів].

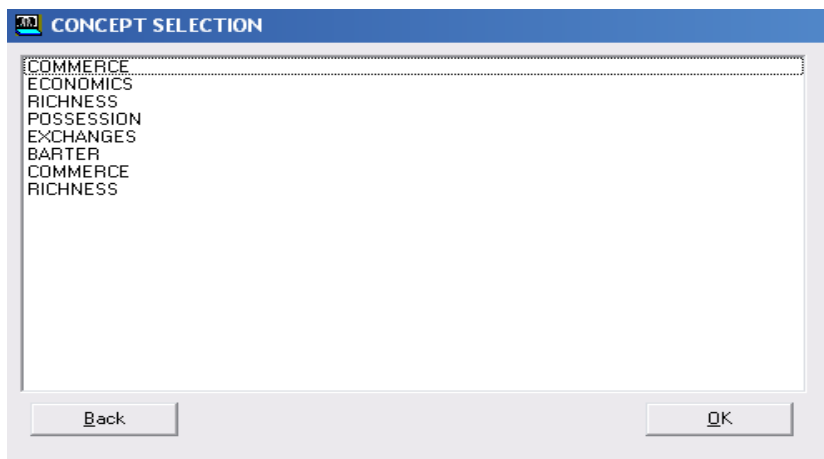
Як бачимо, саме фразеологія стала доволі плідним матеріалом для укладання ономасіологічних словників. Навіть алфавітний пошук фразеологічної одиниці у традиційному семасіологічному словнику пов'язаний зі складністю влучного підбору ключового слова, варіанта фразеологізму (які у фразеології надзвичайно поширені). Переклад фразеологізму часто еквівалентний лише умовно, тобто, або за семантичними, або прагматичними чи конотативними показниками. Тому користувачеві цінніше мати в розпорядженні певну сукупність фразеологічних одиниць (іншими словами – так би мовити “фразеологічний тезаурус”), з якої легше вибрати необхідний засіб вираження думки. Отже, з погляду фразеології семасіологічні алфавітні словники не такі вже явні суперники ономасіологічних [Pamies 1998, с. 207].

На превеликий жаль, словник “AUTOFRAS”, розроблений для операційної системи *Windows*, не працює у версіях, старших за *XP*. Тому проілюструємо основні етапи його функціонування зображеннями знімків екрану.

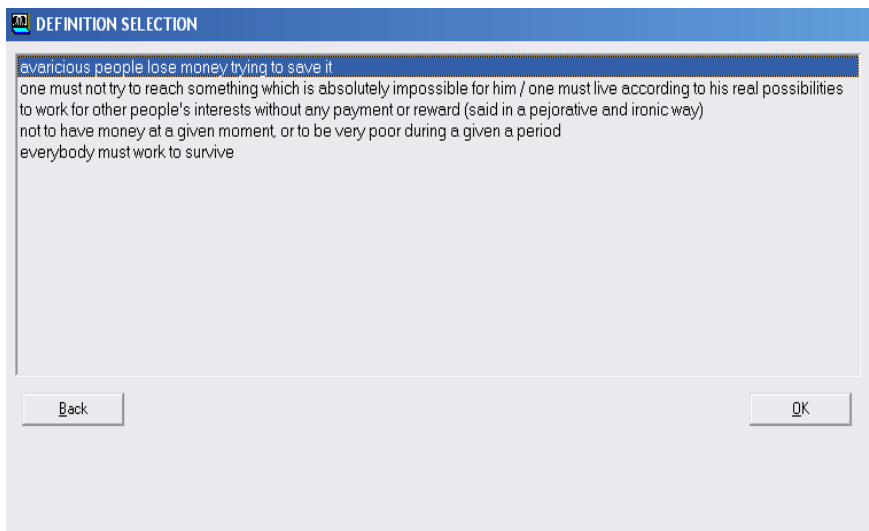
1. У словнику можливий семасіологічний (формальний) та ономасіологічний (концептуальний) пошук. Розглянемо етапи його використання на прикладі пошуку фразеологізмів, пов'язаних з поняттям “гроші”. Оскільки метамова словника – англійська, у пошуковому полі вводимо дескриптор “money”:

The screenshot shows the AUTOFRAS application window. At the top is a blue title bar with the AUTOFRAS logo and name. Below it, the window is divided into two main sections: 'CONCEPTUAL SEARCH' and 'FORMAL SEARCH'. In the 'CONCEPTUAL SEARCH' section, there is a text input field containing the word 'money'. To the right of this field are two buttons: 'OK' and 'Exit'. The 'FORMAL SEARCH' section is currently empty. Below the 'FORMAL SEARCH' section, there is a vertical list of flags representing different languages: Spanish, French, British English, Italian, Portuguese, German, Dutch, Russian, Romanian, and Czech. Each flag is followed by an empty text input field for selecting a specific phrase or translation.

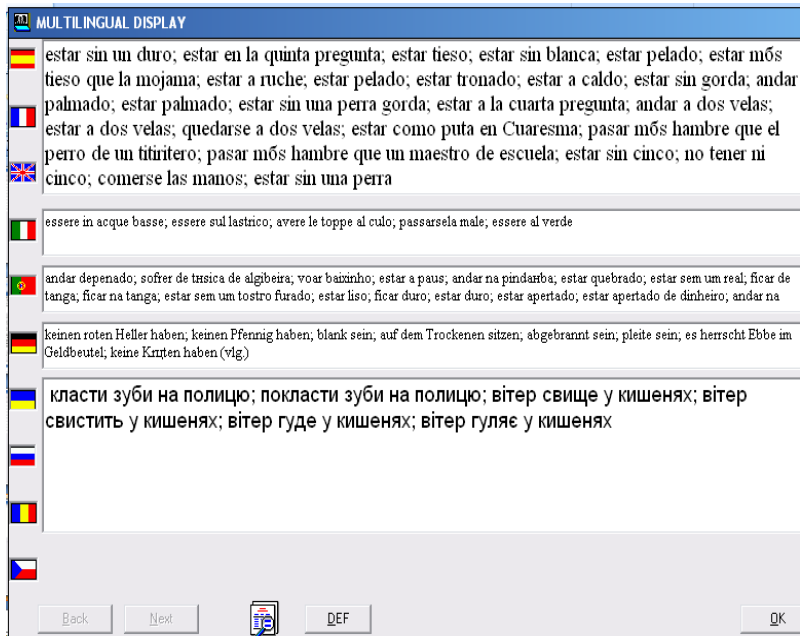
2. На здійснений запит “money” програма виводить на вибір користувача декілька тематичних класів або категорій:



3. У разі вибору “RICHNESS” програма запропонує обрати один з метамовних описів значень фразеологічних одиниць, які потенційно можуть зацікавити користувача. Далі обираємо в якості ілюстративного прикладу одну з дефініцій, а саме: “not to have mone at a given moment or to be very poor during a given period”:



4. Після вибору необхідного користувачеві метамовного опису програма віднаходить варіанти фразеологізму десятима мовами. У цілому словник міститься 1200 фразеологічних одиниць кожною мовою:

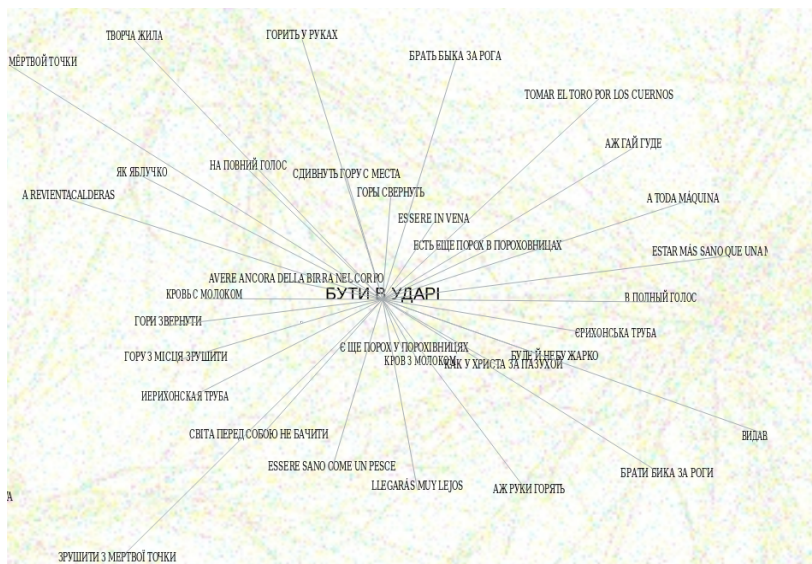


На стадії використання у тестовому режимі і корекції перебуває також багатомовний словник фразеологічних синонімів "IdeoPhrase" [IdeoPhrase]. У цьому ономасіологічному словнику можливо вести пошук фразеологічних одиниць за певною ідеєю, описаною за допомогою метамовних дескрипторів українською та англійською мовами. Далі подаються результати пошуку фразеологізмів, пов'язаних з ідеєю 'скромність':

UKR.: БЕЗ РОДУ Й ПЛЕМЕНИ
 ENG.: CUT SOMEONE DOWN TO SIZE
 ENG.: IN MY HUMBLE OPINION
 ENG.: TO HOLD ONE'S TONGUE
 FRENCH.: À MON HUMBLE AVIS
 FRENCH.: TENIR SA LANGUE
 FRENCH.: VIVRE D'AMOUR ET D'EAU FRAÎCHE
 FRENCH.: VIVRE DE L'AIR DU TEMPS
 HEBR.: טובה ציפור אחת ביד משתיים על העץ
 IT.: ABBASSARE LE ALI
 IT.: BAGNARSI IL BECCO
 IT.: ESSERE UN CINCINNATO

LAT.: OMNIA MEA MECUM PORTO
LAT.: QUANTUM SCIMUS GUTTA EST
RUS.: БЕЗ РОДА И ПЛЕМЕНИ
RUS.: ВОДУ НЕ ЗАМУТИТЬ
RUS.: ВСЁ СВОЁ НОШУ С СОБОЙ
RUS.: ДЕРЖАТЬ ЯЗЫК ЗА ЗУБАМИ
RUS.: ЗНАТЬ СВОЁ МЕСТО
RUS.: ЛУЧШЕ СИНИЦА В РУКЕ, ЧЕМ ЖУРАВЕЛЬ В НЕБЕ
RUS.: НЕ БОГИ ГОРШКИ ОБЖИГАЮТ
RUS.: НЕ НА ВЫСОТЕ
RUS.: НЕ СНИМАТЬ ЗВЁЗД С НЕБА
RUS.: С ХЛЕБА НА ВОДУ ПЕРЕБИВАТЬСЯ
RUS.: ТИШЕ ВОДЫ, НИЖЕ ТРАВЫ
SPAN.: A PALO SECO
SPAN.: CALLARSE LA BOCA
SPAN.: EN MI MODESTA OPINION
SPAN.: MÁS VALE UN PÁJARO EN MANO QUE CIENTO VOLANDO
SPAN.: MORDERSE LA LENGUA
SPAN.: PONER EN SU SITIO
SPAN.: SUSTENTARSE DEL AIRE
SPAN.: TIRAR DE LA LENGUA
SPAN.: VIVIR CON PAN Y CEBOLLAS
SPAN.: VIVIR DEL CUENTO
UKR.: БЕЗ МАНІВЦІВ
UKR.: БЕЗ ПРИКРАС
UKR.: ВОДИ НЕ ЗАМУТИТИ
UKR.: ГОВОРИТИ ЯК Є
UKR.: З ХЛІБА НА ВОДУ ПЕРЕБИВАТИСЯ
UKR.: ЗІРОК З НЕБА НЕ ЗНИМАТИ
UKR.: ЗНАТИ СВОЄ МІСЦЕ
UKR.: НА МОЮ СКРОМНУ ДУМКУ
UKR.: НАЧЕ У ВУСІ
UKR.: НЕ В УСІ ДЗВОНИ ДЗВОНИТИ
UKR.: НЕ ДОВОДИТИ ДО ГРІХА
UKR.: НЕ НА ВИСОТІ
UKR.: НЕ СВЯТІ ГОРШКИ ЛІПЛЯТЬ
UKR.: НЕ СПОЧИВАТИ НА ЛАВРАХ
UKR.: НИЖЧЕ ТРАВИ ХОДИТИ
UKR.: ПОСТАВИТИ НА МІСЦЕ
UKR.: ТРИМАТИ ЯЗИКА ЗА ЗУБАМИ
UKR.: ТЯГТИ ЗА ЯЗИК
UKR.: ЯК ТУРЕЦЬКИЙ СВЯТИЙ

На **схемі 4** приклад результатів пошуку фразеологічних синонімів для одиниці “бути в ударі”:



Як видно з проаналізованих функцій словників, ідея поєднання декількох різномовних ономазіологічних словників дає плідний ґрунт для створення і двомовних перекладних словників. Цей задум покладено й в основу “Зіставного болгарсько-російського словника”. Його упорядник, А.Я.Перванов, на основі спільних дескрипторів поєднує іменники експансивної дії і зазначає, що завдяки специфічній організації макроструктури у зіставному словнику обидва напрямки перекладу стають функціонально рівноцінними. Крім того, ідеографічний характер словника висуває певні вимоги до обсягу словника й вимагає підготовленості користувача [Перванов 2011].

Позаяк ономазіологічні словники є ще доволі рідкісним явищем, у розпорядженні перекладача часто немає таких вартісних ресурсів для пошуку засобів вираження певного значення, смислу. Однак у разі відсутності ономазіологічного словника замість нього можна скористатися будь-якою пошуковою системою мережі Інтернет, хоча й з певними обмеженнями: відповідні засоби вираження можна знайти, наприклад, у системі Google шляхом приблизного опису поняття за ключовими характеристиками.

Окрім лексичного та фразеологічного рівнів, граматики теж заслуговує на увагу ономазіології. Зокрема, створення ономазіологічного словника синтаксем стало б у пригоді широкому колу фахівців, як практиків, так і теоретиків: перекладачам, журналістам, дослідникам мови й перекладу. Можна без перебільшення сказати, що видатний український граматикист І.Р. Вихованець у своїй праці “Нариси з функціонального синтаксису української мови” [Вихованець 1992] перебуває буквально за декілька кроків від створення такого словника.

Таким чином, спільною рисою значної кількості ономазіологічних словників є їхня спрямованість від змісту до форми, що становить надзвичайно корисне знаряддя для розв’язання багатьох труднощів репродуктивного етапу перекладу. За таким принципом побудовано тематичні, інверсійні словники, тезауруси, ідеологічні словники. Найбільшим викликом створення ономазіологічних словників є розробка оптимальної синоптичної схеми-карти для змістових категорій, зручної для користувача у процесі пошуку необхідного засобу вираження. Якщо в паперовому форматі пошук за синоптичною схемою вимагає суттєвих інтелектуальних зусиль, інтуїції та ерудиції, в електронному словнику цей процес полегшується завдяки широким можливостям автоматичного пошуку. До того ж, комп’ютерна техніка дозволяє створювати словники, принципово неможливі у паперовому форматі, зокрема, інверсійні, які теж слід віднести до примітивних ономазіологічних словників.

Найбільшою проблемою розробки ономазіологічних словників є пошук оптимальної класифікації лексики, про що докладніше сказано у наступному підрозділі.

1.7. Принципи класифікації лексики в електронних ономазіологічних словниках

Лексеми в ономазіологічних словниках впорядковані не за алфавітом, а за ідеями. Немає єдиного шаблону щодо систематизації ідей за певним порядком, однак у цій галузі вже існують значні напрацювання. Як показано у попередньому підрозділі, існують спрощені варіанти ономазіологічних словників: тематичні словники, інверсійні словники, тезауруси. У тематичних словниках лексеми впорядковано за темами, однак кількість тем теоретично не обмежена, і розподіл за темою не позбавлений суб’єктивності. В інверсійних електронних словниках дефініція

передусь лемі, однак мало хто з користувачів здатний добре формулювати дефініції необхідного поняття, тим паче ще й так, щоб вони збігалися з дефініцією словника. Отже, пошук за допомогою ключових слів, описових дескрипторів, усе ж, видається раціональнішим.

У тезаурусі критерієм класифікації лексем є наявність семантичного зв'язку між лексемами, зокрема таких: синонімія, антонімія, гіпонімія, гіперонімія, голонімія, меронімія та інш.; однак невідомо наперед, який саме зв'язок дасть корисний результат пошуку, і використання тезауруса вимагає чималої лінгвістичної ерудиції та інтуїції, вміння знаходити і формулювати родові, видові поняття, взагалі добре володіти дефініційним словником (докладніше про нього – у п. 5.3.2.). У власне ономазіологічних словників система пошуку лексем детальніша і більш упорядкована, ніж у тезаурусах.

У зв'язку з нематеріальним характером ідей і складністю їхньої фіксації у матеріальній формі для пошуку дескриптора ідеї в ономазіологічних словниках використовується синоптична схема (класифікаційна схема основних категорій лексики, спрямована від найзагальніших понять до більш специфічних). Загальноприйнятою практикою з-поміж лексикографів є заснування цієї класифікації на основі розподілу царин життя. Архісевою зазвичай виступає поняття “існування”, і тоді класифікація лексики є екзистенційно-центричною. Натомість в основу синоптичної схеми “Ідеологічного словника іспанської мови” Х. Касареса покладено поняття “Бог”, тобто, схема за суттю теоцентрична. Така різниця унаочнює ту обставину, що класифікація ідей багато в чому ґрунтується на суб'єктивній картині світу, і універсальна схема навряд чи можлива. І все ж, наявні словники становлять неабияку вартість для перекладачів, хоча пошук відповідної лексики в них вимагає певних інтелектуальних зусиль. Електронним ономазіологічним словником користуватися легше, ніж паперовим, оскільки широкі можливості формального автоматичного пошуку позбавляють користувача необхідності проходити весь шлях від найзагальніших дескрипторів до більш специфічних і дозволяють віднайти ключову ідею одразу, хоча це не знімає проблему вдалого метамовного формулювання самої ідеї в словнику.

Під час користування ономазіологічним словником можна рекомендувати починати пошук з найзагальніших дескрипторів, що

збігаються з гіперонімом або голонімом лексеми, пошук якої необхідно здійснити.

Доволі просто такий пошук проводити у термінологічних ономазіологічних словниках, адже створювати таксономічну ієрархію у термінології, що ґрунтується переважно на штучно створених схемах і класифікаціях, легко. До прикладу, у двомовному термінологічному тезаурусі фахової мови психології В.В. Тарасової основні ієрархічні відношення представлені гіпогіпернімічними зв'язками [Тарасова 2013, с. 21], так само, як і в таксономічній класифікації назв біологічних видів.

На відміну від термінологічної лексики, для класифікації понять загальної лексики потрібні принципово інші категорії. І навпаки, як зазначає Ю. Караулов, ті категорії, в які доволі легко вписати стихійно створені за час еволюції мови лексеми, не підходять для групування реалій, предметів матеріальної культури, термінів точних понять [Караулов 1976, с. 333]. Тим не менш, класифікаційні схеми для загальної лексики розроблялись і розробляються. У синоптичній схемі Х. Касареса лексика класифікується за 38 рубриками (Табл. 5):

Табл. 5. Синоптична схема класифікації лексики в “Ідеологічному словнику іспанської мови” за ред. Х. Касареса:

Релігія. Фізика і Хімія. Географія, Метеорологія. Астрономія. Геологія, Мінералогія. Ботаніка. Зоологія. Анатомія. Фізіологія. Медицина. Харчування. Одяг. Житло. Чутливість, Відчуття. Почуття. Існування. Зміни. Відношення. Порядок. Випадковість. Простір. Форма. Рух.	Розташування. Час. Кількість. Розуміння. Оцінка. Судження. Воля. Поведінка. Дія. Мова. Мистецтво. Держава. Народ. Звичаї. Право і правосуддя. Власність. Військова справа. Торгівля. Банківська діяльність. Біржа. Сільське господарство. Зоотехніка. Транспорт.
--	--

Ці категорії, у свою чергу, поділяються на 2 000 підкласів або груп, у кожній з яких містяться дотичні до неї лексеми. 38 категорій, показаних у таблиці, об'єднуються у декілька масштабніших ідей: Бог, Людина, Природа та інші. Тобто, класифікація Х. Касареса є теоцентричною. Натомість, П.М. Роже класифікує лексику, виходячи з абстрактних понять, тобто, радше з ідеалістичних позицій.

В аналогічний спосіб, а саме, за допомогою використання ієрархічних класифікаційних схем, побудовано й ідеографічний словник російської мови, упорядкований О.С. Барановим [Баранов 1995]. Український ономасіологічний словник, роботу над яким ще не завершено, розробляється Н.П. Дарчук, завідувачем лабораторії комп'ютерної лінгвістики Київського національного університету імені Тараса Шевченка.

Створюючи класифікаційну схему лексики для бази даних "WordNet" [Wordnet], дослідники з Принстонського університету (Велика Британія) виокремлюють так звані семантичні множники: 'дія, акт, діяльність', 'об'єкт природи', 'тварина, фауна', 'явище природи', 'ознака, властивість', 'особа, людська істота', 'тіло', 'рослина, флора', 'знання', 'володіння', 'подія, випадок', 'процес', 'почуття, емоція', 'кількість', 'їжа', 'відношення', 'група, сукупність', 'форма', 'місце', 'стосунки', 'мотив', 'стан, статус', 'механізм', 'речовина', 'комунікація'. [Troiano 2011]. Ці множники за своїм характером і властивостями співвідносні з синоптичною схемою.

Класифікація лексики за певними категоріями відіграє неабияку роль не лише для укладання ономасіологічних словників, а й під час розмітки текстів (для визначення теми тексту, для правильного встановлення внутрішньотекстових дейктичних зв'язків тощо). Задля здійснення семантичної розмітки корпусу Н.П.Дарчук пропонує такий підхід: "Якщо будувати за текстом частотний словник (ЧС), то найчастотніші лексико-семантичні варіанти, які відображають тематичну домінанту тексту, будуть знаходитися частково у верхній його частині [Баевский 2001, с. 193]. "Однак є значущі, проте низькочастотні лексико-семантичні варіанти, ігнорувати які не можна, оскільки важливими є смислові зв'язки між різними лексико-семантичними варіантами, в тому числі з високочастотними. Більш продуктивним шляхом укладання тематичної моделі тексту є підхід, який базується на виявленні в тексті тематично значимих лексико-семантичних варіантів та групування їх в лексико-тематичні групи за тезаурусним принципом" [Дарчук 2017, с. 24].

На сайті <https://www.english-corpora.org/>, зокрема, в підкорпусі Британського парламенту ("British Parliament Hansard Corpus"), що охоплює тексти від 1803 по 2005 рр., усю лексику розмічено за 38 категоріями: 'світ', 'життя', 'здоров'я і захворювання', 'люди', 'тварини', 'рослини', 'їжа і напої', 'текстильні вироби й одяг', 'фізичні відчуття', 'матерія', 'існування і причинність', 'простір', 'час', 'рух', 'дія', 'властивості', 'надприродне', 'свідомість', 'увага і судження', 'добро і зло', 'емоції', 'бажання', 'наявність і володіння', 'мова', 'суспільство', 'побут і житло', 'озброєний конфлікт', 'влада', 'закон', 'мораль', 'виховання', 'віра', 'комунікація', 'подорож', 'праця і фах', 'торгівля і фінанси', 'дозвілля', які, в свою чергу, розподіляються на декілька рівнів субкатегорій і дрібних семантичних груп [British Parliament].

Н.В. Сніжко, старший науковий співробітник відділу лексикології та лексикографії Інституту української мови НАН України, у 1995 р. здійснила першу в Україні спробу автоматизованого укладання ідеографічного словника української мови, скориставшись результатами комп'ютерного опрацювання 11-томного тлумачного словника української мови. "На основі лексикографічних параметрів, а саме словникових ідентифікаторів, що відтворюють видові й родові семи, а отже, відображають ієрархічні відношення слів та їхніх значень, вдалося з'ясувати, що весь семантичний простір "ПРЕДМЕТНІСТЬ" в українській мові другої половини ХХ століття зводиться до 101 ідеографічного класу: ЛЮДИНА, ТВАРИНА, РОСЛИНА, ЧАС, ПРОСТІР, ДІЯ, ЗНАРЯДДЯ, РЕЗУЛЬТАТ, ЯКІСТЬ, КІЛЬКІСТЬ" та інш." [Сніжко 2016, с. 94]. Отже, 101 іменник є кандидатами для створення відповідних класифікаційних рубрик.

Як уже зазначено, жодна схема чи синоптична карта не охоплює лексичну систему ідеально. Пошук лексеми в ономазіологічному словнику можна образно порівняти з пошуком певного товару в великому торговельному центрі, адже в ньому товари розташовано не за алфавітним порядком (і навіть при всьому бажанні цей задум нездійснений, оскільки один і той самий товар може мати декілька назв). У той самий час самі покупці не завжди знають точну назву товару. Виявляється, що такий товар як, до прикладу, фільтри для кави, може розташовуватися як на полицях для фільтрів, так і в відділах паперових виробів, і в відділі посуду, і в відділі кави. Приналежність одного поняття водночас до різних категорій, згідно з В.П. Берковим [Берков 2004, с. 13], називається принципом

перетинання. Так само, як неможливо розташувати один і той самий товар про всяк випадок у різних секціях магазину, проблематично розмішувати кожне поняття в різні рубрики ономазіологічного словника: з одного боку, це підвищило б шанс знаходження потрібної лексеми, а з іншого – непомірно збільшило б обсяг словника. І хоча у тезаурусі П.М. Роже багатозначні слова часто потрапляють до різних рубрик, В.І. Перебийніс вказує на це як на ваду [Перебийніс 2009, с. 207]. Тим не менш, у цифровому форматі таке багатовсюдне розташування лексем (а точніше – “принцип перетинання”) – цілком здійсненне завдання, більше того, це вже не є недоліком, а, навпаки, може стати перевагою.

Таким чином, ієрархічна класифікація понять, заснована на гіпо-гіперонімічних та меро-голонімічних²⁰ зв'язках, потрібна для забезпечення системного і впорядкованого пошуку необхідної лексеми в ономазіологічному словнику. Основна сьогоденна проблема класифікації полягає в неоднозначності та суб'єктивності класифікації лексики. Провідним підходом виявляється виокремлення найчастіше вживаних лексем узагальнюючого значення у текстах та словниках. Ще однією проблемою упорядкування синоптичної схеми є принцип перетинання, згідно з яким одна й та сама лексема може належати до декількох рубрик водночас. Використання ономазіологічних словників передбачає наявність розвиненої інтуїції, вміння встановлювати різного роду відношення між лексемами, володіння дефініційним словником. Втім, можливості формального комп'ютерного пошуку суттєво полегшують використання ономазіологічного словника на комп'ютерній основі.

Контрольні питання

1. Які існують синоніми і квазісиноніми поняття “слово”? У чому полягає відмінність між ними?
2. У чому полягають відмінності між словоформою і графічним словом?
3. У чому полягають відмінності між слововживанням і словоформою?
4. Чи завжди саме лексичні одиниці є об'єктом упорядкування словників? Проілюструйте відповідь прикладами.

²⁰Меронім — поняття, що виражає частину по відношенню до цілого (голоніма), наприклад: “гілка”-“дерево”.

5. Назвіть синоніми терміна “реєстрове слово” та поясніть переваги й недоліки номінацій означуваного поняття тим чи іншим синонімом.

6. Перерахуйте щонайменше три типи систематизації структурних компонентів у словниках, можливі за допомогою комп'ютерних технологій.

7. Дайте визначення понять: тезаурус, розмітка тексту, лексико-семантичний варіант, глоса, лема.

8. У чому полягають можливості тезауруса, укладеного за допомогою комп'ютерних програм, на відміну від тезауруса у паперовому форматі?

9. Дайте визначення понять: гіпертекст, словникова стаття, ідеографічний словник, колокація, кругове визначення.

10. Чим відрізняється макроструктура від мікроструктури словника?

11. Які альтернативи впорядкування, окрім алфавітного, існують у сучасних електронних словниках?

12. Які компоненти словника можна пов'язати відношенням “ключ” – “значення”?

13. У чому полягає відмінність між поняттям “слово” і “лексема”. Проілюструйте відповідь прикладом.

14. Як розраховуються індекс *TTR*?

15. Яке прикладне використання індексу *TTR*?

16. Які існують можливості використання індексу *TTR* у філологічних дослідженнях?

17. Які параметри можуть міститися у правій частині (глосі) словникової статті в електронних та паперових словниках?

18. Яка роль стилістичних та соціолокальних поміток в електронних та паперових словниках?

19. У чому полягають кількісні і якісні відмінності між паперовими і електронними словниками?

20. Назвіть приклади гіпертекстуальних функцій в електронних словниках, якими ви часто користуєтесь.

21. Якими якісними характеристиками відрізняється сканована версія паперового словника від електронного?

22. У чому полягають проблеми виокремлення лексико-семантичних варіантів лексеми?

23. Чим відрізняються чітко виокремлювані значення лексеми від нечітко виокремлюваних?

24. Якими шляхами можна зменшити кількість кругових визначень у словниках?

25. Чи можливо створити тлумачний словник, повністю позбавлений кругових визначень? Аргументуйте відповідь.

26. Одиниці якого мовного і мовленнєвого рівнів є предметом сучасних електронних словників?

27. Які існують різновиди двомовних перекладних словників?

28. Скільки можливих типів двомовних перекладних словників можна виокремити у заданій мовній парі?

29. У чому полягає специфіка електронних двомовних перекладних словників у порівнянні з паперовими?

30. Назвіть приклади семасіологічних словників.

31. Назвіть щонайменше 7 критеріїв класифікації словників.

32. У чому відмінності між ономасіологічним і семасіологічним словником?

33. Якого характеру труднощі перекладу можуть вирішувати семасіологічні тлумачні словники?

34. Які труднощі перекладу можна розв'язувати за допомогою ономасіологічних словників?

35. Як організована макроструктура ономасіологічного словника?

36. У чому причини недоліків двомовних перекладних словників?

35. Які принципово нові можливості в електронних тлумачних та перекладних словників, у порівнянні з паперовими?

37. За яким принципом розташовуються лексеми у тезаурусі?

38. В яких значеннях вживається термін "атлас лексики"?

39. Як називається словник, в якому можна проводити пошук лексики за дефініцією?

40. Як називається схема лексики, побудована на ієрархічній класифікації, що зазвичай подається у словниках?

41. За яким принципом можна класифікувати лексику у словниках?

42. На скільки ієрархічних рівнів можливо розподілити класифікацію лексики в ономасіологічних словниках?

43. Які принципово нові можливості є в комп'ютерних синоптичних схемах, у порівнянні з паперовими?

44. Чому синоптична схема Х. Касареса називається теоцентричною? Які можливі альтернативні засоби структурування синоптичних схем?

45. Чим відрізняються особливості класифікації термінологічної лексики і лексики загального вжитку?

46. Назвіть приклади словників, автоматично згенерованих комп'ютером. Яке можливе їхнє застосування?

47. Які функції словників, недоречні чи неможливі в паперовому форматі, цілком прийнятні в електронних словниках?

Рекомендовані джерела

1. Берков В.П. Двухязычная лексикография. – М.: Транзиткнига, 2004. – 237 с.
2. Виноградов В.В. Лексикология и лексикография. – М.: Наука, 1977. – 318 с.
3. Гак В.Г. О разных типах двухязычных словарей // Тетради переводчика. – М., 1964 – Вып. 2. – С. 71-78.
4. Городенська К. Граматична лексикографія: стан і перспективи // Українська мова : науково-теоретичний журнал / Нац. акад. наук України, Ін-т укр. мови ; редкол.: П. Гриценко, К. Городенська, О. Яцевська [та ін.]. – Київ, 2018. – № 1 (65). – С. 30-37.
5. Граматичний словник української літературної мови : словозміна / В.І. Критська [та ін.]; відп. ред. Н.Ф. Клименко ; Нац. акад. наук України, Ін-т укр. мови, Ін-т мовознав. ім. О.О. Потебні. – К.: Вид. дім Дмитра Бураго, 2011. – 757 с.
6. Комп'ютерна лексикографія, її предмет та завдання // Карпіловська Е.А. Вступ до прикладної лінгвістики: комп'ютерна лінгвістика. – Донецьк: Юго-Восток, 2006. – С. 54-74.
7. Караулов Ю.Н. Лингвистическое конструирование и тезаурус литературного языка. – М.: Наука, 1981. – 367 с.
8. Клименко Н.Ф. Нові підходи до укладання комп'ютерних словників // Мовознавство. – К., 1996. – № 4-5. – С. 11-15.
9. Комп'ютерна лексикографія // Селіванова О.О. Сучасна лінгвістика: напрями та проблеми: підручник. – Полтава: Довкілля-К., 2008. – С. 660-664.
10. Перванов Я.А. Материалы к русско-болгарскому сопоставительному словарю. [Електронний документ]. Режим доступу: <https://arxiv.org/pdf/1104.2034.pdf>
11. Перебийніс В.С. Частотні словники та їх використання / В.С. Перебийніс, М.П. Муравицька, Н.П. Дарчук. – К.: Наукова думка, 1985. – 204 с.
12. Перебийніс В.І., Сорокін В.М. Традиційна та комп'ютерна лексикографія. Навч. посібник. – К.: Вид. центр КНЛУ, 2009. – 218 с.
13. Селегей В. Электронные словари и компьютерная лексикография. [Електронний документ]. Режим доступу: www.lingvoda.ru

14. Сірук О.Б. Тезаурус дієслів української мови: лінгвістичні проблеми та методика конструювання: Автореф. дис. ...канд. філол. наук: Спец. 10.02.01 / О.Б. Сірук. – К.: КНУ імені Тараса Шевченка, 2006. – 20с.
15. Сніжко Н.В. Зведений словник української мови: джерела і засади укладання / Н. Сніжко // Культура слова. – 2015. – Вип. 83. – С. 90-99. – Режим доступу: http://nbuv.gov.ua/UJRN/Kuls_2015_83_18
16. Сніжко Н. Структурна, функціональна та хронологічна параметризація лексики у зведеному словнику, тезаурусі й електронній картотеці // Українська мова : науково-теоретичний журнал – К., 2016. – № 4 (60). – С. 87-102.
17. Фокін С.Б. Іспанські відповідники прислівника “довго”: перекладознавчий та лексикографічний аспекти // Мовні і концептуальні картини світу. – К., – 2010 – Вип. 31. Том. – С. 172-177.
18. Фокін С.Б. Вибір дескрипторів для ономазіологічного словника: питання методології // П'ята Всеукраїнська наукова конференція романістів. Структурно-семантичні і когнітивно-дискурсивні парадигми сучасного романського мовознавства. – Одеса: Одеський національний університет, 2015. – С. 71-72.
19. Фокін С.Б. Двомовні перекладні словники: причини недосконалості та перспективи оптимізації // Мовні і концептуальні картини світу. – К.: ВПЦ “Київський університет”, 2017. – Вип. 59. – С.214-220.
20. Фокін С.Б. Глибина кругових визначень у словнику як віддзеркалення властивості лексичної системи // Фаховий та художній переклад: теорія, методологія, практика: збірник наукових праць. – К.: Аграр Медіа Груп, 2018. – С. 370-376.
21. The Longman American Defining Dictionary. – [Електронний ресурс]. Режим доступу: http://www.longmandictionariesusa.com/res/shared/vocab_definitions.pdf
22. Casares J. Introducción a la lexicografía moderna / J. Casares. – Madrid: Consejo Superior de Investigaciones Científicas, 1950. – 354 p.
23. Cowie A.P. English dictionary for foreign learners: a history / A.P. Cowie. Oxford: Clarendon Press, 1999. – 235 p.
24. 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>
25. Hai Xu. A Critique of the Controlled Defining Vocabulary in Longman Dictionary of Contemporary English, 2012. – Lexikos 22(1), p. 367-381 – [Електронний ресурс]. Режим доступу: <http://lexikos.journals.ac.za/pub/article/view/1013/530>

26. Luque Durán J.D., Pamies Bertrán A. (eds) *Léxico y fraseología*. – Granada: Mateo Ediciones, 1998. – 225 p.
27. Martí Antonín M., Fernández Monraveta A. & Vázquez García G. *Lexicografía computacional y semántica* – Barcelona: Ediciones de la Universidad de Barcelona, 2003. – 163 p.
28. Smouchtchynska I. *Lexicologie française*. – K.: Видавничий дім Дмитра Буparo, 2015. – 600 с.
29. Stock P.F. Polysemy // *Practical Lexicography. A Reader*. – Oxford University Press. 2008. – P. 153-160.

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

ПЕРЕГЛЯД ЦИХ СТОРІНОК МОЖЛИВИЙ
У ПАПЕРОВОМУ ВАРІАНТІ ПОСІБНИКУ

РОЗДІЛ V ДОПОМІЖНІ ПРОГРАМИ В ДІЯЛЬНОСТІ ПЕРЕКЛАДАЧА І ПЕРЕКЛАДОЗНАВЦЯ

5.1. Функції текстових редакторів, корисні у практиці перекладу

Якщо в вісімдесяті-дев'яності роки XX століття набір тексту на комп'ютері розглядався як альтернатива друкарської машинки, нинішні текстові редактори оснащено широкими можливостями автоматизації часто повторюваних дій, одночасного виконання різних функцій, паралельної роботи з декількома документами та програмами, новітніми знаряддями захисту й відновлення інформації, інтеграції до текстових документів функцій інших програм тощо. Зупинимось докладніше на тих засобах, що безпосередньо корисні в роботі перекладача.

Налаштування автоматичного створення резервних копій. У разі випадкового видалення або псування робочого файлу користувач може відновити дані передостанньої збереженої версії документа. Налаштування директорії, в якій зберігаються резервні копії, знаходяться у таких вкладках:

для LibreOffice: Tools – Options – LibreOffice – Paths

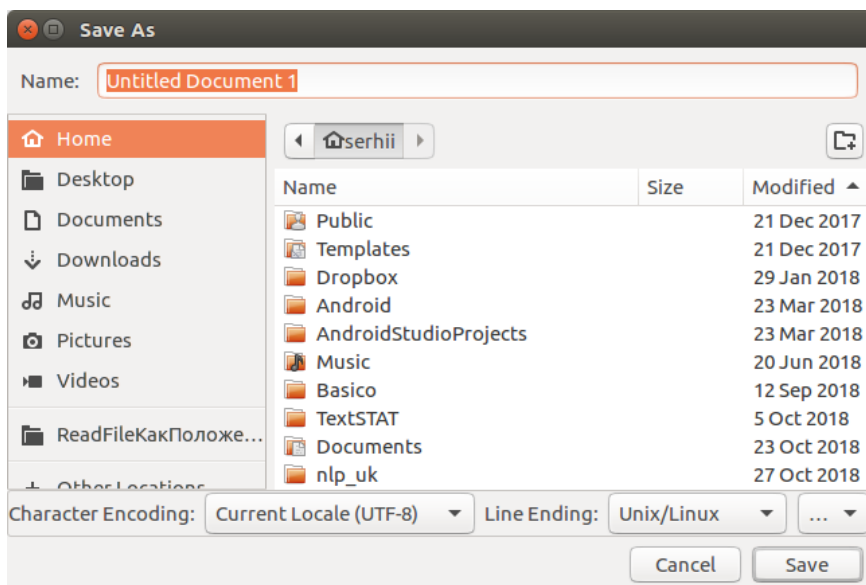
для Microsoft Word: Options – Advanced – Create Backup Copy

Іноколи опція збереження резервних копій у текстовому редакторі є обраною за замовчуванням, тому навіть якщо користувач попередньо не здійснив налаштування резервного збереження копій, у разі втрати даних рекомендуємо перевірити, чи обрана певна директорія, в якій автоматично зберігається попередня версія файлу, оскільки існують шанси знайти в ній загублені дані.

Збереження текстового файлу у різних форматах і системах кодування символів. Хоча зазначена можливість знайома більшості сучасним користувачам, наголосимо на тому, що текстові редактори дозволяють зберігати файли у форматах

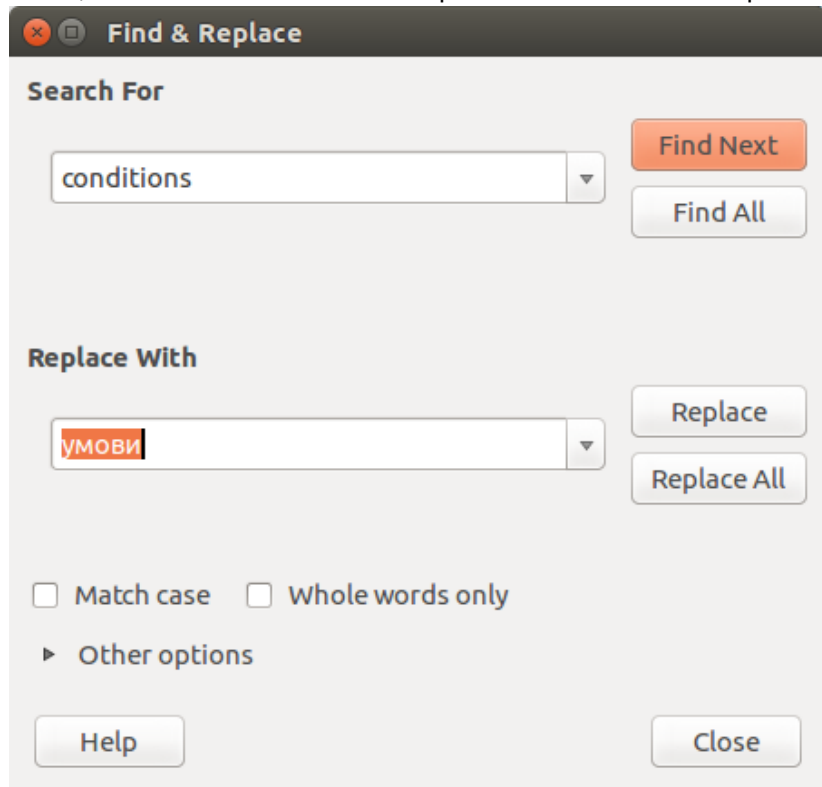
doc, docx, rtf, odt, txt, pdf, html, xml та інш. Обрати формат можна у діалоговому вікні, що викликається функцією “Зберегти як...”.

У роботі з сучасними текстовими редакторами, такими як *Libre Office* та *Microsoft Word*, потреба зміни й налаштування кодування виникає рідко. Однак у разі необхідності збереження текстів у форматі простого тексту *txt*, а також у форматі субтитрів *srt* інколи доводиться змінювати кодування тексту для того, щоб він коректно прочитувався в інших програмах. Практика показує, що латинські літери з діакритичними позначками правильно прочитуються у форматі *UTF-8*, в якому можуть коректно відображатися і кириличні символи, якщо ж ні – для кириличних символів рекомендуємо скористатися кодуванням *Unicode Big Endian* або *ANSI*. У форматі *UTF-16* на символи виділяється вдвічі більше пам'яті, отже, він є більш універсальним, однак для його підтримки в відповідному середовищі необхідно попередньо обрати це кодування в налаштуваннях операційної системи. У текстових редакторах кодування може бути обране під час збереження файлу у форматі *txt*:



Функція пошуку і глобальної заміни у текстових редакторах. Переважна більшість перекладачів широко використовує практично універсальну операцію, єдину для різних програм і навіть поширених операційних систем для комп'ютерів, а саме, наскрізний пошук у файлі, текстовому документі, що найчастіше викликається сполученням клавіш *Ctrl+F*. Однак меншість перекладачів знає про нюанси додаткових параметрів пошуку, що можуть прислужитися у їхній роботі.

По-перше, за допомогою використання функції глобальної заміни, можна автоматично перекласти часто повторювані



терміни, словосполучення. Операція глобальної автозаміни викликається сполученням клавіш *Ctrl+H*.

У той самий час актуальним буває використання і маски пошуку. Маска пошуку відповідає не конкретному заданому сполученню символів, а в узагальненому вигляді відображує цілу множину виразів. Якщо у пошукачі *Google* вираз умов*

відповідатиме графічним словам: “умов”, “умова”, “умови”, “умовний”, “умовити”, “умовляти” і т.п., у текстових редакторах такі знаки як зірочка “*” і знак питання “?” не виступають байдужими символами. Для пошуку усіх графічних слів, що починаються зі сполучення символів “умов” необхідно скористатися регулярним виразом. У редакторі *Libre Office* відповідний запит можна реалізувати за допомогою такого виразу:

Find & Replace

Search For

умов[а-яА-Яіігґєєїї]*

Find Next

Find All

Replace With

Replace

Replace All

☐ Match case ☐ Whole words only

▼ Other options

☐ Current selection only ☐ Backwards

☒ Regular expressions ☐ Search for Paragraph Styles

☐ Similarity search **Similarities...** ☐ Comments

Attributes... **Format...** **No Format**

Help **Close**

Порівняння двох версій текстового документа.
Надзвичайно корисною для перекладача функцією сучасних

текстових редакторів є порівняння двох версій документа. Шляхом пошуку спільних та відмінних рис у двох текстах програма відображає ті зміни, які відбулися у наступній версії, в порівнянні з попередньою. У такий спосіб помітити зміни набагато легше, ніж шляхом суцільного перечитування обох текстів.

Libre Office: Edit – Track Changes – Compare Document
Microsoft Word: Review – Compare

Система контролю версій. Під час виконання масштабної роботи її доцільно розподіляти між декількома перекладачами. Розподіл не обов'язково відбувається послідовно, адже один перекладач може вичитувати, виправляти, вдосконалювати фрагменти, перекладені колегою. Якщо декілька користувачів вносять зміни до певного документа на власних комп'ютерах одночасно, під час збереження може виникнути конфлікт версій. Уникнути цього дозволяє система *GoogleDocs*: вона дозволяє відкривати й редагувати документ одночасно на декількох пристроях; внесені зміни практично миттєво відображаються на моніторах усіх активних користувачів.

Автоматичний розрахунок обсягу тексту. Часто необхідно розраховувати обсяг тексту, щоб оцінити масштаб роботи, кількість необхідного на неї часу, а також розрахувати її вартість. В Україні поширеною практикою є розрахунок сторінки у кількості знаків (1860 знаків з урахуванням пробілів дорівнює сторінці), у країнах Західної Європи прийнято вимірювати довжину тексту у кількості слововживань, відповідно, і тарифи послуг перекладача пропонуються по відношенню до кількості слововживань, а не знаків. Операцію розрахунку обсягу тексту або ж виділеного фрагмента тексту можна здійснити таким шляхом:

Tools – “Word count”

Окрім знання функціональної кнопки, перекладачеві необхідно зважати й на те, що поняття “слова” або, в цьому контексті “графічного слова”, теж не таке однозначне, отже, статистичний підрахунок може вестися по-різному. А від кількості слів залежить гонорар перекладача. Оскільки не тільки пробіл, а й інші символи можуть вважатися роздільниками слів, наприклад, у програмі автоматизованого перекладу на основі електронної перекладацької пам'яті *Déjà Vu* також роздільником вважається

апостроф, у той час як у програмі *Microsoft Word* серед роздільників фігурує також дефіс. Поза тим, у текстовому редакторі не вважаються словами сполучення цифр, оскільки вони не підлягають перекладу²¹.

Переклад субтитрів за допомогою текстових редакторів. Сучасні субтитри характеризуються розмаїттям форматів. Розглянемо два найпоширеніших з них: *ass*, *srt*. Зазвичай субтитри до фільмів містяться в окремих файлах з розширеннями, що відповідають їхньому формату *ass* ("*Alpha Sub Station*"), формату *srt* ("*Sub Rip Text*"). Файли субтитрів можна відкривати, редагувати і перекладати за допомогою спеціалізованих програм, кожна з яких характеризується набором власних функцій і своєрідним інтерфейсом. Розглянемо наразі можливість перекладу з використанням текстових редакторів. Ось приклад формату *ass*:

Dialogue: 0,1:07:18.74,1:07:19.89,Default,,0,0,0,,Hi! How are you?
Dialogue: 0,1:07:20.34,1:07:21.41,Default,,0,0,0,,I'm fine, thanks!

Нижче наводимо також формат субтитрів *srt*:

1107
01:24:28,665 --> 01:24:30,064
Hi! How are you?

1108
01:24:31,145 --> 01:24:32,260
I'm fine, thanks!

З технічного оформлення видно, що кожен субтитр складається з двох функціонально різних частин: часової розмітки та тексту субтитру. Перекладачі з досвідом можуть дещо змінювати й часову розмітку, однак варто робити це за допомогою спеціальних програм, які попереджають про перевантаження субтитрів

²¹Насправді числові дані доволі часто зазнають змін у перекладі з романських та германських мов на українську та навпаки: приміром, в українській мові, так само як і у французькій, російській, між розрядами цілих чисел прийнято ставити пробіл, в іспаномовних текстах — крапку, а в англomовних — кому; десяткові частки в англomовних текстах відокремлюються від цілої частини крапкою.

знаками, надто дрібні інтервали, накладання одного субтитру на інший тощо. Під час перекладу змінюється переважно текстова частина. За умов збереження формату часової розмітки можна також редагувати час початку і закінчення субтитру. Наприклад, останній субтитр можна подовжити на півсекунди у такий спосіб:

01:24:31,145 --> 01:24:32,260

У разі зміни таймеру слід враховувати, щоб кінець субтитру не накладався на наступний субтитр, щоб витримувалися загальноприйняті вимоги до субтитрів: тривалість від 1 до 6 секунд, наявність мінімального часового інтервалу між субтитрами.

Під час перекладу слід дотримуватися таких принципів: субтитр не повинен перевищувати двох рядків; верхній рядок має бути коротшим, ніж нижчий; у рядку має міститися не більше 35-40 символів. Так може виглядати перекладений субтитр:

1107

01:24:28,665 --> 01:24:30,064

Вітаю! Як справи?

1108

01:24:31,145 --> 01:24:32,260

Дякую, усе гаразд.

Найпоширенішою системою кодування для кирилічного алфавіту є *UTF-8*, однак у деяких програмах для коректного прочитання українських літер потрібно користуватися форматом *ANSI*. Кодування обирається на етапі відкривання або закривання файлу і діалоговому вікні, як, наприклад, у редакторі *TextEditor*:

Контроль якості перекладу за допомогою текстових редакторів. Сучасні текстові редактори оснащено знаряддями перевірки орфографії. Вимірювання обсягу тексту теж може прислужитися перекладачеві: суттєві відхилення обсягу по відношенню до тексту оригіналу можуть свідчити про втрату певного фрагмента або інші похибки (зазвичай текст перекладу дещо довший за текст оригіналу). Окремі програми автоматичного перекладу безпосередньо вбудовуються в текстовий редактор. До таких, наприклад, належить *WordFast*, програма машинного перекладу на основі електронної перекладацької пам'яті.

Таким чином, сучасні текстові редактори, завдяки багатій функціональності і автоматизації численних дій (таких як збереження резервних копій, встановлення паролю, приведення у відповідність різних версій документа, розрахунок обсягу тексту), суттєво полегшують працю перекладача. Функції пошуку, глобальної автозаміни дозволяють знаходити не лише задане сполучення літер, а й цілу сукупність виразів водночас, виділяти їх обраним форматом. Це наближує текстові редактори до знарядь штучного інтелекту. Значно детальніше можливості пошуку представлено у програмах опрацювання корпусу (конкордансерах).

5.2. Принципи роботи програм перевірки якості перекладу

Суттєво полегшують працю перекладача програми, що перевіряють орфографію. Це найвідоміша з комп'ютерних функцій, знайома всім нинішнім перекладачам, що виконує формальну перевірку якості перекладу. Цією функцією оснащено практично всі сучасні текстові редактори. Однак сучасні технології здатні зарадити перекладачеві більше. Зокрема, спеціальні програми можуть, за певними формальними критеріями, вказувати на деякі невідповідності оригіналу і перекладу. Часто ці функції закладено в самі системи машинного перекладу (наприклад, такі як Trados, Déjà Vu), однак перевірку якості можна виконати й спеціальним окремим програмним забезпеченням.

З-поміж найтипівіших помилок, що їх здатні розпізнати вказані програми, можна назвати такі: неперекладений фрагмент перекладу, повтор фрагмента оригіналу в перекладі, невідповідність обсягу цільового фрагмента та фрагмента оригіналу (відсоток розбіжності регулюється), наявність неприпустимих символів у фрагменті перекладу, неоднаковість перекладу ідентичних фрагментів оригіналу, відмінність пунктуаційних знаків у кінці речення в оригіналі й перекладі, пробіли між знаками пунктуації, подвійні пробіли, подвійні крапки, невідповідність уживання великої літери на початку речення, незбіг цифр у числових даних, спеціальних тегах, помилки в відтворенні неперекладних назв та позначок (їхній список задається заздалегідь) [Gerasimov].

Згадаємо деякі програми, зауважуючи, що створити вичерпний список бодай основних знарядь – практично нездійсненне

завдання, оскільки деякі з нових продуктів обов'язково з'являться, поки ця праця перебуватиме в друці.

QAD (QA Distiller) – програма здійснює перевірку не лише перекладів, а й комп'ютерної перекладацької пам'яті.

Xbench – окрім зазначених параметрів, здійснює перевірку відповідності послідовностей літер верхнього й нижнього регістрів, перевіряє правильність перекладених термінів відповідно до завантаженої бази даних поширених форматів (*Déjà Vu*, *WordFast*, *Trados*, *TBX*, *TMX* та інш.).

Verifika, окрім указанного вище, також допомагає перевіряти відповідність різного роду форматів як окремих графічних слів, так і числових даних; дозволяє користувачеві створювати й додавати власні параметри звірки оригіналу й перекладу), так само, як і програма *ErrorSpy*.

Корпусні дослідження в галузі перекладу здатні збагатити можливості перевірки. Зокрема, цілком зримим у найближчому майбутньому видається завдання з перевірки відповідності сполучуваності фрагментів у текстах перекладу і у референтному корпусі. Так, наприклад, встановлено факт, що, для забезпечення метафоричності словосполучення перекладачі вживають у тексті ті сполучення, що рідше зустрічаються у референтному корпусі цільової мови [Lewandowska 2012: 20-25]; цікаві висновки щодо стилю перекладу, його лексичного багатства може надати розрахування коефіцієнту *TTR* (докладніше – у підрозділі 1.1.); можлива інтеграція тексту перекладу зі словниками “хибних друзів” перекладача, що за своєю суттю і призначенням є фактично словниками типових помилок перекладу. Ці та інші спостереження щодо особливостей перекладних текстів у зіставленні з оригінальним мовленням відкривають значні перспективи збагачення як програм машинного перекладу, так і знарядь перевірки якості.

5.3. Принципи роботи програм обробки природної мови у системах штучного інтелекту

Найбільш животрепетне питання розробників штучного інтелекту – можливість навчити комп'ютер або машину розуміти смисл повідомлень. І хоча тут не йдеться про комунікацію власне в людському розумінні, дедалі частіше взаємодія між людиною і технікою залежить від успішності спілкування між ними. Під час такого спілкування мовлення має бути зрозумілим для техніки.

Тому або мовлення має адаптуватися людиною, або необхідно розробляти знаряддя для розпізнавання мовлення і його трансформації у зрозумілий для техніки код. Для цього, наприклад, пошукові системи, так само як і багато інших технічних засобів, в яких задіяно опрацювання людського мовлення, мають розрізняти морфологічні характеристики однієї лексеми. Окрім власне граматичного аналізу, однією з нагальних проблем є зняття двозначності (або неоднозначності) повідомлення, окремих його складових. Ці та інші проблеми спілкування людини і техніки розв'язують у різних галузях наук, і не останню роль у цьому процесі відіграє комп'ютерна лексикографія. Розробка спеціальної технічної мови, або лексикону, також дуже необхідна у системах розпізнавання і синтезу мовлення (скажімо, в автоматичному синхронному перекладі), довідкових системах з автоматичними відповідями на питання, у системах автоматизованої вибірки даних з текстів (англ. – “text mining”) тощо. Український вчений Е.Ф. Скороходько ще у 1983 р. запропонував у своїй праці “Семантичні мережі і автоматична обробка тексту” спеціальну методологію для автоматичного реферування текстів [Скороходько 1983, с. 151-186], автоматичного індексування²² тексту, представлення семантичної мережі тексту. Узагальнення підходів до реферування тексту, а також принципи дії й розробки інших програм автоматичного опрацювання текстів описано у підручнику Н.П. Дарчук “Комп'ютерна лінгвістика” [Дарчук 2008, с. 288-293]. Н.П. Дарчук належать провідні розробки в галузі опрацювання українськомовних текстів, зокрема, аналізатор “АГАТ”, а також програми перекладу “РУТА”, “ПЛАЙ”.

Суттєву роль у процесі автоматичної обробки тексту відіграє розпізнавання граматичних категорій, і не лише морфологічних, а й синтаксичних характеристик. Утім, комп'ютерні технології вносять свої корективи, і категорії комп'ютерної граматики можуть відрізнятися від тих, що описано у класичних працях. “Комп'ютерна граMATика – системний, строго впорядкований, формалізований, лінгвостатистичний інтегральний опис знакових одиниць певної мови у вигляді структурних моделей із необхідною і достатньою аналітикою для виконання завдань штучного інтелекту” [Дарчук 2015, с. 21]. Приміром, у корпусі текстів української мови на порталі *mova.info* слова розмічено за такими морфологічними категоріями: ад'єктив, дієслово, числівник,

²²Індексування – різновид розмітки, спеціальний метамовний опис, який найліпше відповідає потенційним запитам пошуку.

присудкове слово, займенниковий іменник, займенниковий прикметник, присвійний займенник, сполучник сурядний, сполучник підрядний, прийменник, прислівник, частка, вигук, скорочення, скорочення з крапкою, аббревіатура. Якщо слово “твій” розмічено як “займенниковий прикметник”, його можна з високою долею ймовірності перекласти на англійську мову займенником “*my*”, а якщо він маркований як займенниковий іменник – то, найвірогідніше, правильним перекладом у контексті речення виявиться форма абсолютного присвійного займенника “*mine*”. У Браунівському корпусі використовується 87 тегів для частиномовної розмітки. Велика кількість маркувань пояснюється тим, що у назві тегу поєднується частиномовна характеристика разом з морфологічною категорією. Зазначені факти свідчать про те, що з метою автоматичної обробки тексту справді необхідна специфічна граматики. Меншою мірою у корпусній лінгвістиці розроблено синтаксичні параметри.

Отже, розпізнавання частиномовної характеристики допомагає і в системах машинного перекладу, і під час смислового аналізу тексту.

В усіх перерахованих вище процесах вкрай важливо розробити зрозумілий для машини лексикон для автоматичної обробки тексту. За Е.Ф. Скороходьком, сучасні тлумачні словники, з огляду на численні логічні помилки та інші недосконалості, не містять достатньо чітких даних про семантичні зв'язки між лексемами і не здатні давати уявлення про “ланцюги семантичних відношень” [Скороходько 1983, с. 151-186]. Це цілком природно: комп'ютери потребують значної кількості даних, які не потрібні людині, адже виводяться шляхом елементарних логічних міркувань або інтуїтивно завдяки емпіричному досвіду, у той час як для комп'ютера мають бути експліцитно вказані численні формальні деталі: статистичні відомості щодо сполучуваності, тип, розряд, семантична група лексеми тощо. Чітко впорядковане семантичне маркування теж є ключовим в автоматичній обробці тексту: іменник “*tabaco/tabacco/tabac*” в іспанській, італійській і французькій мовах, маркований як рослина, позначає “тютюн”, а якщо в контексті він має процесуальну сему (“*El tabaco perjudica seriamente la salud*” – ісп., “*Il tabacco uccide*” – італ., “*Le tabac tue*” – фр.), його значенням є “тютюнопаління”. Крім того, комп'ютерний лексикон має бути зручним для представлення формальних даних та зняття двозначності.

Отже, на нинішньому етапі розробки програм машинного інтелекту завдання комп'ютерної лінгвістики полягає у створенні інструментів зняття різного роду двозначності шляхом розробки й використання алгоритмів автоматичного розпізнавання лексико-граматичних класів слів (частиномовної розмітки), визначення семантичних груп і класів, виявлення синтаксичних ролей і функцій тощо. Розподіл мовних одиниць на класи, групи, необхідний для забезпечення потреб штучного інтелекту, відрізняється від сучасних класифікацій граматичних і лексичних категорій, розроблених для навчання людини. Розглянемо зазначені проблеми представлення даних про мову для комп'ютера у двох наступних підрозділах.

5.3.1. Роль лематизаторів та програм частиномовної розмітки для зняття граматичної двозначності

Одним з найпотужніших формальних методів аналізу тексту, який, до того ж, відносно легко застосовувати на практиці, є використання програм, які співвідносять словоформу з лемою (або початковою словниковою формою). Вони називаються лематизаторами. Одним з прикладів такої програми, розробленої на матеріалі іспанської мови, є “Flexionador y lematizador de palabras del español” [Flexionador y lematizador]. Якщо за допомогою лематизатора проаналізувати іменник “ganando”, він виведе такі результати:

Табл. 9. Результат аналізу дієслівної форми за допомогою лематизатора:

 Початкова форма: ganar	Flexionar	Relaciones
 Категорія: перехідне дієслово; неперехідне прономінальне		
 Флексія: герундіальна		
 Семантична класифікація:		
Матеріальні значення		
 Акціональне		
  Спеціальні лексико-семантичні варіанти в галузі науки, мистецтва, родів діяльності		
Нематеріальні значення		
 Вольові дії		

- ▮▮ Набувати
- ▮▮ Стан або статус
- ▮▮ Досягнення [Flexionador y lematizador]²³

Як бачимо, окрім виявлення граматичної категорійної приналежності особової форми або неособової форми дієслова, віднаходження початкової форми, лематизатори здатні визначати й деякі семантичні характеристики лексеми, зокрема, у показаному прикладі – це приналежність до акціонального дієслова, що виражає процеси набуття та досягнення, і це, безсумнівно, відіграє колосальну роль у знятті неоднозначності. Крім того, зазвичай лематизатори містять і функції флексера (флексер – програма, що генерує список усіх словоформ за всією граматичною парадигмою).

Інтернет не рясніє лематизаторами для англійської мови, очевидно, з огляду на бідність флексій у цій мові. Відповідно, лематизатор для англійської мови елементарно може створити програміст-початківець, програма може працювати за принципом відкидання закінчень *-s*, *-ed*, *-ing*, а також ідентифікації неправильних дієслівних форм та іменникових форм множини.

У програмах обробки тексту використовуються також дещо складніший процес, який називається стеммінг (від англ. “stem” – “основа”). У результаті стеммінгу словоформи в тексті зводяться до основ, отже, окрім відкидання флексії, від слова відкидаються також деякі словотвірні морфеми, насамперед, суфікси. У спрощених алгоритмах опрацювання текстів замість відкидання закінчень за замовчуванням вважається, що дві або три останні літери найчастіше змінюються при словозміні і збігаються з флексією, і тому їх відкидають як несуттєві. Їх прийнято називати псевдофлексією.

Для функціонування лематизатора, як зазначає Н.П. Дарчук, необхідна база даних з так званим “граматичним словником” зі словниковими статтями, що містили б морфологічні варіанти лексеми. Так, графічне слово “твори” за даними розмітки марковане як іменник, воно є множиною називного або знахідного відмінка іменника “твір”; якщо воно помічене як дієслово, воно є наказовим способом другої особи однини дієслова “творити”.

Оскільки лематизатори зводять словоформи до леми, їхнє застосування дуже широке: їх можна використовувати у різноманітних програмах, що опрацьовують людське мовлення: від

²³Переклад наш — С.Ф.

автоматичного реферування до систем машинного перекладу, адже словники, на яких ґрунтуються такі програми, побудовані переважно на лемах, а не на особових чи відмінкових словоформах.

5.3.2. Методи зняття лексичної двозначності шляхом автоматичного аналізу

Однією з центральних проблем машинного перекладу та формалізації людської мови є багатозначність як повідомлення в цілому і окремих його складових. Якщо кількість граматичних структур у мові доволі обмежена, і їхні форми і значення достатньо вичерпно описані, лексичний рівень характеризується надзвичайною масштабністю, динамічністю і практичною неосязністю. Отже, вичерпний опис виявляється нездійсненним завданням, і необхідні методологічні знаряддя, які дозволять знімати неоднозначність шляхом формального аналізу. Як буде показано далі, комп'ютерні лексикографічні ресурси відіграють у цих процесах суттєву роль. Зняття двозначності доцільно здійснювати у таких комп'ютерних ресурсах: в електронних словниках, у текстах та текстових корпусах. Саме з цих міркувань варто розробляти спеціальну метамову для семантичної розмітки текстів, а також лексем у словниках.

Одним з доволі простих прийомів пропонується у положеннях теорії детермінант, розроблених Ю. Марчуком [Марчук 2006, с. 79]. Згідно з нею, графічне слово може бути потенційно двозначним, однак значно рідше двозначністю характеризуються сполучення слів: за умов наявності певних слів-детермінант у найближчому контексті двозначність знімається, завдяки чому люди здатні повноцінно порозумітися. Так, наприклад, український іменник "потяг" може означати як "поїзд", так і "захоплення чимось, схильність до чогось". Звичайно, такі мінімальні контексти як "потяг до літератури", "потяг до Відня" безпомилково визначають лексико-семантичний варіант лексеми завдяки детермінантам "література", "Відень"; натомість прикметник "величезний" вже не дозволяє зняти двозначність у сполученні з іменником "потяг". Конструктивний висновок для комп'ютерної лексикографії полягає в тому, що словники, необхідні для штучного інтелекту, мають містити, окрім самої лексеми і її значень та нюансів, також широку базу детермінант лексико-семантичних варіантів.

Було б украй складно ретельно описувати кожну лексему й вивіряти її потенційні детермінанти, натомість значно простішим з цією метою видається використання автоматичної компіляції словників біграм, триграм (у майбутньому, можливо, n -грам), що значно заощаджує людські ресурси. Критерій вибору біграм і триграм ґрунтується на статистичному методі, адже за його допомогою можна відокремити типові n -грами від випадкових. На лінгвістичному порталі *movs.info* користувачі мають можливість згенерувати на основі корпусу української мови список 2-грам, 3-грам, 4-грам та 5-грам, а також отримати абсолютну частоту кожного сполучення слів [Корпус текстів української мови]; демонстраційна версія словника біграм Російського державного гуманітарного університету теж включає статистичні підрахунки [Автоматическая обработка]. У генераторі n -грам компанії *Google* можна проводити пошук n -грам для таких мов: англійська, іврит, італійська, китайська, німецька, російська, французька [Google n -gram Viewer]. Іншими словами, під час “розуміння” слова машина проводить його аналіз з урахуванням найближчого контексту і порівнює його з контекстом у репрезентативному корпусі (або навіть з корпусом паралельних текстів). Одним з завдань лексикографів є розробка таких ресурсів як для окремих мов, так і для мовних пар.

Словники біграм і триграм слугують насамперед як матеріал для комп'ютерів, адже їх використовує радше вузьке коло користувачів (хіба що інформатики, філологи-дослідники). Побіжно зазначимо, що автоматично укладені словники біграм і триграм можуть стати надійною основою для укладання словників сталих виразів, словників сполучуваності.

У сучасному нейромережевому аналізі змісту тексту (англійською – “text mining”) використання біграм і триграм теж виявляється ключовим. Так, для порівняння ступеня схожості текстів, його окремих фрагментів, потенційних варіантів перекладу один з поширених алгоритмів полягає у порівнянні суто словникового складу, так званий “bag of words” (у перекладі з англійської – “мішок слів”). Звичайно, за такого підходу абсолютно не враховується пунктуація, зв'язки між лексемами та їхній порядок у реченні (а тим паче – й у тексті), хоча для більшості випадків він справді дієвий. І він ще дієвіший, якщо замість порівняння окремих словоформ, під час перекладу аналізуються біграми або триграми фрагментів, а не окремі лексеми.

Перспективним для зняття неоднозначності видається підхід, в якому поєднується аналіз слова та слів у його контексті, коли в програму закладено і потенційні значення слова, і параметри контексту. Найпростіший аналіз контексту полягає у зборі даних щодо лівої і правої дистрибуцій слова та їхніх частот. Шляхом перебору варіантів машина може з високим ступенем імовірності виводити, який саме лексико-семантичний варіант реалізується в тому чи іншому вузькому контексті, біграмі. Цей метод, розроблений італійськими дослідниками Р. Навільї і П. Веларді, називається структурно-семантичним співвіднесенням ("structural semantic interconnection", англ.). Розробники цього методу вказують і на певні його недоліки: набір смислів лексичних одиниць у різних словниках і, відповідно, різних базах даних не збігається; окрема проблема полягає в тому, що цей алгоритм налаштований на надто високу точність зняття двозначності усіх лексем, у той час як часто для смислового аналізу документа достатньо зняття двозначності лише декількох слів запиту [Navigli].

Описаний тут метод дуже ефективний, хоча й далеко не ідеальний, адже детермінанта не обов'язково знімає двозначність: у біграмах "яскравий захід", "пам'ятати захід" детермінанти "яскравий", "пам'ятати" не усуває можливості різних прочитань іменника "захід". Здавалося б, такі казуси можна вирішити за допомогою триграм, однак проблема в тому, що при зростанні кількості компонентів сполучень слів, сама кількість сполучень зростає експоненційно. Якщо припустити, що в словнику мови міститься 100 000 лексем (а обсяг сучасних словників мають приблизно такий порядок), кількість теоретично можливих триграм становитиме 10^{15} , – і це якщо не враховувати граматичну варіативність словоформ²⁴.

Насправді в мовленні ніколи не реалізуються усі теоретично можливі варіанти сполучень, однак комп'ютерам навряд чи під силу відокремити теоретично можливі триграми від теоретично неможливих; більше того, у різного роду творчості автори художньої літератури, публіцистики час від часу вдаються до

²⁴Під час нейромережевого аналізу сукупність слів фрагмента тексту типово вміщується у вектор, розмірність якого дорівнює кількості потенційно можливих лексем, тобто, обсягу словникового складу слів. Однак опрацювання сотиссячного вектора стає непомірним завданням для техніки, тим паче, що в реченні, яке найчастіше стає предметом автоматичного аналізу, вжито лише незначну частину з усього словникового складу слів, отже, реченню найчастіше відповідає вектор з величезною кількістю нульових значень. Цей феномен отримав назва "прокляття розмірності".

навмисного вживання непоєднуваних компонентів, як у мовленнєвих видах мистецтва, так і в інших; свіжим словотвором, поєднанням непоєднуваного можуть характеризуватися взагалі й науково-технічні тексти, у зв'язку з постійним виникненням неологічних словосполучень. Отже, на нашу думку, поглиблення аналізу за рахунок збільшення довжини аналізованих сполучень слів, є екстенсивним, мало перспективним шляхом розвитку алгоритмів зняття двозначності.

Зняття лексичної двозначності шляхом граматичного аналізу. Розмітка тексту нерозривно пов'язана з процесом зняття двозначності. У комп'ютерній лінгвістиці набуває актуальності розподіл омонімів на внутрішньочастиномовні та міжчастиномовні [Дарчук 2015, с. 9], позаяк відомість про семантичні й граматичні категорії слова значно спрощує зняття неоднозначності. Так, у таблиці нижче показано, як працює демонстраційна версія програми розмітки текстів *FreeLing 4.0* на прикладі іспанського речення “Usted no nada nada²⁵”, що містить гру слів [An open source]:

Табл. 13. Результат автоматичної розмітки речення програмою FreeLing 4.0

Usted	No	Nada	Nada
usted	no	nadar	nada
PP2CS0P	RN	VMIP3S0	PI0CS
		01904293-v	
		01960911-v	

Програма присвоює кожному графічному слову певний код, який закладений розробникам в базу даних і заснований на статистичних критеріях сполучуваності. “PP”, наприклад, означає “propnombbre personal” – “особовий займенник”, “VMIP” – “неперехідне дієслово руху, особова форма”. Тим не менш, як

²⁵Словоформа “nada” може бути заперечним займенником (“нічого”) або ж особовою формою дієслова “nadar” (“плавати”). Переклад речення: “Ви зовсім не плаваєте”; букв.: “Ви не плаваєте нічого”. Лематизатор, закладений у програму, розпізнає цю міжчастиномовну омонімію.

показано на прикладі аналізу біграм, статистичний аналізі не є бездоганим. В реченні іспанською мовою "Hay que ver como **nada** ese chico" програма все ще вважає форму "nada" займенником, хоча в цьому контексті це омонімічне дієслово:

Табл. 14. Результат автоматичної розмітки речення програмою FreeLing 4.0

Hay	que	ver	como	nada	ese	chico	.
haber	que	ver	como	nada	ese	chico	.
VMIP	CS	VMN	CS	PI0C	DD0M	NCMS	Fp
3S0		0000		S00	S0	000	

Якщо граматичні одиниці практично вичерпно описані у довідниках, які, відповідно, доволі легко піддаються переведенню у формат електронних баз даних, лексика становить значно складніший, масштабніший і, без перебільшення можна сказати, невичерпний об'єкт, тому його аналіз потребує принципово інших підходів.

Проблема каталогізації лексико-семантичних значень у словниках. Справжнім викликом для сучасної лексикографії є, безперечно, створення списку або каталогу лексико-семантичних варіантів (або, як їх часто називають, "смислів", "семем", "семантичних множників", згідно з Ю. Карауловим [Караулов 1980], що цілком еквівалентно терміну А. Вежбицької "semantic primes" [Goddard, Wierzbicka 2014, с. 12]).

Не перше століття учених не полишає думка вивести базові лексичні одиниці. Наприклад, Л. Ервас Пандуро у 1787 р. довів універсальність 63 слів на матеріалі сотні мов (див. його працю "*Vocabolario poliglotta, con prolegomeni sopra più de CL lingue*" – "*Багатомовний словник з коментарями щодо більше 150 мов*"; до них належать назви частин тіла, а також деякі явища і предмети зовнішнього світу [цит. за. 2002, с. 59]. А.Б. Долгопольський вивів 15 термінів-претендентів на універсальні, що доведено на матеріалі двохсот мов усіх континентів [цит. за Ramies, 2002, с. 59]. Однак універсальні одиниці не обов'язково слугують плідними узагальнюючими поняттями: навряд чи лексику чи фразеологію можна класифікувати за групами: очі, тулуб, ноги тощо.

Однак залежно від сфери застосування та галузі науки зустрічаються й інші назви. Створення списку всіх лексико-семантичних варіантів без прив'язки до певної мови є мрією будь-якого лексикографа, адже такий список став би ідеальною

метамовою опису лексичної семантики або навіть мовою-посередником для машинного перекладу. Щонайменше, цей список можна було б використати для створення багатомовного універсального словника. Тоді кожній лексемі можна було б приписати одне або, найчастіше, декілька чітко визначених значень, а це в свою чергу дозволило б усувати двозначність усіх слів у тексті та здійснювати переклад, вільний від двозначних тверджень. Десятки, а то й сотні вчених світу намагалися прийняти цей виклик. Мабуть, найвідомішим досягненням на цій ниві стала праця групи дослідників під керівництвом польсько-австралійської вченої А. Вежбицької, що спричинилася до виявлення універсальних семантичних категорій.

Табл. 10. Універсальні семантичні категорії, згідно з А. Вежбицькою, К. Годдаром

I~ME, YOU, SOMEONE, SOMETHING~THING, PEOPLE, BODY	substantives
KIND, PART	relational substantives
THIS, THE SAME, OTHER~ELSE	determiners
ONE, TWO, SOME, ALL, MUCH~MANY, LITTLE~FEW	quantifiers
GOOD, BAD BIG, SMALL	evaluators and descriptors
KNOW, THINK, WANT, DON'T WANT, FEEL, SEE, HEAR	mental predicates
SAY, WORDS, TRUE	speech
DO, HAPPEN, MOVE,	actions, events, movement, contact
BE (SOMEWHERE), THERE IS, BE (SOMEONE)'S, BE (SOMEONE/SOMETHING)	location, existence, possession, specification
LIVE, DIE	life and death
WHEN~TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME, FOR SOME TIME, MOMENT	time
WHERE~PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE, INSIDE, TOUCH	space
NOT, MAYBE, CAN, BECAUSE, IF, VERY, MORE, LIKE~AS	logical concepts

[Goddard, Wierzbicka 2014, с. 12].

Як зазначають самі К. Годдар та А. Вежбицька, ця схема була перекладена багатьма мовами й перевірена на їхньому матеріалі, чим підтвердила свою чинність. Втім, у зв'язку з багатозначністю певних універсальних слів, а також з тим, що слова, на перший погляд рівнозначні в різних мовах, характеризуються відмінним набором сем і дистрибуцією (так званою алолексією), окремі з виявлених універсальностей окремих семантичних компонентів була поставлена під сумнів, а деякі з них не підтверджено й досі [Goddard, Wierzbicka 2014, с. 13-17].

Лексикографи дедалі більше намагаються вписати розрізнення лексико-семантичних варіантів і відтінків значень у строгі наукові рамки з чіткими критеріями, застосування яких складається в метод або комплекс методів. Окрім питання про якісний набір самих складників (так би мовити, “елементарних часток” значення), які можна визнати універсальними, гостро постає і проблема їхньої кількості у складі лексеми. А в комп'ютерний час наявність чіткого кількісного діапазону дало б змогу оптимізувати процес розрізнення лексико-семантичних варіантів слова. Питання розрізнення значень лексеми у комп'ютерних словниках розглядається у підрозділі 1.5.

Окрім виокремлення значень певної лексеми, злободенним питанням лексикографії є створення списку елементарних сем, за допомогою яких можна створювати дефініції слів. Частково проблему обмеження елементарних значень лексикографи намагалися вирішити під час формулювання дефініцій у тлумачних словниках. Упорядники словників цілком свідомі того,

що дефініції лексичних одиниць замкнені самі на собі, і найпростіше, чим вони можуть зарадити користувачеві – це звузити кількість лексем, що використовується в дефініціях, до списку, за яким в англійській лексикографічній традиції закріпилася назва “defining vocabulary” (“дефініційний словник”).

Описуючи дефініційний словник, зокрема, редколегія словника *The Longman American Defining Vocabulary* зазначає, що “слова для дефініцій обиралися виключно з заданого списку, а якщо в дефініції й міститься якесь стороннє слово, воно позначалося великими заголовними літерами. Дефініційний словник для тлумачень ретельно відібрано після вивчення таблиці частоти вживання англійських слів. Більше того, слова в дефініціях використовуються в їхньому основному значенні”²⁶ [The Longman]. У той самий час китайський вчений Хай Ксю з Гуандунського університету іноземних мов і зовнішньої торгівлі (КНР) застерігає, що підхід, заснований на використанні найчастіших слів у дефініціях, не позбавлений деяких недоліків через те, що, окрім часто вживаних слів, лексикографу для пояснення значення лексем також потрібні узагальнюючі поняття високого рівня абстракції, граматичні терміни. Тому вже в п’ятій версії *Longman Defining Dictionary* список слів, хоча й коротший на 154 одиниці в порівнянні з першою версією і становить у цілому 2094 позиції, покриває більшу кількість лексики в корпусі текстів, ніж попередній, здавалося б, масштабніший дефініційний словник [Hai Xu 2012]. І справді, такі слова як “суть”, “властивість”, “специфіка” не є найчастішими у пересічних мовців, однак безсумнівно часто зустрічаються в дефініціях понять. При всьому тому, дослідник вказує на недбалості в використанні як заявленого списку, так і значень, в яких використовуються дефініційні слова, а також їхньої частинимовної належності; поза тим, окремим проблемним питанням залишається використання в дефініціях словосполучень, що можуть мати самостійне значення, відмінне від значень слів, що входять до його складу [Hai Xu 2012]. Дійсно, значення сполучення слів на зразок “мати рацію” не так вже й легко вивести зі значень слів, що входять до їхнього складу, тому визначення рамок дефініційного словника виявляється значно складнішим, ніж видається на перший погляд.

Обсяг дефініційного словника теж поки що лишається питанням відкритим, хоча в основному лексикографи обмежуються приблизно порядком двох тисяч лексем. Аргументом на користь цієї кількості

²⁶Переклад наш – С.Ф.

може бути те, що опанування двох тисяч базових лексем, які не так вже й складно опанувати навіть в іноземній мові, дасть ключ до розуміння значенні усіх інших. А.П. Кові зауважує, що нічого магічного у числі 2 000 немає, адже на практиці кількість використаних для визначень слів суттєво занижується, і насправді лексем потрібно значно більше заявленого числа [Cowie 1999, с. 110].

Ще одне питання, яке варте розв'язання – чи можуть деякі значення мати універсальний характер, чи вони універсальні тільки частково, у той час як деякі з них неминуче марковані культурою мовця. Як зазначає Ю. Марчук, укладачі словників виходять з того, що у людства одне й те саме коріння, природа органів людського тіла та інтелекту в усіх народів, саме тому між лексикою та граматиною різних мов чимало схожих рис [Drounin 1866, цит. за Марчук, 197], хоча, як сказано вище, виявлення посправжньому універсальних лексичних значень не дало масштабних результатів.

Тим не менш, у термінографії, виходячи з принципів гармонізації, нормалізації та інтернаціоналізації, часом справді забезпечується майже універсальне впорядкування деяких терміносистем, як наприклад, у Міжнародній класифікації хвороб ВООЗ [Міжнародна класифікація хвороб]:

Табл. 11. Назви сколіозу, згідно з Міжнародною класифікацією хвороб ВООЗ (іспанською мовою):

M41	Escoliosis
M41.0	Escoliosis idiopática infantil
M41.1	Escoliosis idiopática juvenil
M41.2	Otras escoliosis idiopáticas
M41.3	Escoliosis toracogénica
M41.4	Escoliosis neuromuscular
M41.5	Otras escoliosis secundarias
M41.8	Otras formas de escoliosis
M41.9	Escoliosis, no especificada

[Código Internacional de Enfermedades]

Табл. 12. Назви сколіозу, згідно з Міжнародною класифікацією хвороб ВООЗ (українською мовою):

M41.0	Дитячий ідіопатичний сколіоз
M41.1	Ювенільний ідіопатичний сколіоз
M41.2	Інший ідіопатичний сколіоз
M41.3	Торакогенний сколіоз
M41.4	Нервово-м'язовий сколіоз
M41.5	Інший вторинний сколіоз
M41.8	Інші форми сколіозу
M41.9	Сколіоз, не уточнений

[Міжнародна класифікація хвороб]

Таким чином, зняття двозначності – важливий крок на шляху розмітки текстів. Розмітка графічних слів у тексті за семантичними параметрами суттєво збільшує можливості штучного інтелекту.

Семантична розмітка текстів. Ідея створення універсальних шифрів (або дескрипторів), дуже продуктивна для терміносистем, виявляється неприйнятною для опису лексики загального вжитку. Тим не менш, деякі розвідки в цій галузі виявилися доволі перспективними і заслуговують на згадку. Н. Добріч в 2010 році розробив систему тегів для морфологічних рис, семантичних і синтаксичних характеристик, а також для супровідних елементів [Dobrić 2010]. З-поміж релевантних семантичних характеристик дослідник виокремлює такі: 'людина', 'тварина', 'конкретна', 'машина', 'місце', 'кількість'. Для ілюстрації використання цих сем під час зняття двозначності розглянемо приклад. Іменник "toros" в іспанській мові може позначати сукупність тварин, хоча у множині він також метонімічно позначає "кориду", "бій биків", на відміну від української та інших мов. Таким чином, відсутність/наявність тегу <animal> допомогло б знімати двозначність і правильно перекладати форму "toros". Один з класиків світового масштабу в галузі лексичної семантики Ю.Д. Апресян пропонує значно більш деталізовану класифікацію типів значень, зокрема, 12 основних і 39 додаткових типів значень для іменників, 20 для дієслів, 23 для прикметників; до прикладу, для іменників вчений виокремлює такі регулярні значення: "дія – суб'єкт дії", "стан – причина стану", "властивість – суб'єкт властивості", "дія – об'єкт дії", "дія – результат дії" тощо [Апресян 1995, с. 193-214]. Хоча зазначені типи не дозволяють розрізняти лексеми в рамках однієї семантичної групи, їхнє застосування у семантичній розмітці має дозволити успішно знімати двозначність певної заданої лексеми.

Кожен з тегів можна було б зашифрувати цифрами або літерами, що дозволило б створити своєрідний каталог лексико-семантичних варіантів (одному лексико-семантичному варіанту відповідав би набір різних категорій). Однак це тільки на перший погляд полегшує життя лінгвістам і кібернетикам. “Відверто кажучи, просто вражає, якщо не обурює, як окремі лінгвісти (а також психологи та когнітивісти в цілому) схильні вірити, що як тільки вони перекладають те чи інше слово іншим “технічним” еквівалентом, завдання з семантичного аналізу на цьому завершуються; мало що може бути більш оманливим” [Goddard, Wierzbicka 2014, с. 10]. Проблема в тому, що наразі не відомо, який з пропонованих різними вченими списків сем найдоречніший для зняття неоднозначності лексеми в тексті. Крім того, завжди існує небезпека, що під один і той самий металінгвістичний опис можуть підпадати різні лексико-семантичні варіанти і лексеми. М. Марті Антонін, А. Фернандес Монравета, Г. Васкес Гарсія у свою чергу зазначають, що семантична розмітка корпусу вручну надто трудомістка і дорога, а готових шаблонів з лексичною інформацією, які могли б гарантувати якість результату, не існує [Martí Antonín].

Лексичне анотування у корпусах для зняття двозначності.

Більшість анотованих корпусів обмежуються частиномовною розміткою графічних слів, наприклад, анотований корпус Королівської академії іспанської мови [Corpus de Referencia del Español Actual. Versión Anotada], Генеральний регіонально анотований корпус української мови [ГРАК]. Злободенним дискусійним питанням розмітки корпусу залишається проблема вибору категорійної класифікації лексики, адже різні дослідники пропонують різні схеми. Н.П. Дарчук, дослідниця з Київського національного університету імені Тараса Шевченка, яка детально займається проблематикою семантичної розмітки текстів на матеріалі української мови, зазначає, що провідну роль у цих процесах відіграє використання тезаурусів [Дарчук 2015, 19]. Дослідження з розмітки текстів, у тому числі автоматичної й напівавтоматичної, були популярні на початку двотисячних років. А. Кіряков, Б. Попов, І. Терзієв, Д. Манов, до прикладу, скористалися лексико-семантичною онтологією, що охоплює приблизно 250 класів і 100 властивостей, що дало змогу укласти від 10^5 до 10^6 описів [Kiryakov 2004, с. 50-53].

Розробники проекту українськомовних інформатизованих ресурсів “lang-uk” розпочали створення загального анотованого

корпусу українською мовою, що включає розмітку сутностей. Наразі у розмітці використовуються такі сутності: ПЕРС (іменники та іменникові синтагми на позначення осіб), ЛОК (іменники та іменникові синтагми на позначення локативів), ОРГ (іменники та іменникові синтагми на позначення організацій), РІЗН – інші сутності [проект uk-lang]. На сайті Британського національного корпусу, зокрема, в підкорпусі Британського парламенту, що охоплює тексти від 1803 по 2005 рр., усю лексику розмічено за такими семантичними групами: світ, життя, здоров'я і захворювання, люди, тварини, рослини, їжа і напої, текстильні вироби й одяг, фізичні відчуття, матерія, існування і причинність, простір, час, рух, дія, властивості, надприродне, свідомість, увага і судження, добро і зло, емоції, бажання, наявність і володіння, мова, суспільство, побут і житло, озброєний конфлікт, влада, закон, мораль, виховання, віра, комунікація, подорож, праця і фах, торгівля і фінанси, дозвілля, які, в свою чергу, розподіляються на декілька рівнів субкатегорій і дрібних семантичних груп [British Parliament]. Однак детальна класифікація не вирішує всіх проблем, адже, приміром, одне й те саме поняття, як, наприклад, “exaggerate” зустрічається в зазначеному ресурсі у шести різних категоріях: свідомість, увага/судження, властивість, відчуття/емоції, відносне розташування у просторі, протяжність у просторі.

Анотування теми тексту й типу дискурсу. Окрім власне питання семантичної класифікації лексики, постала проблема можливості автоматичної розмітки текстів. У світі вже розроблено значну кількість відповідних програм, що мають функцію автоматичної розмітки текстів.

У нинішній час питання семантичної розмітки, доволі популярне в теоретичних розвідках першого десятиліття ХХІ сторіччя, вийшло в поле прикладних розробок, пов'язаних з необхідністю добування даних з текстів у мережі Інтернет, оптимізації їхнього пошуку. З цією метою необхідно було розробити класифікаційну сітку понять, явищ, характеристик, які підходили б для автоматичної розмітки різних текстів відповідними тегами. Огляд публікацій та надбань свідчить, що розробка питання поступово перейшла до комп'ютерних техніків і програмістів, а основними класифікаційними категоріями могли слугувати найтипівіші запити користувачів, класифікації сфер життя, запропоновані на провідних веб-сторінках і пошукових системах. При всій раціональності і практичній цінності такого підходу мусимо

визнати, що інтернетні класифікації зовсім не обов'язково відповідають найзагальнішим категоріям нашого мислення, а часто можуть бути продиктовані орієнтуванням на торгівлю і споживацтво. Поза тим, у проблематиці лексико-семантичного анотування тексту слід враховувати ще один вагомий чинник сучасної науки: вчені часто працюють на межі свого фаху і галузей інших наук; однак підхід програмістів до опрацювання текстів природної мови, так званий інженерний підхід, відрізняється від принципів аналізу, що розробляються лінгвістами. Як зазначає Н.П. Дарчук, розробник проблематики семантичної розмітки тексту в Україні: "саме лінгвістові треба підготувати і сформулювати вихідні дані (словники, граматики, алгоритми) так, щоб їх можна було оформити мовою програми, зрозумілої для комп'ютера, дотримуючись принципів гнучкості й багатоетапності аналізу" [Дарчук 2015, с. 19].

Розмітка здійснюється і з метою тематичного анотування текстів. У науковій літературі пропонуються різні ієрархії так званих 'типових сутностей' ('entity types'), розробленої для анотування частин текстів. Згадаємо деякі з них. Категорії *BBN* (*Bolt, Beranek and Newman*), запропоновані у 2002 р., використовуються для відповіді на питання і складаються з 29 типів та 64 субтипів [Brunstein 2002]. Ієрархія, що її запропонував С. Секайн у 2002 р., складається з 200 типів [Sekine 2016]. Пізніше, у 2011 р. А. Ріттер, С. Кларк, М. Етціоні розробили ієрархію, засновану на вільній базі даних типових сутностей у соціальних медіа під час приголошшливого експерименту щодо розпізнавання назв сутностей. Таке розпізнавання базувалося на формальному аналізі (наприклад, використання великої літери, що найвірогідніше охоплює власні назви, хоча й, як зазначають дослідники, у Твіттері вживання великої/малої літери набагато менш упорядковане, ніж у видрукованих книгах). Так, ієрархія, розроблена С. Секайном (Нью-Йоркський університет, США) у 2002 р., складається з 200 субтипів [Sekine 2016]. Спочатку дослідники розмітили вручну 2 400 інтернетних постів за 10 типами, взятими з "Freebase", що слугувало потім для тренування програми: з цією метою долучався словник, а також дані щодо дистрибуції тих чи інших типів [Ritter 2011].

Окрім тематичних критеріїв, тексти можна класифікувати і за іншими важливими з погляду досліджень критеріями, зокрема, соціолінгвістичними. Соціолінгвістична розмітка могла б дозволити робити автоматичну вибірку з матеріалу заданих характеристик і

досліджувати, приміром, мовлення чоловіків, жінок, підлітків, освічених, неосвічених осіб тощо. Корисною була б і розробка критеріїв автоматичного визначення соціолінгвістичних характеристик тексту з метою його розмітки. Сьогодні одним з перспективних і плідних напрямків досліджень є дискурс як об'єкт аналізу. Сучасні корпуси ще не містять тип дискурсу серед критеріїв пошуку, хоча така розмітка широко прислужилася б сучасним дослідникам. Проблема дискурсивної розмітки не лише в тому, що поняття це відносно нове, неоднозначно потрактоване, а подекуди й не сприйнята, а й тим, що один і той самий текст може належати до декількох дискурсів водночас. Отже, під час класифікації необхідно враховувати так званий "принцип перетинання" (термін В.П. Бєркова), тобто, використовувати багатокатегорійну класифікацію.

Таким чином, неоднозначність мовних і мовленнєвих одиниць виявляється основною проблемою розпізнавання природної мови комп'ютером. У комп'ютерній лінгвістиці розв'язання цієї проблеми ведеться за такими напрямками: зняття граматичної двозначності, зняття лексичної двозначності, семантична розмітка лексем у словниках, слів у текстах і корпусах. Для здійснення цих кроків необхідна розробка чіткої стандартизованої метамови опису граматичних і лексичних одиниць, їхніх сполучень. Суттєві і доволі достовірні висновки для зняття двозначності можна отримувати шляхом статистичного аналізу оточення лексеми в найближчому контексті (біграм, триграм або більших за масштабами сполучень слів).

5.3.3. Застосування комп'ютерної лінгвістики у системах штучного інтелекту

Практичне застосування природної мови у системах штучного інтелекту пов'язано насамперед з потребами комунікації техніки та її користувачів. Тому лінгвісти, інформатики, програмісти працюють над тим, щоб техніка коректно інтерпретувала людське мовлення і формулювала свої повідомлення теж доступними і зрозумілими для людей мовними знаками. Звісно, спосіб опрацювання мовлення роботом відрізняється від розуміння мови людиною. У той час як спілкування між людьми часто обумовлене обміном почуттями, комунікація технічного обладнання завжди зосереджена на виконанні певних дій, команд, операцій або їхньому невиконанні залежно від поставленого завдання. Тобто, спілкування з технікою переважно складається з команд та

інструкцій. Тому словникові статті, задіяні у комп'ютерних програмах, мають описуватися за допомогою спеціальної метамови, погодженої з інформатиками. У цьому підрозділі зосередимось на специфіці такої метамови.

З огляду на багатозначність лексичних одиниць, граматичних структур, людська мова недосконала для застосування в комп'ютерній техніці. Тому програмісти розробляють мови програмування; у мовах програмування високого рівня деякі фрагменти коду часом можуть дуже близько нагадувати природну мову людини, однак правила вживання команд, операцій і функцій у них строго лімітоване, обмежено спеціальним форматом. Приклади метамови, що застосовуються у базах даних, містяться у таблицях 6, 8 і схемі 9. Поки що метамови різного програмного забезпечення можуть суттєво відрізнятися, однак у цілому, з метою уніфікації опису лексики у різних базах даних можна рекомендувати користуватися стандартами Міжнародної організації зі стандартизації: ISO-1087, ISO-12620, ISO-24610, а також "Загальної онтології для опису мови" ("General Ontology for Linguistic Description", що часто скорочують до акроніма "GOLD") [General Ontology].

Комунікацію з технікою вести справді легко, коли йдеться про виконання простих стандартизованих команд. З іншого боку, життєвий досвід підказує, що невідготівлені користувачі цілком безпорадні перед виконанням елементарних операцій, приміром, зі створення текстового файлу, вирівнювання рядків, зміни шрифту, мови, друку тексту тощо. Таким чином, або користувачеві потрібні додаткові знання, або інтерфейс і мову повідомлень техніки необхідно краще адаптувати до мовлення людини.

Окремою проблемою є розуміння команд та інструкцій не лише в буквальному формулюванні, а й також у вигляді перифраз. Таким чином, необхідно вкладати в бази даних комп'ютерів відомості про семантичні зв'язки між лексемами, словосполученнями, реченнями, іншими формами мовленнєвих актів, механізми розпізнавання мовлення, лексичні й синтаксичні аналізатори, репрезентативні корпуси текстів. Деякі спрощені механізми формальних перифраз використовуються з метою розширення комп'ютерної перекладацької пам'яті [Gupta].

Над проблемою комунікації людей з технікою працює багато вчених. В Україні у цьому напрямку веде дослідження І.П. Біскуп [Біскуп]. Не вдаючись до перерахування нескінченного бібліографічного списку праць у цій надто динамічній галузі,

кількість яких щодня поповнюється, підкреслимо, що створення комп'ютерного лексикону, сумісного з мовою програмування і зрозумілої як для техніки, так і для людини у цій галузі, вкрай важливе, так само як і створення механізмів перифразування та знання двозначності.

Дуже цікавою розробкою на цій ниві є система автоматичного пошуку відповідей на питання "Вольфрам Альфа" [WolframAlpha]. На відміну від пошукачів, таких як *Google*, що у відповідь на запит видає список дотичних до ключових слів ресурсів, "Вольфрам Альфа" часто дає прямі відповіді на питання. Метамовою програми є англійська. У цій галузі становить інтерес веб-ресурс, гра "Акінатор", яка шляхом задавання до 25 питань "вгадує" ім'я і прізвище задуманого персонажа, відомої особи.

Одне з провідних застосувань комп'ютерної лінгвістики у системах штучного інтелекту – машинний переклад, про який докладніше сказано у наступному підрозділі.

5.3.4. Комп'ютерна лексикографія і системи машинного перекладу

На сьогодні багато програм машинного перекладу або вже містять у своїй базі електронні словники, або здатні інтегрувати нові лексикографічні ресурси. Зауважмо, що значна кількість сучасних електронних словників виходить за межі слова і словосполучення, як, наприклад, проект "Glosbe" [Glosbe] або "Linguee" [Contextual Search Engine Linguee]); і навпаки, багато сучасних програм комп'ютерного перекладу можуть бути використані як двомовні перекладні словники, такі як система "Google Translate". Це ілюструє, як різниця між електронними перекладачами та словниками поступово стирається, і не буде дивно, якщо у найближчому майбутньому ці знаряддя об'єднаються в одну комплексну систему.

В укладанні програм комп'ютерного перекладу лексикографія відіграє провідну роль, оскільки програмі потрібен спеціальний словник. В.С. Перебийніс так описує вимоги до комп'ютерного словника, задіяного у системі перекладу в Мовно-інформаційному фонді Інституту мовознавства УАН: "З тлумачного словника випикується слово з його тлумаченням. Якщо слово має кілька тлумачень, воно випикується стільки разів, скільки тлумачень у нього є. За тлумаченням слову приписується номер семантичної групи. Потім за синонімічним словником випикується синоніми до

реєстрового слова, яким можна приписати те саме тлумачення. Розташовуються вони за спадом семантичної близькості й усі разом утворюють семантичну групу, якій присвоюється певний номер. Далі зі словників вхідних мов, на які буде здійснюватися переклад, виписуються слова, що мають те саме значення, що й українське слово, і їхні синоніми, які теж розташовуються за спадом семантичної спільності. Цій семантичній групі присвоюється той самий номер, що й відповідній українській семантичній групі. За спільністю номерів семантичних груп знаходиться відповідник українського слова в даній мові” [Перебийніс 2009, с. 192].

Спеціально оформлюватися має і словникова стаття: “1) реєстрове слово у вихідній формі; 2) вказівка на лексичні омоніми, номери яких містяться в іншому словнику, на який у даній статті робиться посилання; 3) лексико-граматичний клас та його код (відсилка до таблиці словозмінних типів, їх ознак і номерів); 4) номер парадигматичного класу (відсилка до таблиці словозмінних типів, їх ознак і номерів); 5) тлумачення (номер семантичної групи та ранг слова у цій групі, у якій слова розташовані за спадом семантичної близькості)” [Перебийніс 2009, с. 193-194]. Інформація про слово, цікава і потрібна людині, відрізняється від тих відомостей, що необхідні комп’ютеру під час виконання перекладу. Тому, відповідно, перекладні словники, що виступають модулями машинного перекладу, до прикладу, в системі перекладу з англійської на російську “АМПАР”, містять відомості щодо формальних характеристик (слова, словосполучення змінні і незмінні, найбільш незмінна основа слова, омографи з відповідною частиномовною характеристикою), а також лексико-семантичні та семантико-граматичні характеристики частин мови (повна/неповна видова парадигма дієслова, спеціальні семантичні типи іменників, розряди прикметників) [Дарчук 2008, с. 311-317]. Уточнення, вдосконалення алгоритмів може передбачати зміну розмітки, її деталізацію; лексико-семантичні критерії можуть теж варіюватися: одне й те саме дієслово може, за різними класифікаціями, потрапляти і до дієслів мовлення, і до дієслів комунікації, вольової дії (наприклад, “наказувати”).

Оскільки існують численні теоретичні і практичні праці, присвячені машинному перекладу, зосередимось тут на ролі лінгвістики у створенні систем комп’ютерного перекладу.

Якісним стрибком у розвитку машинного перекладу стало втілення в життя ідеї використання комп’ютерної перекладацької

пам'яті, що реалізовано у широко відомих програмах "Trados", "Déjà Vu", "WordFast" та інших. Перекладацька пам'ять, або, як її часто називають за скороченням англійської назви, *ТМ* ("translation memory"), складається зі списку перекладених раніше фрагментів; можна сказати, що така пам'ять за своєю структурою нагадує величезний двомовний словник текстових фрагментів різного рівня.

На наш погляд, у контексті комп'ютерної лексикографії така перекладацька пам'ять становить неоціненну базу даних, що може бути матеріалом для створення словника. Однак з іншого боку, оскільки такі еквіваленти складаються зазвичай з перекладених речень (у пунктуаційному розумінні), а чимало речень сягають суттєвої довжини, вони часто просто обтяжують пам'ять. Адже дуже мало ймовірно, що колись трапиться таке саме або бодай на 90% схоже, складне поширене речення, а надто, якщо воно містить декілька підрядних або сурядних зв'язків. Л.Н. Беляєва стверджує, що потенціал таких паралельних текстів величезний, якщо провести вручну їхню розмітку та подрібнення [Беляєва, с. 87-88]. У свою чергу підкреслимо, що вказані процеси цілком можливо автоматизувати: достатньо мінімальних зусиль, щоб перетворити довгі фрагменти на багатющий двомовний глосарій біграм, триграм... *n*-грам та простих речень, потенціал використання яких вже значно у вищій, ніж у великих фрагментів.

Програма "ParsEqv" власне і призначена для подрібнення фрагментів перекладацької пам'яті; ступінь точності роботи програми становить від 80 до 95% [ParsEqv]. З аналогічною метою можна використовувати програми вибірки термінів з паралельних текстів (як, наприклад, "Araya", "SDL Multiterm Extractor"), хоча подібні програми мало поширені і характеризуються дуже високою вартістю на ринку програмного забезпечення; при використанні цих програм поза увагою залишаються словосполучення, сталі вирази, цілі речення. Результатом роботи програми є список двомовних еквівалентів, які можна перетворити у двомовний перекладний глосарій.

Ще один часто використовуваний принцип програм машинного перекладу – статистичний, який використовується, наприклад у системі "GoogleTranslate". Оскільки, як показує практика, модель послівного перекладу зазвичай призводить до незадовільних результатів, розробники почали використовувати модель "пофрагментного перекладу", а саме, перекладу за допомогою еквівалентів для словосполучень, біграм, триграм... *n*-грам. Бази даних для еквівалентів таких сполучень слів постійно поповнюються

по мірі того як з'являються нові переклади, і під час перекладу програма обирає той еквівалент для сполучення слів, який найчастіше зустрічається в базі даних. Логічно припустити, що, чим ширша база даних, тим частіше вибір комп'ютера буде влучним.

Новітнім методом, іншими словами, останнім словом у комп'ютерних технологіях стало використання штучних нейронних мереж, що не обминуло і галузь машинного перекладу. Як стверджує Д. Корбут, якщо раніше система "Google Translate" використовувала величезні бази даних для кожної мовної пари, то тепер використовується одна-єдина система для всіх мов. Наразі в системі перекладу 103 мови; проблемою цього підходу є те, що для якісного функціонування всіх цих комбінацій текстів (їхнє число дещо більше 10 000) необхідний масштабний матеріал для тренування штучної нейронної мережі, з якого вона виводитиме правила, відповідності; тому й якість перекладу за таким алгоритмом залежатиме від того, наскільки поширена та чи інша мовна пара, наскільки є можливість детально "натренувати" програму [Corbut 2017].

З-поміж програм машинного перекладу, що містять українську мову, слід назвати: "Pragma 6", яка включає знаряддя для перекладу як онлайн, так і в окремо встановлюваному програмному забезпеченні, вона працює з 51 мовою, серед яких англійська, французька, іспанська, італійська, каталонська, галісійська, баскська, німецька [Online Translator]. Система "Systran" працює в українсько-англійському та англо-українському напрямках перекладу. Для перекладу з російської мови на українську і навпаки найпопулярнішими є програми "Proling Office" і "Language Master". Перекладацьке програмне забезпечення, у якому задіяні найпоширеніші романські мови, детально описано у посібнику О.А. Ткаченка "Машинний переклад і редагування текстів для ОКР "магістр" програми "Переклад: французька, іспанська та італійська мови". Докладно принципи роботи, алгоритми й особливості упорядкування структурних модулів програм машинного перекладу розглянуто у праці Н.П. Дарчук "Комп'ютерна лінгвістика" [Дарчук 2008, с. 324-335].

Контрольні питання

1. Наведіть принцип контролю різних версій документа під час виконання спільного перекладацького проекту.
2. Назвіть, щонайменше, 4 формальні параметри, за якими може здійснюватися контроль якості перекладу.

3. Виконання яких часто повторюваних процесів можна оптимізувати під час використання сучасних текстових редакторів?
4. Які функції текстових редакторів допомагають під час збереження даних?
5. Чим відрізняється розрахунок обсягу тексту у текстових редакторах і в програмах машинного перекладу?
6. Які особливості формату тексту і файлу необхідно враховувати під час перекладу субтитрів за допомогою текстових редакторів?
7. Які корисні функції текстових редакторів можна використати на етапі редагування перекладу?
8. У чому полягають основні проблеми комунікації людини і комп'ютерної техніки?
9. Чим має відрізнятися лексикон машини від лексичної системи природної мови?
10. Як називається програма, що зводить словоформу до початкової словникової форми?
11. Як розпізнавання граматичних характеристик допомагає знімати лексичну двозначність при автоматичному аналізі текстів?
12. Що таке "міжчастиномовні омоніми"? Яким чином їх можна розпізнавати під час автоматичного аналізу тексту?
13. Які існують труднощі семантичної розмітки тексту?
14. Які існують класифікації лексико-семантичних категорій, застосування яких доречно для автоматичного аналізу тексту?
15. У чому причини недосконалості машинного перекладу?
16. За якими критеріями можливо здійснювати формальну перевірку якості перекладу?
17. За яким принципом побудована перекладацька комп'ютерна пам'ять?
18. Якого типу метамовна розмітка тексту допомагає в автоматичному аналізі тексту?
19. Назвіть 3-4 поширені алгоритми систем машинного перекладу?
20. Які існують шляхи збагачення перекладацької комп'ютерної пам'яті?

РЕКОМЕНДОВАНІ ДЖЕРЕЛА

1. Navigli R., Velardi P. Structural Semantic Interconnections: a Knowledge-Based Approach to Word Sense Disambiguation // IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). – 2005. n° 27(7) – p. 1075-1086.

2. Баранов А.Н. Категории искусственного интеллекта в лингвистической семантике. Фреймы и сценарии // Актуальные проблемы прикладного языкознания. – М., 1987.
3. Беляева Л.Н. Корпусная лингвистика и перевод: потенциал и ограничения / Л.Н. Беляева // Труды международной конференции "Корпусная лингвистика – 2011". – СПб: Изд-во филологического ф-та СПбГУ, 2011. – С. 86-91.
4. Біскуп І.П. Лінгвістичне моделювання знань у штучних мовах і мовах програмування // Мовні і концептуальні картини світу. К. – 2010. – Вип. 32. – С. 47-51. Данилюк І. Корпус текстів для вивчення граматичної службовості / І.Данилюк // Лінгвістичні студії. – Вип. 26. – Донецьк : ДонНУ, 2013. – С. 224-229.
5. Дарчук Н.П. Комп'ютерна лінгвістика (автоматичне опрацювання тексту): підручник. – К.: ВПЦ "Київський університет", 2008. – 351 с.
6. Дарчук Н.П. Лінгвістичне забезпечення автоматичних систем аналізу українськомовного тексту (на прикладі системи автоматичного граматичного аналізу тексту АГАТ) : автореф. дис. ... д-ра філол. наук / Дарчук Наталія Петрівна ; М-во освіти і науки України, Київ. нац. ун-т ім. Тараса Шевченка, Ін-т філології. – К., 2015. – 34 с.
7. Караулов Ю.Н. Лингвистическое конструирование и тезаурус литературного языка. – М.: Наука, 1981. – 367 с.
8. Марчук Ю.Н. Компьютерная лингвистика. Учебное пособие. – М., 2006 – 317 с.
9. Скороходько Е.Ф. Лінгвістичні основи автоматизації інформаційного пошуку / Е.Ф. Скороходько. – К., 1970. – 242 с.
10. Скороходько Э.Ф. Семантические сети и автоматическая обработка текста. – К.: Наукова думка, 1983. – 128с.
11. Ткаченко О.О. Машинний переклад і редагування текстів для ОКР "магістр" програми "Переклад: французька, іспанська та італійська мови" – К.: КНУ імені Тараса Шевченка, 2017. – 144 с.
12. Фокін С.Б. Засоби автоматизованого збагачення комп'ютерної перекладацької пам'яті // Мовні і концептуальні картини світу. – К.: ВПЦ "Київський університет", 2016.– Вип. 57. – С. 384-391.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

Теоретичні праці

1. Андрієвська Е.М. Сучасна електронна лексикографія в роботі перекладача: навчальний посібник у 2-х част. – К.: КНУ ім. Тараса Шевченка, 2016. – Ч.1. – 87 с.; Ч. 2. – 47 с.
2. Апресян Ю.Д. Лексическая семантика. Синонимические средства языка. – М.: Наука, 1995. – 367 с.
3. Баевский В.С. Лингвистические, математические, семиотические и компьютерные модели в истории и теории литературы. – М.: Языки славянской культуры, 2001. – 336 с.
4. Баранов А.Н. Категории искусственного интеллекта в лингвистической семантике. Фреймы и сценарии // Актуальные проблемы прикладного языкознания. – М., 1987.
5. Беляева Л.Н. Корпусная лингвистика и перевод: потенциал и ограничения / Л.Н. Беляева // Труды международной конференции "Корпусная лингвистика – 2011". – СПб: Изд-во филологического ф-та СПбГУ, 2011. – С. 86-91.
6. Берков В.П. Двухязычная лексикография. – М.: Транзиткнига, 2004. – 237 с.
7. Біскуп І.П. Лінгвістичне моделювання знань у штучних мовах і мовах програмування // Мовні і концептуальні картини світу. К. – 2010. – Вип. 32. – С. 47-51.
8. Боровинський І.М. Відтворення образної своєрідності поезій англійських та американських романтиків в українських перекладах [Автореферат] : автореф. дис. ... канд філологічних наук: 10.02.16 / Боровинський І.М. – К.: КНУ імені Тараса Шевченка, 2011. – 19 с.
9. Бродська Р. Лексико-стилістичний аналіз тексту методом статистичної лексикографії: навч. посібник / Р. В. Бродська ; Чернівецький національний ун-т ім. Юрія Федьковича. – Чернівці: Рута, 2003. – 200 с.
10. Василик А.Р. Стратегія М. Рудницького в контексті українського художнього перекладу ХХ ст.: автореф. дис. ... канд. філол. наук: 10.02.16 / Василик Анастасія Романівна ; М-во освіти і науки, молоді та спорту України ; Київ. нац. ун-т ім. Т. Шевченка. – Київ, 2012. – 20 с.

11. Виноградов В.В. Лексикология и лексикография. – М.: Наука, 1977. – 318 с.

12. Вихованець І. Р. Лексикографічно-граматичне опрацювання прийменників української мови / І.Р. Вихованець // Українська мова: наук. журнал. 2002. – № 4. – С. 90-96.

13. Вихованець І.Р. Нариси з функціонального синтаксису української мови. – К.: Наукова думка, 1992. – 224 с.

14. Воеводина Т.В. О соответствии перевода жанровым традициям ПЯ в свете социологической теории перевода // Тетради переводчика. – М.: – 1979. – Вып.16. – С.83-91.

15. Гак В.Г. О разных типах двуязычных словарей // Тетради переводчика. – М., 1964 – Вып. 2. – С. 71-78.

16. Гак В.Г. Сопоставительная лексикология: на материале французского и русского языков / В.Г. Гак. – М.: Междунар. Отношения, 1977. – 264 с.

17. Гетьман З.О., І.С. Орлова. Теорія і практика перекладу з іспанської мови на українську: навчальний посібник для студ. вищих навч. закладів / З.О. Гетьман, І.С. Орлова. – К.: ВПЦ "Київський університет", 2010. – 223 с.

18. Городенська К. Граматична лексикографія: стан і перспективи // Українська мова : науково-теоретичний журнал / Нац. акад. наук України, Ін-т укр. мови ; редкол.: П. Гриценко, К. Городенська, О. Яцевська [та ін.]. – Київ, 2018. – № 1 (65). – С. 30-37.

19. Гринев С.В. Основы лексикографического описания терминосистем: автореф. дис. докт. филол. наук. М., 1990. – 43 с.

20. Данилюк І. Корпус текстів для вивчення граматичної службовості / І.Данилюк // Лінгвістичні студії. – Вип. 26. – Донецьк : ДонНУ, 2013. – С. 224-229.

21. Дарчук Н.П. Комп'ютерна лінгвістика (автоматичне опрацювання тексту): підручник. – К.: ВПЦ "Київський університет", 2008. – 351 с.

22. Дарчук Н.П. Дослідницький корпус української мови: основні засади і перспективи / Н.П. Дарчук // Вісник Київського нац. ун-ту ім. Т. Шевченка. Серія: Літературознавство. Мовознавство. Фольклористика. – К.: ВПЦ "Київський університет", 2010. – № 21. – С. 45-49.

23. Дарчук Н.П. Лінгвістичне забезпечення автоматичних систем аналізу українськомовного тексту (на прикладі системи автоматичного граматичного аналізу тексту АГАТ) : автореф. дис. ... д-ра філол. наук / Дарчук Наталія Петрівна ; М-во освіти і науки України, Київ. нац. ун-т ім. Тараса Шевченка, Ін-т філології. – К., 2015. – 34 с.

24. Дарчук Н.П., Лангенбах М.О., Сорокін В.М., Ходаківська Я.В. Паралельний корпус текстів ПарКУМ. Науковий часопис Національного педагогічного університету ім. М.П. Драгоманова. Серія 9. Сучасні тенденції розвитку мов, 2017. – Вип. 15. – С. 28-35.

25. Дарчук Н.П. Можливості семантичної розмітки корпусу української мови (КУМ) / Н.П. Дарчук // Науковий часопис Національного педагогічного університету імені М.П. Драгоманова. Серія 9 : Сучасні тенденції розвитку мов : зб. наук. праць. – Київ : Вид-во НПУ імені М.П. Драгоманова, 2017. – Вип. 15. – С. 18-28.

26. Денисов П.Н. Полисемия и классификация лексики / П. Н. Денисов // Словарные категории. – М., 1988. – С. 15-23.

27. 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.

28. Дорошенко О.Ю. Відтворення особливостей художнього дискурсу Хуліо Кортасара в українському перекладі : автореф. дис. ... канд. філол. наук : 10.02.16 / Дорошенко Ольга Юріївна. – М-во освіти і науки України, Київ. нац. ун-т ім. Тараса Шевченка. – Київ, 2017. – 18 с.

29. Засєкін С.В. Психолінгвістичні універсалії перекладу художнього тексту / С. В. Засєкін. – Луцьк: Волин. нац. ун-т ім. Лесі Українки, 2012. – 274 с. – [Електронний документ]. Режим доступу: http://esnuir.eenu.edu.ua/bitstream/123456789/3214/3/monograph_editor.pdf

30. Ідеофраз, Ономазіологічний багатомовний електронний словник фразеологічних синонімів // С.Б. Фокін. Свідectво про реєстрацію авторського права на твір № 78269 від 12.04.2018 Міністерства економічного розвитку і торгівлі України.

31. Комп'ютерна лексикографія, її предмет та завдання // Карпіловська Е.А. Вступ до прикладної лінгвістики: комп'ютерна лінгвістика. – Донецьк: Юго-Восток, 2006. – С. 54-74.

32. Калустова О.М. Відтворення частки ‘же’ при перекладі іспанською як методологічна проблема // Проблеми семантики слова, речення та тексту. Зб. наук. праць. – К., 2002. – Вип. 9. – С. 130-135.

33. Калустова О.М. Електронні бази паралельних та співвідносних текстів у перекладознавчих дослідженнях // Вісник Сумського держ. ун-ту. Філологічні науки. – Суми, 2004. – №1 (60). – С. 130-136.

34. Калустова О.М. Мова перекладу та перекладна мова // Науковий журнал "Культура народів Причорномор'я". – Сімферополь: – 2004. – № 50. – С. 105-109.
35. Караулов Ю.Н. Общая и русская идеография / Юрий Николаевич Караулов. – М.: Наука, 1976. – 355 с.
36. Караулов Ю.Н. Частотный словарь семантических множителей русского языка. – М., 1980. – 280 с.
37. Караулов Ю.Н. Лингвистическое конструирование и тезаурус литературного языка. – М.: Наука, 1981. – 367 с.
38. Карпіловська Є. Українська комп'ютерна лінгвістика сьогодні. [Електронний документ станом на 23.04.2013]. Режим доступу: <http://www.polit.ua/lectures/2011/01/19/karpilovskaya.html>
39. Касарес Х. Введение в современную лексикографию. – М.: Изд-во иностранной лит-ры, 1958. – 354 с.
40. 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>
41. Клименко Н.Ф. Нові підходи до укладання комп'ютерних словників // Мовознавство. – К., 1996. – № 4-5. – С. 11-15.
42. Ковалёва К.И. Оригинал и перевод: качественные и количественные различия // Перевод в современном мире: Сб. статей. – М.: Всерос. центр переводов, 2001. – С. 27-39.
43. Ковтуненко Л.С. Комп'ютерні аспекти лексикографічних систем // Мовознавство. – К., 1996. – № 4-5. – С. 28-34.
44. Крупнов В.Н. Лексикографические аспекты перевода: Учеб. пособие для ин-тов и ф-тов. иностр. яз. – М.: Высш. шк., 1987. – 192 с.
45. Комп'ютерна лексикографія // Селіванова О.О. Сучасна лінгвістика: напрями та проблеми: підручник. – Полтава: Довкілля-К., 2008. – С. 660-664.
46. Комунікативна ситуація, комунікативний акт, дискурс // Селіванова О.О. Сучасна лінгвістика: напрями та проблеми: підручник. – Полтава: Довкілля-К., 2008. – С. 660-664.
47. Куконіна Н. А. Зіставлення словника перекладу зі словником оригіналу як перекладознавча проблема (на матеріалі перекладів шекспірівського "Гамлета"): автореф. дис. на здобуття наукового ступеня канд. філол. наук, спец.: 10.02.16 – перекладознавство / Н. А. Куконіна, Київський національний ун-т ім. Тараса Шевченка. – К., 2002. – 20 с.

48. Курс письмового перекладу. Іспанська мова: навч. посіб. для студ. вищ. навч. закл. / [О.М. Калустова, І.М. Шиянова, К.А. Танич, Н.Ю. Квасюк, М.О. Коломієць. – К: ВПЦ "Київський університет", 2012. – 271 с.

49. 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.

50. Марчук Ю.Н. Компьютерная лингвистика. Учебное пособие. – М., 2006 – 317 с.

51. Найда Ю. Анализ значения слов и составление словарей / Юджен Найда // Новое в зарубежной лингвистике. – М., 1962. – Вып. 2. – С. 45-71.

52. Найда Ю. Наука переводить // Вопросы языкознания. – М., 1970. – № 4. – С. 3-14.

53. Некряч Т.Є. Деталь у цілісній картині світу Е. Гемінґвея: проблеми перекладу / Т. Некряч, Р. Довганчина // Вісник Київського національного університету імені Тараса Шевченка. Іноземна філологія. – 2014. – Вип. 47. – С. 19-23. – [Електронний документ]. Режим доступу: http://nbuv.gov.ua/UJRN/VKNU_if_2014_1_8

54. Новикова М.А. Проблема индивидуального стиля в теории художественного перевода: автореф. ... дис. д-ра филол. наук. – Л.: Ленинград. гос. ун-т им. А.А. Жданова, 1980. – 27 с.

55. Носенко И.А. Начала статистики для лингвистов / И. А. Носенко. – М.: Высш. школа, 1981. – 156 с.

56. Орлова І.С. Термінознавство. Іспанська мова. Навчальний посібник. – К.: Освіта України, 2013. – 136 с.

57. Перванов Я.А. Материалы к русско-болгарскому сопоставительному словарю. [Електронний документ]. Режим доступу: <https://arxiv.org/pdf/1104.2034.pdf>

58. Перебийніс В.С. Частотні словники та їх використання / В.С. Перебийніс, М.П. Муравицька, Н.П. Дарчук. – К.: Наукова думка, 1985. – 204 с.

59. Перебийніс В.І., Сорокін В.М. Традиційна та комп'ютерна лексикографія. Навч. посібник. – К.: Вид. центр КНЛУ, 2009. – 218 с.

60. Перебийніс В.І., Бобкова Т.В., Гриднєва Л.М., Лебедев К.М. Морфологічне кодування англо-українського корпусу паралельних текстів офіційних документів // Науковий вісник Чернівецького університету № 441-443. – Чернівці: "Рута", 2009. – С. 166-171.

61.Перебийніс В.С. Статистичні методи для лінгвістів. – Вінниця: Нова книга, 2013. – 176 с.

62.Пиотровский Р.Г. Информационные измерения языка / Р.Г. Пиотровский. – Ленинград, 1968. – 116 с.

63. Проект lang-uk. – [Електронний документ]. Режим доступу від 21.10.2018: <http://lang.org.ua/uk/corpora/>

64. Пшенична Л., Шевченко В, Шишкіна Н. Термінологічна культура фахівця – освітньо-навчальний проєкт / Л. Пшенична, В. Шевченко, Н. Шишкіна. – [Електронний ресурс]. Режим доступу: <http://ena.lp.edu.ua:8080/bitstream/ntb/15263/1/8-34-41.pdf>

65. Рецкер Я.И. Плагиат или самостоятельный перевод? (Об одной судебной экспертизе) / Я.И. Рецкер // Тетради переводчика. – М., 1963. – Вып. 1, 1963. – С. 42–64.

66. Румянцева Е.А. Аспекты разработки двуязычных учебных электронных словарей [Електронний документ]. Режим доступу: <http://www.dialog-21.ru/digests/dialog2006/materials/html/Rumyanceva.htm>

67. Савчин В. Лукашіана Анатолія Перепаді / В. Савчин // Всесвіт: журнал іноземної літератури. – № 11/12, 2011. – С. 215-221.

68. Седлорова А.І. Мовний та позамовний контекст перекладацького доробку В. Підмогильного (на матеріалі перекладів з французької мови): автореф. дис. ... канд. філол. наук : 10.02.16 / Седлорова Аліна Ігорівна; Київ. нац. ун-т ім. Тараса Шевченка. – Київ, 2013. – 19 с.

69.Селегей В. Электронные словари и компьютерная лексикография. [Електронний документ]. Режим доступу: www.lingvoda.ru

70.Семантична деривація // Селіванова О.О. Сучасна лінгвістика: термінологічна енциклопедія. – Полтава: Довкілля-К., 2006. – С. 528.

71.Сервантес Сааведра М. Премудрий гідальго Дон Кіхот з Ламанчі (Частина I) / пер. М. Лукаша. – К.: Дніпро, 1985. – 704 с.

72. Сірук О.Б. Тезаурус дієслів української мови: лінгвістичні проблеми та методика конструювання: Автореф. дис. ...канд. філол. наук: Спец. 10.02.01 / О.Б. Сірук. – К.: КНУ імені Тараса Шевченка, 2006. – 20 с.

73. Скороходько Е.Ф. Лінгвістичні основи автоматизації інформаційного пошуку / Е.Ф. Скороходько. – К., 1970. – 242 с.

74.Скороходько Э.Ф. Семантические сети и автоматическая обработка текста. – К.: Наукова думка, 1983. – 128 с.

75. Сніжко Н.В. Зведений словник української мови: джерела і засади укладання / Н. Сніжко // *Культура слова*. – 2015. – Вип. 83. – С. 90-99. – Режим доступу: http://nbuv.gov.ua/UJRN/Kuls_2015_83_18
76. Сніжко Н. Структурна, функціональна та хронологічна параметризація лексики у зведеному словнику, тезаурусі й електронній картотеці // *Українська мова : науково-теоретичний журнал* – К., 2016. – № 4 (60). – С. 87-102.
77. Сухенко К.М. Фактор частотности и перевод // *Теорія і практика перекладу. Республ. міжвід. наук. зб.* – К., 1981. – Вип. 5. – С. 80-83.
78. Український національний корпус. – [Електронний документ]. Режим доступу: http://lcorp.ulif.org.ua/virt_ulnc/
79. Тарасова В.В. Лексикографічна концепція двомовного термінологічного тезауруса фахової мови психології // *Науковий вісник Національного університету біоресурсів і природокористування України: збірник наукових праць. Серія: філологічні науки*. – К.: Національний університет біоресурсів і природокористування України, 2013. – Вип. 186, ч. 2. – С. 23-30.
80. Ткаченко О.О. Нові електронні ресурси на допомогу перекладачам // *Мова і культура*. – К.: Видавничий дім Дмитра Бурого, 2011. – Вип. 14. – Т. VI (152). – С. 357-366.
81. Ткаченко О.О. Програмне забезпечення для перекладача-фрілансера // *Мовні і концептуальні картини світу: зб. Наук. пр. Вип.2(44)*. – К.: ВПЦ «Київський університет», 2013 р. – С. 309-316.
82. Ткаченко О.О. Огляд програмного забезпечення для перекладачів станом на 2009 рік // *Мова і культура*. – К.: Видавничий дім Дмитра Бурого, 2010. – Т. 135. – С. 111-126.
83. Ткаченко О.О. Машинний переклад і редагування текстів для ОКР “магістр” програми “Переклад: французька, іспанська та італійська мови” – К.: КНУ імені Тараса Шевченка, 2017. – 144 с.
84. Флорин С. Чем словари не удовлетворяют переводчика? / С.Флорин // *Мастерство перевода*. – М.: Советский писатель, 1975. – С. 397-398.
85. 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.

86. Фокін С.Б. Іспанські відповідники прислівника “довго”: перекладознавчий та лексикографічний аспекти / С.Б. Фокін // Мовні і концептуальні картини світу. – К., – 2010 – Вип. 31. Том. – С. 172-177.

87. Фокін С.Б. Критерій достатності корпусу при розрахуванні фактору частотності в перекладі / С.Б. Фокін. // Вісник Харківського університету: Романо-германська філологія. Методика викладання іноземних мов. – Харків: Харківський національний університет імені В.Н. Каразіна, 2011. – Вип. 68. – С. 200-207.

88. Фокін С.Б. Дистрибутивний аналіз при укладанні двомовних перекладних словників (на прикладі українсько-іспанських відповідників поля “освіта” // Проблеми семантики, прагматики та когнітивної лінгвістики – К.: Київський національний університет імені Тараса Шевченка, 2012. – Вип. 21. – С. 490-500.

89. Фокін С.Б. Хибні друзі перекладача як наслідок словотвірної інтерференції // Мовні і концептуальні картини світу. – К.: ВПЦ “Київський університет”, 2013. – С. 192-200.

90. Фокін С.Б. Частотні параметри перекладної мови в загальній та частковій теоріях перекладу // Фаховий та художній переклад: теорія, методологія, практика. Матеріали доповідей VI Міжнародної науково-практичної конференції 05.04-06.04.2013 – К.: Аграр Медіа Груп, 2013. – С. 389-396.

91. Фокін С.Б. Дисфемізми в художньому перекладі: еквівалентність у мікро- та макроконтексті / С.Б. Фокін. // Мовні і концептуальні картини світу. – К.: ВПЦ “Київський університет”, 2014. – Вип. 47. – Ч. 2. – С. 473-484.

92. Фокін С.Б. Маркери дисфемізмів у художньому тексті (на матеріалі іспанської та української мов) / С.Б. Фокін. // Проблеми семантики, прагматики та когнітивної лінгвістики. – К.: Логос, 2014. – Вип. 25. – С. 462-472.

93. Фокін С.Б. “Улюблене” слово перекладача: статистичний та макростилістичний аспекти // Мова і культура. Науковий журнал. – К.: Видавничий дім Дмитра Бураго, 2014. – Вип. 17. – Т. III (171). – С. 292-298.

94. Фокін С.Б. Автоматизація вибірки анафоричних повторів для перекладознавчих і контрастивних досліджень // Вісник Київського національного університету. Серія: Іноземна філологія. – К.: ВПЦ “Київський університет”, 2015. – Вип.1(48).– С. 24-28.

95. Фокін С.Б. Паралелізм та розбіжності суфіксів *-ин(-in)/-ina, -um(-im)/-ita* в українському та іспанському термінотворенні // Фаховий та художній переклад: теорія, методологія, практика: збірник наукових праць. – К.: Аграр Медіа Груп, 2015. – С. 316-322.

96. Фокін С.Б. Переваги корпусів малого обсягу для досліджень макроявищ у перекладі // Мовні і концептуальні картини світу. – К.: ВПЦ “Київський університет”, 2015. – Вип. 51. – С. 590-597.

97. Фокін С.Б. Вибір дескрипторів для ономазіологічного словника: питання методології // П'ята Всеукраїнська наукова конференція романістів. Структурно-семантичні і когнітивно-дискурсивні парадигми сучасного романського мовознавства. – Одеса: Одеський національний університет, 2015. – С. 71-72.

98. Фокін С.Б. Контрастивне дослідження категорії повторюваності: нові завдання для корпусної лінгвістики // Мовні і концептуальні картини світу. – К.: ВПЦ “Київський університет”, 2015.– Вип. 55. – Ч. II. – С. 348-356.

99. Фокін С.Б. Макрохарактеристики мовлення персонажу в оригіналі і перекладі (на матеріалі діалогів Дон Кіхота і Санчо Панси) // Вісник Київського університету. Серія: Іноземна філологія. – К.: ВПЦ “Київський університет”, 2016. – Вип. 49. – С. 30-34.

100. Фокін С.Б. Статистичні ознаки імітації стилю перекладача (на матеріалі використання префіксального словотвору в перекладі “Дон Кіхота” М.Лукаша і А.Перепаді // Фаховий та художній переклад: теорія, методологія, практика: збірник наукових праць. – К.: Аграр Медіа Груп, 2016. – С. 391-398.

101. Фокін С.Б. Засоби автоматизованого збагачення комп'ютерної перекладацької пам'яті // Мовні і концептуальні картини світу. – К.: ВПЦ “Київський університет”, 2016.– Вип. 57. – С. 384-391.

102. Фокін С.Б. Двомовні перекладні словники: причини недосконалості та перспективи оптимізації // Мовні і концептуальні картини світу. – К.: ВПЦ “Київський університет”, 2017. – Вип. 59. – С. 214-220.

103. Фокін С.Б. Автоматизована вибірка міжмовних еквівалентів з паралельних текстів: алгоритм та перспективи застосування // Фаховий та художній переклад: теорія, методологія, практика: збірник наукових праць. – К.: Аграр Медіа Груп, 2017. – С. 399-405.

104. Фокін С.Б. Комп'ютерна лексикографія і переклад: іспанська та українська мови. Навчальний посібник. – К.: Капрі, 2017. – 152 с.

105. Фокін С.Б. Глибина кругових визначень у словнику як віддзеркалення властивості лексичної системи // Фаховий та художній переклад: теорія, методологія, практика: збірник наукових праць. – К.: Аграр Медіа Груп, 2018. – С. 370-376.

106. Хорохорин Л.Г. Некоторые замечания о многозначности и недифференцированности слов в двуязычных словарях // Тетради переводчика. – М., 1964 – Вып. 2. – С. 91-99.

107. Чередниченко О.І. Динаміка французького словника і комп'ютерна лексикографія // Про мову і переклад. – К.: Либідь, 2007. – 248 с.

108. Щерба Л.В. Опыт общей теории лексикографии // Языковая система и речевая деятельность. / Лев Владимирович Щерба. – Л.: Наука, 1974. – С. 290–297.

109. Широков В.А. Комп'ютерна лексикографія / В.А. Широков. – К.: Наукова думка, 2011. – 351 с.

110. Шиянова І.М. Засоби відтворення функціональної перспективи речення в іспансько-українському перекладі : автореф. дис. ... канд. філол. наук: 10.02.16 / Шиянова Ірина Михайлівна ; М-во освіти і науки, молоді та спорту України, Київ. нац. ун-т ім. Тараса Шевченка. – Київ, 2013. – 20 с.

111. The Longman American Defining Dictionary. – [Електронний ресурс]. Режим доступу: http://www.longmandictionariesusa.com/res/shared/vocab_definitions.pdf

112. Aburmada I. La lengua de la ciencia y los diccionarios terminológicos. – [Електронний документ]. Режим доступу: http://digital.csic.es/bitstream/10261/22978/1/BFEU_2008.pdf

113. 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.

114. Alesina N.M., Vinogradov V.S. Teoría y práctica de la traducción. El español. – К.: Вища школа, 1993. – 207 p

115. Alonso Ramos M., Sanromán Vilas B. Construcción de una base de datos de colocaciones léxicas. – [Електронний документ]. Режим доступу: <http://www.sepln.org/revistaSEPLN/revista/26/alonso.pdf>

116. Alvar Ezquerro M. Criterios de diseño para la creación de corpórea // Alvar Ezquerro, M.- Villena Ponsoda, J.A. (Coords.) Estudios para un corpus del español. – Málaga: Universidad de Málaga, 1994. – P. 31-40.

117. Atril Solutions. Using Regular Expressions. – [Електронний документ]. Режим доступу: <https://atrilolutions.zendesk.com/hc/en-us/articles/205539861-Useful-regular-expressions>

118. Baker M. Corpus-Based Translation Studies: The Challenges that Lie Ahead // Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C.Sager. Amsterdam: John Benjamins, 1997. – P. 175-186.

119. Biçici E. Dynamic translation memory: Using statistical machine translation to improve translation memory fuzzy matches / E. Biçici, M. Dymetman // *Computational Linguistics and Intelligent Text Processing*, 2008. – P. 454-465.
120. Bowker L. Towards a Methodology for a Corpus Based Approach to Translation Evaluation / L. Bowker // *Meta. Revue des Traducteurs*. – 2001. – n° 46, Vol- 2. – [Электронный документ]. Режим доступа: <http://www.erudit.org/revue/meta/2001/v46/n2/002135ar.pdf>
121. Burnard L. Metadata for Corpus Work / L. Burnard. – [Электронный ресурс]. Режим доступа: <http://users.ox.ac.uk/~lou/wip/metadata.html#HDR>
122. Brunstein A. Annotation guidelines for answer types. BBN technologies. – [Электронный ресурс]. Режим доступа: <https://catalog ldc.upenn.edu/docs/LDC2005T33/BBN-Types-Subtypes.html>
123. Cabré M.T. La terminología y la traducción especializada // *Manual de documentación y terminología para la traducción especializada*. – Madrid: Arco/Libros. – p. 89-122.
124. Casares J. Introducción a la lexicografía moderna / J. Casares. – Madrid: Consejo Superior de Investigaciones Científicas, 1950. – 354 p.
125. Cervantes Saavedra M. Don Quijote de La Mancha / M. Cervantes Saavedra. – Madrid: Alfaguara, 2004. – 1251 p.
126. Chantal Pérez Hernández M. Explotación de los corpórea textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento // *Estudios de Lingüística del Español*. – Universidad de Málaga. – 2002. – Vol. 18.: – [Электронный документ]. Режим доступа: <http://elies.rediris.es/elies18/>
127. Corbut D. Machine Learning Translation and the Google Translate Algorithm, 2017. – [Электронный документ]. Режим доступа: <https://blog.statsbot.co/machine-learning-translation-96f0ed8f19e4>
128. Cowie A.P. English dictionary for foreign learners: a history / A.P. Cowie. Oxford: Clarendon Press, 1999. – 235 p.
129. Dancette J. L'intégration des relations sémantiques dans les dictionnaires spécialisés multilingues : du corpus ciblé à l'organisation des connaissances. – 2011 – [Электронный документ]. Режим доступа: <http://id.erudit.org/iderudit/1006177ar>
130. Dobrić N. Word Sense Disambiguation Using ID Tags – Identifying Meaning in Polysemous Words in English // *The 29th International Conference on Lexis and Grammar/LGC*, –2010. – 97-105.

131. Ebeling J. Contrastive linguistics, translation and parallel corpora // *Meta: Revue des Traducteurs*. – 1998. – Vol. 43, numéro 4.: – [Электронный документ]. Режим доступа: <https://www.erudit.org/fr/revues/meta/1998-v43-n4-meta169/002692ar.pdf>
132. Esquerre R. Los diccionarios de frecuencia en español https://cvc.cervantes.es/ensenanza/biblioteca_ele/aepe/pdf/boletin_10_06_74/boletin_10_06_74_03.pdf
133. Fokin S.B. TTR changes in different directions of translation // *Translation Journal*. – 2013. – Vol. 7. – № 1. – [Электронный документ]. Режим доступа: <http://translationjournal.net/journal/63ttr.htm>
134. Fokin S. Palabras prefijadas en ucraniano como recurso de la lengua de traducción: aspecto macroestilístico / S.B. Fokin // *Mundo Esloveno*. – Granada: Universidad de Granada, 2014. – nº 13. – P. 193-204.
135. Fokin S.B. Construcciones causativas "hacer(se)+Inf.", "dejar(se)+Inf." como recurso del texto de traducción: micro- y macroparámetros" // Wilk-Racińska, J., Szyndler, A., Tatoj *Relecturas y nuevos horizontes de los estudios hispánicos*. –Vol. 4º: Lingüística y Didáctica de la Lengua Española. – Katowice: Wydawnictwo Uniwersytetu Śląskiego, 2016. – P. 51-60.
136. 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.
137. General Ontology for Linguistic Description. – [Электронный ресурс]. Режим доступа: <http://www.linguistics-ontology.org/>
138. Gerasimov A. Review of Translation Quality Assurance Software – [Электронный ресурс]. Режим доступа: <http://www.translationdirectory.com/articles/article1253.htm>
139. Goddard C. Words and meaning / C. Goddard, A. Wierzbicka. – New-York: Oxford University Press, 2014. – 314 p.
140. Corpas Pastor G. Localización de recursos y compilación de corpus via Internet: aplicación para la didáctica de la traducción médica especializada // *Manual de documentación y terminología para la traducción especializada* / ed. C. Gonzalo García; V. García Yebra. – Madrid: Arco/Libros; 2004. – P. 223-258.
141. Goffin R. L'eurolecte: oui, jargon communautaire: non // *Meta: journal des traducteurs / Meta: Revue des Traducteurs*. – Vol. 39, №4. – 1994. – P. 636-642.
142. Gupta R. Improving Translation memory Matching and Retrieval Using Paraphrases/ R. Gupta, C. Orasan, M. Zampieri, M. Vela, Genabith J. van, R. Mitkov. // *Machine Translation*, November 2016. – P. 1-22.

143. Hai Xu. A Critique of the Controlled Defining Vocabulary in Longman Dictionary of Contemporary English, 2012. – Lexikos 22(1), p. 367-381 – [Электронный ресурс]. Режим доступа: <http://lexikos.journals.ac.za/pub/article/view/1013/530>
144. Humboldt-Universität zu Berlin. – [Электронный документ]. Режим доступа: https://www.linguistik.hu-berlin.de/en/institut-en/professuren-en/korpuslinguistik/links-en/korpora_links
145. Hutchis J. Compendium of Translation Software; 2010. – [Электронный документ]. Режим доступа: <http://www.hutchinsweb.me.uk/Compendium.htm>
146. Hassani G.A Corpus-Based Evaluation Approach to Translation Improvement // Meta. Revue des Traducteurs. Volume56, Issue2, Juin, 2011, p. 351-373.
147. ISO-1087-1:2000. Terminology work – [Электронный документ]. Режим доступа: Vocabulary https://edisciplinas.usp.br/pluginfile.php/312608/mod_resource/content/1/ISO_1087-1_2000_PDF_version_%28en_fr%29_CPDF.pdf
148. ISO-12620 Computer Applications in Terminology – Data categories <http://semanticweb.kaist.ac.kr/>
149. ISO-24610-1 Language resource management – Feature structure http://www.iso.org/iso/catalogue_detail.htm?csnumber=37324
150. Renales T. Kirchner J.V. Diccionario multilingüe de especies marinas para el mundo hispano. Reseña // Punto y coma. – n 11. – 1992. – [Электронный документ]. Режим доступа: <http://ec.europa.eu/translation/bulletins/puntoycoma/11/pyc119.htm>
151. Kiryakov A., Popov B., Terziev I., Manov D., Ognyanoff D. Semantic annotation, indexing, and retrieval // Journal of Web Semantics, 2004. – № 2(1). – P. 49-79.
152. Kutuzov A. Change of word types to word tokens ratio in the course of translation (based on Russian translations of k. Vonnegut's novels / A. Kutuzov [Электронный документ]. – 2010. – Режим доступа: <http://arxiv.org/ftp/arxiv/papers/1003/1003.0337.pdf>
153. Lexicografía // García de Quesada M. Estructura definicional terminográfica en el subdominio de la oncología clínica. [Электронный документ]. Режим доступа: <http://elies.rediris.es/elies14/cap141.htm>
154. Les corpus et la recherche en terminologie et en traductologie / M. Campenhoudt, R. Temmerman (dir) // Meta: Revue des Traducteurs. – Montreal, 2011. – Vol. 56. – n° 2. – 223-464 p.
155. Lewandowska-Tomaszczyk B. Explicit and tacit. An interplay of the quantitative and qualitative approaches to translation /

B. Lewandowska-Tymoszczyk / University of Łódź // Quantitative Methods in Corpus-Based Translation Studies: a practical guide / ed. By Michel P. Oakes and Mendji. – Amsterdam: John Benjamins Publishing, 2012. – P. 1-34.

156. Lozanova Bogomilova E. La diversidad lingüística y cultural. Experiencia en la elaboración del banco de datos terminológico multilingüe BT-FRASJURE / Bogomilova Lozanova, Elena; Quiroz, Gabriel; Giraldo, John Jairo; de Andrade, Ana Lúcia // La terminología en el siglo XXI: contribución a la cultura de la paz, la diversidad y la sostenibilidad: actas del IX Simposio Iberoamericano de Terminología RITMER04, 2006. –P. 675-686.

157. Luque Durán J.D., Pamies Bertrán A. (eds) Léxico y fraseología. – Granada: Mateo Ediciones, 1998. – 225 p.

158. Maia B. Word Order and the First Person Singular in Portuguese and English // Meta. Revue des Traducteurs. Montreal, 1998. – 43(4). – P. 589-601.

159. Malmkjaer K. Love thy Neighbour: Will Parallel Corpora Endear Linguists to Translators? // Meta, v. 43, nº. 4, pp. 534-541.

160. Manning Ch.D. Introduction to Information Retrieval /P. Raghavan, H. Schütz. – Cambridge: Cambridge University Press. – [Электронный документ]. Режим доступа: <https://nlp.stanford.edu/IR-book/>

161. Martí Antonín M., Fernández Monraveta A. & Vázquez García G. Lexicografía computacional y semántica – Barcelona: Ediciones de la Universidad de Barcelona, 2003. – 163 p.

162. Medina Guerra A. Lexicografía española. – Barcelona: Ariel S.A., 2003 – 427 p.

163. Melby A., Budin G. & Wright S.-E. Terminology Interchange Format (TIF) // Sendebarr: Revista de la Facultad de Traducción e Interpretación de Granada. – Granada, 1993. – nº4 – P. 69-110.

164. Moral R., Fábregas I. El análisis del léxico en los diccionarios – [Электронный документ]. Режим доступа: [vhttps://es.slideshare.net/RdelMoral/anlisis-del-lxico-en-los-diccionarios-delmoral-fbregas](https://es.slideshare.net/RdelMoral/anlisis-del-lxico-en-los-diccionarios-delmoral-fbregas)

165. Mudraya O. Automatic Extraction of Translation Equivalents of Phrasal and Light Verbs in English and Russian / O. Mudraya, S. Piao, P. Rayson, S. Sharoff., B. Babych, L. Lofberg // Phraseology: an interdisciplinary perspective. – Amsterdam: John Benjamins Publishing, 2008. – P. 293-309.

166. Navigli R., Velardi P. Structural Semantic Interconnections: a Knowledge-Based Approach to Word Sense Disambiguation // IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). – 2005. n° 27(7) – P. 1075-1086.
167. Nida E., Taber R. The theory and practice of translation. – Leiden.: E.J.Brill, for United Bible Societies, – 1982. – 218 p.
168. Ortega y Gasset J. Sobre el fascismo // Obras completas. Tomo III. – Madrid: Alianza editorial, Revista de occidente, 1983. – P. 489-497.
169. Øveras L. In Search of the Third Code: an Investigation of Norms in Literary Translation // Meta. – 1998. – Vol. 43. – №. 4. – P. 571-588.
170. Pamies A. Fraseología y metáfora: aspectos tipológicos y cognitivos / A. Pamies Bertrán, E.-M. Iñesta Mena. – Granada: Granada Lingvística, 2002. – 288 p.
171. Perebeynoss V., Khidekel S. Frequency of Language Units as a Reflection of their Systemic and Functional Properties // Journal of Quantitative Linguistics. – Vol. 11 – №1-2, April/August 2004. – P. 3-25.
172. Punto y Coma: Boletín de los traductores españoles de las instituciones de la Unión europea. – [Електронний документ]. Режим доступу: http://ec.europa.eu/translation/spanish/magazine/es_magazine_es.htm
173. Ritter A., Clark S., Etzioni M. and O. Named entity recognition. An experimental study // Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, EMNLP 2011, 27-31 July 2011, John McIntyre Conference Centre, Edinburgh, UK L. – [Електронний документ]. Режим доступу: <https://homes.cs.washington.edu/~mausam/papers/emnlp11.pdf>
174. Savory Th. The Art of Translation. – London: Jonathan Cape, 1957. – 159 p.
175. Sekine S. Sekine's Extended Named Entity Hierarchy. – [Електронний документ]. Режим доступу: – <https://nlp.cs.nyu.edu/ene/>
176. Smouchtchynska I. Lexicologie française. – К.: Видавничий дім Дмитра Бураго, 2015. – 600 p.
177. Stock P.F. Polysemy // Practical Lexicography. A Reader. – Oxford University Press. 2008. – P. 153-160.
178. Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication. – [Електронний документ]. <https://benjamins.com/catalog/term>

179. Thierry Fontenelle. Using a Bilingual Dictionary to Create Semantic Networks // Practical Lexicography. A Reader. – Oxford University Press. 2008. – P. 169-189.
180. Traduction. Terminologie. Rédaction. – [Електронний документ]. Режим доступу: <https://www.erudit.org/en/journals/ttr/>
181. Troyano J. A. ITALICA. Wordnet. – [Електронний документ]. Режим доступу: <http://www.lsi.us.es/~troyano/documentos/wordnet.pdf>
182. Vázquez Ayora G. Introducción a la traductología / G. Vázquez Ayora. – Georgetown : Georgetown University School of Languages and Linguistics, 1977. – 488 p.
183. Verba G.G, Guetman Z.O. Manual de traductología (español-ucraniano) = Підручник з перекладознавства (іспанська та українська мови) / Galina Verba, Zoia Guetman. – Вінниця: Нова Книга, 2013. – 303 с.
184. Villers E.M. Lexicographie – Montréal: Presses de l'Université de Montréal, 2014. – 70 p.
185. Zhu Q. Automatic extraction of translations from web-based bilingual materials / Q. Zhu, D. Inkpen, A. Asudeh // Machine Translation 21:3, 2007. – P. 139-163.
186. Wierzbicka A. Semantic Primes and Universals. – Oxford-New-York: Oxford University Press, 2004. – 512 p.

Лексикографічні, термінологічні ресурси і бази даних

187. Автоматическая обработка текста: ресурсы Российского Государственного Гуманитарного Университета (онлайн демо) <http://aot.ru/demo/bigrams.html>
188. Антоненко-Давидович Б.Д. Як ми говоримо. – 4-е вид. перероб. і доп. – К.: “Українська книга”, 1997. – 336 с.
189. Архів української періодики онлайн. – [Електронний документ]. Режим доступу: <http://uk.glosbe.com/https://libraria.ua/>
190. Багатомовний перекладний ідеографічний словник. – [Електронний документ]. Режим доступу: <http://translation.sensagent.com>
191. Багатомовний юридичний словник-довідник / І.О. Голубовська, В.М. Шовковий, О.М. Лефтерова та ін. – К.: Видавничо-поліграфічний центр “Київський університет”, 2012. – 543 с.
192. Баранов О.С. Идеографический словарь русского языка. – М.: ЭТС, 1995 – 820 с.

193. Бирик С.П. Словник епітетів української мови / С.П. Бирик, С.Я. Єрмоленко, Л.О. Пустовіт; [за ред. С.Я. Єрмоленко]. – К.: Довіра, 1998. – 430 с.
194. Всесвітній словник української мови. – [Електронний документ]. Режим доступу: <https://uk.worldwidedictionary.org/>
195. Гетьман З.О., Орлова І.С., Паламарчук О.Л., Стрельчук Г.П., Шевченко М.В. Українсько-іспанський словник лінгвістичної термінології. – К.: Освіта України, 2011. – 356 с.
196. Глуховцева К.Д. Лінгвістичний атлас лексики народного побуту українських східнослов'янських говірок. – Луганська: Альма-матер, 2003. – 182 с.
197. Головащук С.І. Складні випадки наголошення. Словник-довідник / С.І. Головащук. – К.: Либідь, 1995. – 192 с.
198. Городенська К.Г. Граматичний словник української мови. Сполучники / К.Г. Городенська. – К.; Херсон: Вид-во ХДУ, 2007. – 340 с.
199. ГРАК, Генеральний регіонально анотований корпус української мови / М. Шведова, Р. Фон Вальденфельс, С. Яригін, М. Крук, А. Рисін, М. Возняк. – Київ, Осло, Йена, 2017-2018. – [Електронний ресурс]. – Режим доступу: uacorporus.org
200. Граматичний словник української літературної мови : словозміна / В.І. Критська [та ін.]; відп. ред. Н.Ф. Клименко ; Нац. акад. наук України, Ін-т укр. мови, Ін-т мовознав. ім. О.О. Потебні. – К.: Вид. дім Дмитра Бураго, 2011. – 757 с.
201. Гринчишин Д., Капелюшний А., Сербенська О., Терлак З. Словник-довідник з культури української мови. – Львів: Фенікс, 1996. – 368 с.
202. Грицелак В.П. Довідник транслітерації іноземних власних назв українською мовою: для журналістів, географів та картографів, туристів тощо. – Львів: Карти і Атласи, 2009. – 56 с.
203. Довідник ІНКОТЕРМС українською. Особливості використання. <http://met-online.kiev.ua/menu/dovidka/inko/>
204. Еспансько-український словник / Уклад. О. Буній. Буенос-Айрес – Рим: Український Католицький університет, 1993. – 993 с.
205. Испанско-русский словарь – [Електронний документ]. Режим доступу: <https://www.diccionario.ru>
206. Інверсійний словник української мови / Уклад. С.П. Бевзенко, О.І. Бондар та ін. – К.: Наукова думка, 1985.
207. Іспансько-український словник / М. Жердинівська. – К.: Основи, 1996. – 226 с.

208. Іспансько-український, українсько-іспанський словник / за ред. В. Бусела. – К.-Ірпінь: Перун, 2013. – 1120 с.
209. Іспансько-український словник економічних термінів / Уклад. Е.В.Шевкун, О.О.Курченко. – К.: Корнійчук, 2001. – 341 с.
210. Іспансько-українсько-російський та українсько-російсько-іспанський словник водогосподарських термінів / Уклад. С.М. Кропивко, І.М. Кропивко. – Рівне: НУВГП, 2008. – 244 с.
211. Клименко Н.Ф., Карпіловська Є.А., Карпіловський В.С., Недозим Т. І. Словник афіксальних морфем української мови / Н.Ф. Клименко, Є.А. Карпіловська, В.С. Карпіловський, Т.І. Недозим. – К.: Ін-т мовознавства ім. О. Потебні, 1998. – 435 с.
212. Колекція корпусів англійської мови. – [Електронний документ]. Режим доступу: <https://www.english-corpora.org/>
213. Коломієць В. Корпус анотацій наукових статей із комп'ютерної лінгвістики / В. Коломієць, В. Орел // Комп'ютерна лінгвістика: сучасне і майбутнє. Матеріали Міжнародної науково-практичної конференції. – К. : КНЛУ, 2012. – С. 32-34.
214. Коломієць Н.Ф., Регушевський Е.Ф. Словник фразеологічних синонімів. – К.: Радянська школа, 1988. – 200 с.
215. Корпус текстів української мови. – [Електронний документ]. Режим доступу: <http://www.mova.info/corpus.aspx?l1=209>
216. Корпус української мови. – [Електронний документ]. Режим доступу. – [Електронний документ]. Режим доступу: <http://korpus.org.ua/>
217. Корпус русского литературного языка <http://www.narusco.ru/>
218. Кошарний К.О. Іспансько-український військовий термінологічний словник. – К.: ПП Кльоц, 2007. – 56 с.
219. Міжнародна класифікація хвороб – [Електронний документ]. Режим доступу: <http://mkh10.com.ua/>
220. Национальный корпус русского языка.– [Електронний документ]. – <http://www.ruscorpora.ru/>
221. Обернений частотний словник сучасної художньої прози / Уклад. Т. О. Грязнухіна, Н. П. Дарчук, Є. А. Карпіловська та ін. – К.: Спалах, 1998. – 960 с.
222. Открытый корпус русского языка. – [Електронний документ]. Режим доступу: opencorpora.org/
223. Пошукач слів української мови. – [Електронний документ]. Режим доступу: <http://uk.words-finder.com/>

224. Сахно І.П. Словник сполучуваності слів української мови: Найуживаніша лексика / Уклад. І.П.Сахно, М.М.Сахно. – Дніпропетровськ: Вид-во Дніпропетр. ун-ту, 1999. – 544 с.
225. Семантичний словник української мови. – [Електронний документ]. Режим доступу: <http://www.mova.info/semvoc.aspx?l1=193>
226. Словарь української мови / упор. Б. Грінченко: Т. 3 – К. : Вид-во Академії наук Української РСР, 1958. – 479 с.
227. Словарь Мультитран. – [Електронний документ]. Режим доступу: <http://www.multitrans.ru/>
228. Словник українсько-іспанський, іспансько-український / Уклад. В. Ф. Сахно, С. А. Коваль; заг. ред. В. Т. Бусла. – Київ-Ірпінь: ВТФ "Перун", 1997. – 534 с.
229. Словник фразеологічних синонімів. – [Електронний документ]. Режим доступу: <http://www.rozum.org.ua/index.php?a=term&d=24&t=546>
230. Словники України. – [Електронний документ]. Режим доступу: <http://lcorp.ulif.org.ua/dictua/>
231. Словник української мови в 11-ти томах. – К.: Наукова думка, 1970-1980.
232. Словник української мови (СУМ). Академічний тлумачний словник. – [Електронний документ]. Режим доступу: <http://sum.in.ua/>
233. Смолій В.А., Федорченко В.К., Цибух В.І. Енциклопедичний словник-довідник з туризму / Передмова В.М. Литвина. – К.:Видавничий Дім "Слово", 2006. – 372 с.
234. Ставицька Л. Короткий словник жаргонної лексики української мови. – К.: Критика, 2003. – 336 с.
235. Тарновська О. Найбільш уживані іспанські прислів'я та приказки з українськими та російськими відповідниками. – К.: Логос, 2004. – 119 с.
236. Тезаурус з комп'ютерної лексикографії / упорядн. О.Б. Сірук. – [Електронний документ]. Режим доступу: <http://www.mova.info/Page3.aspx?l1=188&vocid=1>
237. Тлумачний словник української мови. Томи 1-8 <http://services.ulif.org.ua/expl/Entry/index?wordid=1&page=0>
238. Українсько-іспанський словник / Уклад. М. Жердинівська. – К.: Основи, 1993. – 239 с.
239. Український національний лінгвістичний корпус. – [Електронний документ]. Режим доступу: http://lcorp.ulif.org.ua/virt_unlc/

240. Український зворотний словник / Василь Ніньовський – Мюнхен-Едмонтон: Український технічно-господарський інститут, 1969. – 492 с.
241. Фокін С.Б. Хибні друзі перекладача: іспанська, українська мови. Словник-довідник. – К.: Капрі, 2015. – 112 с.
242. Частотний словник сучасної української художньої прози / За ред. Перебийніс В.С. – К.: Наук. думка, 1981. – Т. 1. – 863 с.; Т. 2. – 855 с.
243. Частотні словники лінгвістичного порталу Mova.info. – [Електронний документ]. Режим доступу: <http://www.mova.info/corpus2.aspx>
244. Шевкун Е.В. Іспансько-російсько-український словник граматичних лексем. – К.: Освіта України, 2014. – 284 с.
245. Юридична енциклопедія: у шести томах / за ред. Ю.С. Шемшученка (голова редкол.) та ін. – [Електронний ресурс]. – Режим доступу: <http://leksika.com.ua/legal/>
246. An open source for language analyzer FreeLing 4.0 // TALP Research Center. – [Електронний ресурс]. – Режим доступу: <http://nlp.lsi.upc.edu/freeling/demo/demo.php>
247. Banche Dati, Corpora e Archivi Testuali // Accademia della Crusca. – [Електронний документ]. Режим доступу: <http://www.accademiadellacrusca.it/it/link-utili/banche-dati-dellitaliano-scritto-parlato>
248. British Parliament “Hansard” Corpus – [Електронний ресурс]. – Режим доступу: <https://www.hansard-corpus.org/semTags1.asp>
249. Buscador de ideas relacionadas y afines. – [Електронний документ]. Режим доступу: <http://www.ideasafines.com.ar/buscador-ideas-relacionadas.php>
250. Buscador Urgente de Dudas de FUNDEU. – [Електронний документ]. Режим доступу: <http://www.fundeu.es/>
251. British National Corpus – [Електронний документ]. Режим доступу: <https://www.english-corpora.org/bnc/>
252. CAES, Corpus de Aprendices de Español. – [Електронний документ]. Режим доступу: <http://galvan.usc.es/caes/search>
253. Casares. J. Diccionario ideológico de la lengua española. – Barcelona: Editorial Gustavo Gili, S.A., 1994 – 887 p.
254. CDH, Corpus del Diccionario Histórico del Español. – [Електронний документ]. Режим доступу: <http://www.rae.es/recursos/banco-de-datos/cdh>

255. Centre National des Ressources Textuelle et Lexicales – [Электронный документ]. Режим доступа: <http://www.cnrtl.fr/corpus/>
256. Clave, Diccionario de Uso del Español Actual. – [Электронный документ]. – Режим доступа: <http://clave.smdiccionarios.com/app.php>
257. COCA, Corpus of Contemporary American English. – [Электронный документ]. Режим доступа: <https://corpus.byu.edu/coca/>
258. Código Internacional de Enfermedades – [Электронный документ]. – Режим доступа: <http://cie10.tiddlyspot.com/>
259. CODIS, Corpus Dinamico dell'Italiano Scritto. – [Электронный документ]. Режим доступа: <http://corpora.dslo.unibo.it/CODIS/>
260. Contextual Search Engine Linguee. – [Электронный документ]. Режим доступа: <http://www.linguee.com>
261. Reverso Context. – [Электронный документ]. Режим доступа: context.reverso.net/
262. CORDE, Corpus Diacrónico del Español. – [Электронный документ]. Режим доступа: <http://www.rae.es/recursos/banco-de-datos/corde>
263. CORIS, Corpus di Riferimento dell'Italiano Scritto. – [Электронный документ]. Режим доступа: http://corpora.dslo.unibo.it/coris_ita.html
264. CORPES XXI, Corpus del Español del Siglo XXI. – [Электронный документ]. Режим доступа: <http://www.rae.es/recursos/banco-de-datos/corpes-xxi>
265. Corpora Collection of Leipzig University. – [Электронный документ]. Режим доступа: <http://corpora.uni-leipzig.de>
266. Corpus de Referencia del Español Actual. Versión Anotada – [Электронный документ]. Режим доступа: <http://web.frl.es/CREA/view/inicioExterno.view>
267. CREA, Corpus de Referencia del Español Actual (CREA). – [Электронный документ]. Режим доступа: <http://www.rae.es/recursos/banco-de-datos/crea>
268. Dámaso Martínez C. Roberto Arlt. Apunte biográfico. – [Электронный документ]. Режим доступа: http://www.cervantesvirtual.com/portales/roberto_arlt/autor_apunte/
269. David B. Givens. The Nonverbal Dictionary of Gestures, Signs & Body Language Cues From Adam's – Apple – Jump to Zygomatic

Smile. – Spokane, Washington: Center for Nonverbal Studies Press. – 2008. – 781 p.

270. Delisle J. et al. Terminologie de la traduction. Translation terminology. Terminología de la traducción. Terminologie der Übersetzung. – Amsterdam/Philadelphia, John Benjamins publishing company, 1999. – 433 p.

271. DeReKo – [Электронный документ]. Режим доступа: <http://www.ids-mannheim.de/cosmas2/>

272. Deutscher Wortschatz Project. – [Электронный документ]. Режим доступа: <http://wortschatz.uni-leipzig.de/>

273. Diccionario de Colocaciones del Español. – [Электронный документ]. Режим доступа: <http://www.dicesp.com/paginas>

274. Diccionario de gestos españoles. – [Электронный документ]. Режим доступа: <http://lecture.ecc.u-tokyo.ac.jp/~cueda/gakusyu/gestos/index.html>

275. Dictionnaire de l'Académie Française. [Электронный документ]. – Режим доступа: <http://artfl.atilf.fr/dictionnaires/ACADEMIE/index.htm>

276. Dictionnaire Larousse. [Электронный документ]. – Режим доступа: <https://www.larousse.fr/dictionnaires/francais>

277. DIRAE, Diccionario Inverso de la Real Academia Española. – [Электронный документ]. Режим доступа: <http://dirae.es/>

278. DPD, Diccionario Panhispánico de Dudas. – [Электронный документ]. Режим доступа: <http://www.rae.es/recursos/diccionarios/dpd>

279. DRAE [Diccionario De La Real Academia Española] – [Электронный документ]. Режим доступа: dle.rae.es/

280. DWDS Core Corpus <http://www.dwds.de/resource/kerncorpus/>

281. El Sistema de Clasificación Decimal Universal. – [Электронный документ]. Режим доступа: <http://www.taranco.eu/cdu/cdu-esquema.htm>

282. EuroTermBank – [Электронный документ]. Режим доступа: <http://www.eurotermbank.com/>

283. Flexionador y lematizador de palabras en español // Grupo de Estructuras de Datos y de Lingüística Computacional <http://www.gedlc.ulpgc.es/investigacion/scogeme02/lematiza.htm>

284. Fish Base. – [Электронный документ]. Режим доступа: <http://www.fishbase.org>

285. FranText. – [Электронный документ]. Режим доступа: <https://www.frantext.fr/>

286. Glosbe: The multilingual online dictionary. – [Электронный документ]. Режим доступа: <http://uk.glosbe.com/>
287. Grande Dizionario Hoepli Italiano. – [Электронный документ]. Режим доступа: http://www.grandidizionari.it/Dizionario_Italiano.aspx
288. Google n-gram Viewer – [Электронный документ]. Режим доступа: <https://books.google.com/ngrams#>
289. Hamburg Dependency Treebank. – [Электронный документ]. Режим доступа: <http://hdl.handle.net/11022/0000-0000-7FC7-2>
290. IdeoPhrase. Multilingual Dictionary of Phraseological Synonyms. – [Электронный документ]. Режим доступа: <http://postup.zzz.com.ua/IdeoPhrase.html>
291. InterActive Terminology for Europe (IATE). – [Электронный документ]. Режим доступа: <http://iate.europa.eu/SearchByQuery.do>
292. Korpus Języka Polskiego. – [Электронный документ]. Режим доступа: <https://sjp.pwn.pl/korpus>
293. Leeds Collection of Internet Corpora. – [Электронный документ]. Режим доступа: <http://corpus.leeds.ac.uk/internet.html>
294. LIMAS-Korpus – [Электронный документ]. Режим доступа: <http://www.korpora.org/Limas/>
295. LingvoSoft Online Explanatory Dictionary. – [Электронный документ]. Режим доступа: <http://www.lingvozone.com/>
296. Listado de frecuencias del Corpus de Referencia del Español Actual. – [Электронный документ]. Режим доступа: <http://corpus.rae.es/lfrecuencias.html>
297. Merriam-Webster Thesaurus. – [Электронный документ]. Режим доступа: <https://www.merriam-webster.com/thesaurus>
298. Moral R. del. Diccionario ideológico: atlas léxico de la lengua española. – М.: Herder, 2009. – 666 p.
299. Narodowy Korpus Języka. – [Электронный документ]. Режим доступа: [Polskiego http://nkjp.pl/](http://nkjp.pl/)
300. One Look Reverse Dictionary. – [Электронный документ]. Режим доступа: <https://www.onelook.com/reverse-dictionary.shtml>
301. Online Translator Pragma6. – [Электронный документ]. Режим доступа: <http://online.translate.ua/>
302. Oxford English Dictionary. – [Электронный документ]. Режим доступа: <http://www.oed.com/>
303. PAISÀ (Piattaforma per l'Apprendimento dell'Italiano su Corpora Annotati) – [Электронный документ]. Режим доступа: <http://www.corpusitaliano.it/it/contents/funding.html>

304. Pamies Bertrán A., Kálustova O. Guía básica de fonética y fonología con equivalencias en cinco idiomas. – Granada: Método, 2000. – 172 p.

305. Pamies Bertrán A., Iñesta Mena E.M., Bálmacz M., Kálustova O.M. Multilingual Electronic Phraseological Dictionary "AUTOFRAS" // Tempus Language Toolbox. – Granada, 1998 (CD-version).

306. ParsEqv, комп'ютерна програма вибірки еквівалентних фрагментів з паралельних текстів // Свідectво про реєстрацію авторського права на твір №76866 від 15.02.2018 Міністерства економічного розвитку і торгівлі України.

307. Polsko-Ukraiński Korpus Równoległy. – [Електронний документ]. Режим доступу: <http://www.domeczek.pl/~polukr/index.php?option=search>

308. Refranero Multilingüe del Centro Virtual Cervantes. – [Електронний документ]. Режим доступу: <https://cvc.cervantes.es/lengua/refranero/Busqueda.aspx>

309. Reverse Dictionary – [Електронний документ]. Режим доступу: <http://reversedictionary.org/>

310. Routledge encyclopedia of translation studies / Ed. by Mona Baker; Ass. by Kirsten Malmkjaer. – London; New York: Routledge, 2001. – XVIII, 654 p.

311. Seco M. Diccionario de dudas y dificultades de la lengua española. – Madrid: Espasa-Calpe, 1998. – 595 p.

312. Słownik języka polskiego. – [Електронний документ]. Режим доступу: [PWN sjp.pwn.pl](http://pwn.sjp.pwn.pl)

313. Standardization of Terminology // International Information Centre for Terminology (INFOTERM). – [Електронний документ]. Режим доступу: http://www.infoterm.info/standardization/standardization_of_terminologies.php

314. Termdat, Banco de Datos de la Administración Federal de Suiza [Електронний документ]. Режим доступу: <https://www.termdat.bk.admin.ch>

315. Termdat. – [Електронний документ]. Режим доступу: <https://www.termdat.bk.admin.ch>

316. Tesauro Multilingüe de la Unión Europea. – [Електронний документ]. Режим доступу: <http://eurovoc.europa.eu/>

317. TextSTAT, Simple Text Analysis Tool. – [Електронний документ]. Режим доступу: <http://neon.niederlandistik.fu-berlin.de/en/textstat/>

318. The iWeb Corpus. – [Электронный документ]. Режим доступа: <https://corpus.byu.edu/iweb/>
319. THERSAURUS UNESCO. – [Электронный документ]. Режим доступа: <http://vocabularies.unesco.org/browser/thesaurus/es/search>
320. Turton N.D., Heaton J.B. Longman Dictionary of Common Errors. – Harlow: Langensch.-Hachette, M (January 1, 2000). – 375 p.
321. UNITERM, The United Nations Terminology Database. – [Электронный документ]. Режим доступа: <http://unterm.un.org/>
322. Visual Thesaurus by Thinkmap Inc. – [Электронный документ]. Режим доступа: <http://www.visualthesaurus.com/app/view>
323. Wierzbicka A. Semantic and Lexical Universals. – Amsterdam/Philadelphia: John Bejamisn Publishing C., 1994. – 512 p.
324. Wolfram Alpha: Computational Knowledge Engine. – [Электронный документ]. Режим доступа: www.wolframalpha.com
325. Wordnet. A lexical database for English. – [Электронный документ]. Режим доступа: <https://wordnet.princeton.edu>

ПРЕДМЕТНИЙ ПОКАЖЧИК

4-грами 192

5-грами 192

абсолютна частота 119, 164, 169, 192

абстрактні поняття 17,66

автоінтертекстуальність 151

автоматизація вибірки 33, 95, 107, 124, 133, 153, 164, 165, 187, 203, 208

автоматизована вибірка повторів 165

автоматизований пошук 15-20, 22, 26, 34, 38, 45, 49, 54, 55, 58-62, 64, 80, 85, 99, 100, 115, 116, 120-135, 138, 139, 164

автоматична розмітка 174, 195, 202

автоматична психолінгвістична експертиза тексту 154

автоматичне визначення параметрів тексту 174

автоматичне індексування тексту 187

автоматичне реферування тексту 187

автоматичне розпізнання помилок перекладу 185

автоматичний аналіз смислу тексту 42, 82, 95, 96, 187, 193, 192

автоматичний граматичний аналіз 187

автоматичний розрахунок обсягу тексту 182

акціональні дієслова 17

алолексія 197

анаграма 34

анадиплозис 165

аналітичність текстів перекладу 146

анафора 117, 140, 154, 165

анафорична розмітка 117

анафоричний повтор 154, 165

англійська мова 24, 31, 33, 37, 39, 40, 46, 47, 50, 53, 54, 58, 77, 85, 87, 88, 89, 90, 92, 100, 116, 134, 164, 191, 206, 207, 209

анотований корпус 116, 124, 128, 130, 134, 136, 137, 138, 139, 201

антоніми 46, 53, 64, 83, 105

арабська мова 134

артикуль 124, 151, 163

архісема 64

архітектоніка тексту 158

атлас лексики 49

багатокатегорійна класифікація 118, 204

Див. також: **принцип перетинання**

багатокомпонентні терміни 87, 90

багатомовний тезаурус 99, 102

багатомовні словники 16, 19, 21, 57, 60, 79, 89, 99, 196

багатомовні колекції корпусів 131, 133

багатомовні корпуси 115, 132, 133, 165

багатослівний запит пошуку 122

байдужі символи 121, 122, 142, 181

баскська мова 209

безеквівалентна лексика 26

біграми 23, 41, 127, 139, 192, 194, 195, 204, 208
білінгвізм 85
білоруська мова 128
біологічний вид 65, 77, 79, 103
болгарська мова 48, 62, 128
браунівський корпус 131, 134, 188
вживання великої/малої літери 203
вибірка з автоматизованої бази 95, 107, 132, 141
вибірка термінів на основі аналізу структури 107, 132
вибіркі даних з текстів
 Див.: **добування даних**
вибіркова середня 169
видо-родові відношення 53, 67, 77
 Див. також: **гіпо-гіперонімічні відношення**
видова парадигма дієслова 207
визначення типу тексту 174
визначник 12, 20, 24, 31, 77, 90
використання корпусів у практиці перекладу 37, 38, 85-94, 115, 116, 117, 120-140
вихідна мова 14
вихідний текст 14
відбивка (графічний прийом) 158
відмінкові форми 35, 37, 126
відносна похибка 169, 171
відносна частота 119, 169
відносні прикметники 107
відносні семи 81
відношення часткового-цілого 52, 65, 68, 77, 105
 Див. також: **голо-меронімічні відношення**
відповідність мовній нормі цільового тексту 145
відповідність традиціям цільової культури 148
відхилення від норм мови 140, 146, 163
відхилення від норм орфографії 163
візуальний аналіз 93
візуальний тезаурус 50
вказівні займенники 163
власноруч створений корпус 141, 159, 160, 166, 167
внутрішньочастиномовні омоніми 194
вокабула 12
встановлення еквівалентних відношень між термінами 85-94
вторинна номінація 82
втрати і компенсації в перекладі 150, 163, 175
галісійська мова 209
гапакс легоменон 155
гасло 12
 Див. також: **лема**
геологія (терміни) 76
герундіальні перифрази 148, 164
герундій 127, 148, 164, 165, 190
гіперонім 64, 65, 68, 81, 82, 87, 89, 90, 96, 100, 105, 106
гіперонімічний пошук 89

гіперпосилання 17, 19, 38
гіперструктура 16
гіпертекстуальність 17, 19, 25
гіпо-гепорнімічні відношення 64, 65, 68, 77, 82, 87, 105, 106
глобальна автозаміна 180
гласа 14, 15, 16, 17, 18, 26, 42, 54, 56, 57
гласарій 15, 16, 208
голо-меронімічні відношення 65, 89, 105, 106
граматична експлікація 147
граматична розмітка 8, 25, 99, 117
граматична сполучуваність 23
граматичний ономазіологічний словник 63
граматичний рід 97
граматичний словник 30, 63, 191
граматичні маркери 151
графічне оформлення тексту 186
графічне слово 11
графічні маркери в тексті 158
греколатинізми 89
грузинська мова 134
двозначність 77, 110, 187, 190, 192, 194, 196, 200, 201, 206
двомовний термінологічний тезаурус 65
двомовні паралельні тексти 25
двомовні перекладні електронні словники 21, 23, 25, 27, 30, 78, 85
двомовні перекладні термінологічні словники 78
двослівний запит пошуку 116
дейктичні зв'язки 66
декомпресія граматичних конструкцій 158
декриптори 19, 22, 56, 64, 83
денотативна еквівалентність 103
денотативне значення 24, 30, 31, 77, 81, 103, 140
deskрипція 139
деталізація значень лексеми 29, 39, 43, 44, 78
детермінанта 192, 194
дефініції 59, 64, 76, 80, 81, 82, 93, 95, 97, 100 198
дефініції за екстенсіоналом 81, 98
дефініції за інтенсіоналом 81, 98
дефініції термінів 76
дефініційний словник 64, 198, 199
дидактичне призначення корпусу 85-94, 140-142
дискурс 22, 36, 80, 116, 141, 145, 149, 150, 154, 158, 165, 167, 202, 204
дискурсивні маркери 150-153
дискурсивно орієнтований корпус 154
дистрибутивний аналіз 90, 91, 203
дистрибутивний аналіз з калькуванням 91
дистрибуція слова 90, 91, 193, 203
дисфемізми 162
диференційні семи 81
діалектизми 157
дієслівне керування 22, 122, 140

дієслово 11, 12, 17, 18, 22, 24, 26, 37, 41, 122, 124, 134, 140, 146, 151, 153, 164, 165, 185, 188, 190, 191, 195, 200, 207

добування даних з тексту 42, 82, 95, 96, 187, 193, 192

Див. також: **text mining**

довжина речення 165

домінанта тексту 150, 154, 156

допоміжні дієслова 146, 164, 165

допоміжні слова 146, 147, 165

дослідницькі паралельні корпуси 128-130

достовірність вибірки 119, 164, 168-172

дублети 106

евфонічність 156

еквівалентність 31, 39, 58, 77, 86, 90, 103, 105, 106, 163

екзистенційно-центрична синоптична схема 64

експансивна дії (семантика) 62

експлікація 24, 145, 146, 147, 151, 189

екстенціонал 30, 31, 81

екстралінгвістична розмітка 117, 166

електронні бази даних 66, 97, 100, 101, 103, 108, 109, 112

елементарні семи 197

еліпсис 140

енциклопедичні функції 95

епонім 31

євролект 76

жанрово-стилістичні доміанти 150, 154

завищення частоти явища в перекладі 148, 165

заголовне слово у словнику 12, 17, 191

Див. також: **лема**

займенник 117, 122, 164, 188, 195

займенниковий іменник 188

займенниковий прикметник 188

закінчення -s, -ed, -ing 191

закон Ципфа 146

заниження частоти вживання явища у перекладі 148, 165

запит пошуку 122

збільшення обсягу тексту в перекладі 146, 147

зв'язок лексичної та граматичної семантики 37

зведений словник 18, 27, 33, 39, 153

звуження значення лексеми 39

звукосимволізм 156

зірочка (байдужий символ) 121

зіставний словник 62

знак питання (байдужий символ) 121

зняття граматичної двозначності 190

зняття двозначності 187, 190, 191, 194, 200

зняття лексичної двозначності 191

зсув значення 39

ідентифікаційний код терміна 79, 100

ідеологія 46, 154

ідіолектна розмітка корпусу 8

ідіюстиль 14, 156, 157

Див. також: індивідуальний стиль перекладача, стиль ієрархічні відношення між лексемами 65, 67
іменник 17, 22, 24, 31, 37, 39, 40, 41, 43, 53, 72, 77, 84, 93, 108, 122, 124, 134, 140, 153, 159, 162, 188, 189, 190, 191, 192, 194, 200, 202, 207
іменикове керування 22, 122, 140
імітація індивідуального стилю перекладача 152
інверсійний словник 18, 26, 63
інверсійний двомовний словник 26
індексування тексту 187
індивідуальний стиль перекладача 14, 151-154
індикатори позамовної дійсності в тексті 150
інтегральний словник 27
Див. також: зведений словник
інтегрована форма лексеми 126
інтенціонал 30
інтерактивність словника 23, 25
інтервал довіри в лінгвостатистиці 169
інтердискурсивність 151
інтертекстуальність 151
інтерференція вихідної мови в перекладі 145, 147
інтроспективний аналіз 140
іспанська мова 18, 20, 26, 31, 34, 36, 40, 40, 47, 48, 54, 65, 77, 79, 87, 116, 122, 134, 135, 190, 209
італійська мова 31, 33, 40, 50, 87, 92, 100, 134, 137, 209
калька 87, 92
каталонська мова 209
категорійна класифікація лексики
Див.: класифікація лексики
квазіпаралельні тексти 87, 132
квазісиноніми 53, 77, 104
квазіфлексії 121
керування (синтаксис) 22, 122, 140
китайська мова 134
кількісні та якісні параметри корпусів 119, Додаток 1
класифікації лексики 49, 63, 66, 78, 201, 202, 203
клівлендський семантичний код 53
когнати 12, 89, 142
код достовірності термінологічної картки 100
код терміна 75, 79, 100, 199
кодування символів текстового файлу 179
конкретні іменники 93
колекції корпусів 133
колігація 23
колокація 21, 107, 140, 146, 157, 162
компенсації в перекладі 163
комп'ютеризований словник 18
комп'ютерна граматики 188
комп'ютерна перекладацька пам'ять 124, 205, 207
комп'ютерний лексикон 206
компаративні явища в контрастивістиці 168
компенсація (прийом перекладу) 150, 163, 165

композиція тексту 146, 154, 156
комунікація між людиною і комп'ютером 97, 204
конкорданс 22, 37, 87, 119, 122, 123, 133, 141, 155, 162
конкордансер 119, 122, 123, 141
конкорданси паралельних текстів 87
конотація 33, 58, 150, 151, 152, 163, 174
контекстуальні синоніми 82
контрастивні дослідження 106, 129, 142, 147-150
контроль якості перекладу 184, 185, 186
контроль версій текстового документа 182
концепти 30
координація (відношення) 106
корпус *ad hoc* 141
корпуси паралельних текстів 127-129
корпуси співвідносних текстів 92, 141, 147-149
критерії співвідносності текстів 141
кругова дефініція 82
культурно маркований термін 76
курсив 158
ланцюги семантичних відношень 189
латвійська мова 134
латинська мова 93
лексико-граматичний клас 174, 207
 Див. також: **артикль, дієслово, займенник, іменник, прикметник, частини мови**
лексико-семантична онтологія 201
лексико-семантична розмітка 25, 201
лексико-семантична сполучуваність 16, 22, 23, 27, 38, 41, 44, 46, 88, 92, 100, 115, 120, 122, 124, 126, 140, 156, 157, 159, 162, 165, 186, 189, 192, 193, 194, 195
лексико-семантичний варіант 16, 31, 39, 41, 42, 43, 53, 78, 190, 192, 196, 201
лексико-тематична група 66
лексична двозначність 191
лексична експлікація 147
лексична розмітка 25, 201
лексична семантика 196, 200
лексичне портретування слова 43
лексичний повтор 13, 151, 154, 156, 165
лексичний атлас 49
лема 12, 15, 16, 21, 26, 51, 54, 64, 99, 123, 190
лематизатор 13, 129, 160, 190, 191, 195
ліва частина словникової статті 14, 15
 Див. також: **лема**
лінгвістична розмітка 117, 128
лінгвістичне конструювання 18, 23, 25, 29, 155
лінгвостатистика 169
ЛСВ
 Див.: **лексико-семантичний варіант**
макроконтекст 163
макрорівень 163

макростилістика 151, 156
макроструктура словника 14, 18, 20, 21, 22, 56, 62
малапропізм 163
маркери 150, 154, 158, 163, 165
маркерів психічних станів 134, 154, 180
маска пошуку 124
маска слова 122
машинний переклад 99, 101, 196, 209
 Див. також: **програми машинного перекладу**
меро-голонімічні відношення 64, 68, 75, 82, 89, 105, 106
 Див. також: **голо-меронімічні відношення**
метадані 117
металінгвістичні дані 78, 201
метаслова 60, 99, 192, 196, 205, 206
метамовний опис 99
метамовні дескриптори 59, 60
метаслова семантичної розмітки 192
метатекстова розмітка 128
метафора 39, 88, 91, 157, 162, 186
метафоричність словосполучення 186
метод Арапова-Ратцевої 19
метод лексичного портретування слова 43
метод Шрейдера 19
методи зняття лексичної двозначності 191
методи статистики в перекладознавстві
 Див.: **статистичний аналіз**
методика встановлення еквівалентних відношень між різномовними лексемами 85
методика опрацювання емпіричного матеріалу з фрагментів нерівномірного обсягу 171
метонімія 40, 200
механізми творення значень слів 39
механізмів втрат і компенсацій 163
міжнародні вимоги щодо нормалізації термінології 97
міжчастиномовні омоніми 194, 195
мікроструктура словника 14, 15, 16, 20, 21
мінікорпус 151
міра індивідуального в перекладі
 Див.: **індивідуальний стиль перекладача**
множинність дефініцій терміна 98
множинність перекладів 149
мова-посередник 92, 196
морфеми 30, 94, 99, 117, 124, 187, 200
морфологічна маска слова 124
морфологічна розмітка 117
морфологічна структура терміна 94, 99
морфологічний тег 200
морфологічні характеристики лексеми 187
морфолого-синтаксична розмітка 99
мультимедійні компоненти словника 22, 23
навчальний корпус 140-142

навчання перекладу 140
 наголос 33, 35, 117, 139
 назви біологічних видів 65, 77, 79, 103
 напівавтоматична розмітка 201
 напівдопоміжні дієслова 146, 165
 напрямок перекладу 146
 нейромережевий аналіз 194
 нелексичні словники 36
 несементні явища в тексті (аналіз) 150
 німецька мова 8, 48, 50, 100, 128, 134, 137, 193, 209
 новогрецька мова 134
 нормалізація термінології 76, 77, 79, 97, 104, 199
 об'єктивне/суб'єктивне в перекладі 154
 обернений словник 18, 34, 36, 54, 160
 обернений частотний словник 160
 обсяг дефініційного словника 199
 обсяг корпусу 158, 160
 обсяг словника 12, 16, 20, 23, 27, 32, 34, 48, 68
 обсяг тексту перекладу 145, 146, 149, 184, 185
 односторонній запит пошуку 116, 120
 означений артикль 163
 означення 42
 okazіоналізм 42, 44, 153
 омографи 207
 омоніми 11, 12, 78, 80, 122, 161, 194, 207
 ономазіологічний словник синтаксем 63
 ономазіологічні словники 22, 35, 45-68, 79, 80, 86, 102
 описові семи 81
 орфографія 35, 115, 140
 орфографічні словники 35
 орфоепічні словники 35
 основа слова 191
 особова форма дієслова 126
 парадигматичний клас
 Див.: лексико-граматичний клас, частини мови
 паралельні тексти 87, 125-129, 132, 145, 162
 паралельні корпуси 127-129
 параметри корпусів 115, 128, 167
 параметри пошуку в корпусі 128, 180
 параметризація словників 15
 паремії 57
 пейоративна лексика 157
 Див. також: дисфемізми
 перевірка достовірності вибірки 119, 164, 168-172
 перевірка орфографії 140, 185
 перевірка сполучуваності лексем 140
 перевірка узусу мовних одиниць 140
 перевірки достатності корпусу
 Див.: перевірка достовірності вибірки
 перевірка якості перекладу 185
 переклад субтитрів 183

перекладацька стратегія 45, 153, 156, 157
 перекладацький дискурс 167
 перекладацькі корпуси 25, 125, 127
 перекладний дискурс 149, 158
 перекладний ономасіологічний словник 62
 переміщення значення 39
 перенесення значення 39, 42
 Див. також: **метафора, метонімія, механізми творення значення**
 перервана дефініція 82
 перестановка глоси і леми 26
 перестановка оригінального і перекладного текстів у
 комп'ютерній перекладацькій пам'яті 25
 Див. також: **інверсійний словник**
 перифрази 148, 164, 205, 206
 підготовка до перекладу конференцій 159
 підкорпус 115, 117, 119, 128., 132
 підмова 80, 149
 піктографічні словники 22
 повтор (текстова категорія) 13, 19, 51, 145, 151, 154-157, 165
 погіршення значення 39
 подвійні латинські назви 79
 позамовні чинники перекладу 26, 146, 149
 покращення значення 39
 полісемія 80
 польська мова 33, 48, 88, 116, 128, 134, 138, 162
 поля термінологічної картки 98, 100
 поняттєві схеми Дж. Ф. Сови 107
 поняття 30, 54, 62, 64, 65, 67, 68, 76, 78, 81, 82
 порівняння двох версій документа 182
 португальська мова 134
 порядок слів 151
 посилення значення 39
 послаблення значення 39
 початкова словникова форма 126
 Див. також: **лема**
 пошук в означеннях 54
 пошук за анаграмою 34
 пошук за лемою 134
 пошук лексеми 129
 пошук словоформи 120, 129
 пошук у корпусах текстів 120
 пошук у текстовому файлі 180
 права частина словникової статті 15
 прагматема 30
 прагматичний рівень 36
 префікс при- 151, 152, 153
 префікс *anti-* 89
 префікс при- 151
 прикметник 140, 151, 159, 188, 192, 200, 207
 прикметникове керування 22, 122, 140
 принцип перетинання 67, 68, 118, 204

Див. також: **багатокатегорійна класифікація, класифікація лексики, семантична розмітка**
принципи пошуку в корпусах текстів 120
проблема кількості лексико-семантичних варіантів 43
проблеми пошуку засобів вираження цільовою мовою 45
програм частиномовної розмітки 190
програми машинного перекладу 26, 101, 108, 124, 157, 174, 185, 191, 196, 206
програми машинного перекладу на основі комп'ютерної перекладацької пам'яті 124
програми розмітки текстів 195
просодична розмітка 117
псевдофлексія 191
психологія (терміни) 65
психолінгвістична експертиза тексту 154
пунктуаційний маркер 165
пунктуація 140, 154, 165, 185
реалія 89, 103
реверс оригінального і перекладного текстів у комп'ютерній перекладацькій пам'яті 25
регістр у словниках 99, 106
Див. також: **соціолокальні помітки у словнику**
регулярні вирази 120-125, 139, 142, 181
регулярні значення 200
реєстрове слово 12, 17, 99
Див. також: **лема**
резервна копія файлу 178, 185
реляційна база даних 108
репрезентативний корпус 164
репродуктивний етап перекладу 21, 27, 30, 45, 63, 86
ретороманська мова 100
референтний корпус 158
рецептивний етап перекладу 21, 27, 30, 45, 86
речення 82, 108, 117, 124, 137, 150, 158, 165, 185, 195, 208
ритм художньої прози 150
різномовні співвідносні тексти 145
родові семи 67
розбіжність частоти явищ в оригіналі і перекладі 168
розв'язання труднощів перекладу за допомогою корпусів 115
роздільник 110, 121, 182
розмітка 15, 40, 66, 95, 117, 120, 128, 200, 202
Див. також: **метамовна розмітка, семантична розмітка, синтаксична розмітка, соціолінгвістична розмітка, тематична розмітка, частиномовна розмітка**
розмітка за типом дискурсу 202
розмітка мікроструктури 15
розмовна лексика 157
розмітка структурних частин словника 15, 16, 157
розмовники 36
розпізнавання граматичних категорій 174, 187
розпізнавання лексико-граматичних класів 174

Див. також: **частиномовна розмітка**
розпізнавання семантичних розрядів 174
розпізнавання семантичних характеристик 190
розпізнавання синтаксичних функцій 174
розпізнавання теми тексту 66, 156, 203
розрахунок лексичного багатства тексту 13
розрахунок обсягу тексту 182
розрізнення значень лексеми 41, 197
розряди прикметників 207
розширення значення лексеми 39
романо-германські мови 26
романські мови 30, 39, 87, 209
російська мова 66, 128, 134, 138, 155, 207, 209
рубрики ономазіологічного словника 65, 68
семантизація 21
семантична відстань 19
семантична близькість 19, 207
семантична маска словосполучення 124
семантична мережа тексту 187
семантична розмітка 117, 200, 201, 203, 207
семантичний аналіз 86, 90, 92
семантичний словотвір 39
семантичний синтез 86
семантичний тег 200
семантичні розряди 17
семантичні відношення 53
семантичні групи 206
семантичні множники 66, 196
семантичні розряди 16
семантичні типи іменників 207
семантичні функції 82, 95, 106
семасіологічні словники 21, 32, 33, 45, 79
семема 196
семіотичний закон простоти 37
середня частота 119, 152, 169
символ-роздільник 110, 121, 182
синонім 27, 28, 33, 46, 53, 60, 64, 80, 81, 83, 85, 93, 100, 102, 104, 105, 134, 156, 157
синонімічне багатство мови перекладу 156
синонімічний повтор 151
синоптична схема 64, 65, 66, 67, 78
синтагматичні відношення 20, 22, 99
синтагматичні словники 22
синтаксема 30, 63
синтаксична розмітка 117
синтаксична структура тексту 158
синтаксичні конструкції 164
система автоматичного пошуку відповідей на питання 206
система контролю версій електронного документа 182
систематизація 51
системи комп'ютерного перекладу 207

Див. також: **програми машинного перекладу**
системи штучного інтелекту 15, 18, 44, 82, 172, 185-212
сільське господарство (терміни) 77
складний синтаксис 148
слова широкої семантики 147, 167
словник біграм 193
словник жестів 25
словник загального користування 27, 78
словник мови перекладу 155
словник новацій 33
словник помилок 35
словник синонімів і антонімів 46
словник словотвірної та морфемної будови слова 36
словник сполучуваності 23, 38, 46, 193
словник споріднених ідей 35, 51
словник сумнівів 35
словник фразеологічних синонімів 60
словник синтаксем 63
словники для перекладачів 27
словники крилатих виразів 48
словники мовленнєвого етикету 36
словники ситуативних виразів 36
словники цитат 36
словник культури 36
словникова стаття 12, 14, 82, 103, 191, 207
слововживання 13, 158, 165
словосполучення 25, 127, 192, 198, 206
словотвір 36, 89, 99, 153, 165, 191, 194
словоформа 13, 120, 124, 165, 190, 194
сміслові зв'язки між лексико-семантичними варіантами 66
сортування термінів 102
соціолектна розмітка 8
соціолінгвістична розмітка 116, 203
соціолінгвістичні характеристики 32, 117
соціолокальні помітки у словнику 16, 82, 157
спеціалізовані словники 78, 94
співвідносні тексти 132, 168
співвідношення між формою і змістом у словнику 32
співіпоніми 53, 105, 106
сполучуваність слів 16, 22, 23, 27, 38, 41, 44, 46, 52, 88, 92, 100, 115, 120, 122, 124, 126, 140, 156, 157, 159, 162, 165, 186, 189, 192, 193, 194, 195
спрошення тексту в перекладі 145
стандартне відхилення у статистиці 119, 152, 169
статистичний аналіз 40, 92, 119, 152, 168-171
статус термінологічної картки 100
створення резервних копій текстового файлу 178
стеммінг 191
стилістичні помітки у словнику 82
стилістично високі лексеми 148, 155
стилістично знижена лексика 148, 163

стилістично знижена фразеологія 155

стиль

Див.: **індивідуальний стиль, ідіостиль**

стратегія перекладу 45, 153, 156, 157

структура речення 150

структурна розмітка 128

структурно-семантичне співвіднесення 193

субординація (відношення) 106

субтитри 183

суплетивні форми 122

суспільно-політичні чинники в перекладі 149

сутності (семантика) 203

суфікси 163, 164, 191

суфікси зневажливої семантики 163

таксономія 56, 65, 78

творення значень слів 39

тег (у розмітці) 17, 40, 95, 98, 202

тезауруси 49, 35, 51, 53, 63, 64, 201

тезаурусні функції 53

тема-рематичні зв'язки 150, 151

тематична домінанта тексту 66

тема тексту 66, 156, 203

тематична розмітка 66, 156, 203

тематичне впорядкування термінології 78

тематичні параметри корпусів 116

тематичні словники 48, 63

теорія детермінант 192

теоцентрична модель синоптичної схеми 64

терміни з предметним значенням 93

терміни-реалії 89, 103

термінологічна картка 78, 81, 98, 100

термінологічне бюро 100

термінологічні бази даних, доступні онлайн 99

термінологічні картки 80

термінологічні синоніми 74, 77, 80, 93, 100, 102, 103, 104, 106

термінологічні словники 78, 94

терміносистема 82, 94, 200

типи розміток 117

типи користувачів словника 27

типологічні характеристики цільового та оригінального текстів 146

тлумачний словник 17, 21, 32, 33, 40, 42, 43, 46, 81, 83, 86, 106, 105, 153, 198

токен 13

Див. також: **слововживання**

транскрипція з адаптацією 89

транскрипція 89, 99

тренування штучної нейронної мережі 209

триграми 127, 192, 194, 208

труднощі репродуктивного етапу перекладу 21, 27, 30, 45, 63, 86

труднощі рецептивного етапу перекладу 21, 27, 30, 45, 86

турецька мова 128
 узус 140, 142, 159
 укладання корпусу 141
 укладання малого корпусу 166
 українська мова 24, 32, 48, 49, 51, 54, 57, 60, 63, 66, 67, 77, 81, 87, 89,
 91, 105, 116, 129, 128, 130, 134, 160, 201, 209
 “улюблене” слово 148, 156, 165
 “універсалії” перекладу 145, 147
 універсальна десяткова класифікація 79
 універсальні семантичні категорії 196
 уніфікації опису лексики 205
 уніфікація опису термінологічних даних 74, 97, 99
 упорядкування малих корпусів 118
 фактор частотності у перекладі 168
 фауна (терміни) 93
 фільтри пошуку 38, 99, 120, 128, 135, 139
 фінська мова 134
 флексер 191
 флексія 161, 191
 Див. також: квазіфлексія
 флора (терміни) 93
 фонемна структура терміна 94
 фонетичне слово 11
 формат енциклопедичної статті 96
 формат опису термінологічних одиниць 94
 формат субтитрів *ass* 183
 формат субтитрів *srt* 183
 формат текстового файлу 178
 формат термінологічної картки 98, 100
 формати субтитрів 179, 183
 фразеологічний тезаурус 58
 фразеологічні одиниці 24, 47, 57, 60, 155
 фразеологічні синоніми 19, 57, 60
 французька мова 31, 34, 40, 48, 50, 77, 87, 92, 100, 116, 134, 139, 209
 функції реверсу оригінального і перекладного текстів у
 комп'ютерній перекладацькій пам'яті 25
 функціональні стилі 118
 хибні друзі перекладача 88, 142, 186
 хронологічні параметри корпусів 116
 художні тексти 14, 163, 36, 130, 160, 152, 156, 160, 164, 167
 художній дискурс 167
 художня проза 36, 150, 160
 цільова мова 14
 цільовий текст 14
 часова розмітка 183
 частиномовна маска 134
 частиномовна розмітка 8, 94, 115, 123, 124, 125, 174, 188, 189, 191,
 198, 201
 частиномовний склад терміна 94
 частиномовна маска словосполучення 124
 частини мови 8, 94, 115, 123, 124, 125, 174, 189, 191, 198, 201, 207

Див. також: **артикуль, дієслово, займенник, іменник, прикметник**
часткові теорії перекладу 160, 169
частота 22, 32, 36, 66, 130, 145, 148, 154, 165
числівник 151, 188
штучний інтелект 15, 18, 44, 82, 172, 185-212
штучні нейронні мережі 151, 173, 174, 209
юридичний словник 79
японська мова 134
bag of words 193
cooccurrence 52
CQL 123
defining vocabulary 198
entity types 203
hapax legomenon 155
JSON 110
machine learning 172
n-грами 127, 192, 208
semantic primes 196
SQL 109
structural semantic interconnection 193
TBX 97
Termdat 76
text mining 187, 193, 202
TM 208
TTR 13, 146, 150, 165, 186
Word sketch engine 123
XML 96

ДОДАТОК І. Параметри основних текстових корпусів

Кількість мільйонів слова- живань	Охоплен ий часовий період, роки	Категорії підкорпусів, фільтри пошуку	Лінгвіст ична розмітк а	Лематиз атор	Пошук виразів з кількох слів	Пошук за регуля рними вираза ми	Наявність переклад них текстів у корпусі	Синхр онний зріз мови	Діахрон ічний зріз мови	Примітки
Українська , Корпус української мови mova.info										
100		Функціонально- стилістико-тематичні категорії: законодавчі, наукові, публіцистичні, поетичні, фолклорні, художні прозові тексти; соціолінгвістичний критерій: стать автора.	Ні	Так	1 та два слова	Ні	Ні	Так	Ні	Корпус автоматично генерує частотний словник обраного підкорпусу; є можливість налаштування користувачького корпусу.
Українська , ГРАК (Генеральний регіональний анотований корпус української мови) http://www.parasolcorpus.org/bonito/run.cgi/first_form										

290	1918-2018	За автором, назвою тексту, датою, регіоном, жанром, містом публікації, видавництвом, роком публікації,	Так	Так	Словоформ а, лема, словосполучення, речення	Так	Так	Так	Так	Автоматично генерується частотний словник підкорусу; пошуку мовою CQL.
Англійська, BNC (British National Corpus) https://corpus.byu.edu/bnc/										
100	1980-початок 1990-х	За жанром, типом мовлення, темою, за частиномовною та граматичною характеристикою слів.	Так	Так	Словоформа, лема, словосполучення, речення, синонімічні зв'язки, морфолог. характ.	Так	Ні	Так	Ні	Потрібна реєстрація користувача.
Англійська, COCA (Corpus of Contemporary American English) https://corpus.byu.edu/coca/										
560	1990-2017	За жанром, типом мовлення, за частиномовною та граматичною характеристикою слів.	Так	Так	Словоформа, лема, словосполучення, речення.	Так	Ні	Так	Ні	Потрібна реєстрація користувача, корпус можна завантажувати на персональний комп'ютер.

Англійська, iWeb Corpus https://corpus.byu.edu/iweb/										
14 000		За жанром, типом мовлення, за частиномовною та граматичною характеристикою слів.	Так	Так	Словоформа, лема, словосполучення, речення.	Так	Ні	Так	Ні	Можливості залежатимуть від програми опарцювання корпусу.
Кількість мільйонів слововживань	Охопленний часовий період, роки	Категорії підкорпусів, фільтри пошуку	Лінгвістична розмітка	Лематизатор	Можливість пошуку виразів з кількох слів	Можливість пошуку за регулярними виразами	Наявність перекладних текстів у корпусі	Синхронний зріз мови	Діахронічний зріз мови	Примітки
Іспанська, CREA (Corpus de Referencia del Español Actual) http://corpus.rae.es/creanet.html										
160		За датою, темою, назвою тексту, автором, регіоном, типом видань, типом мовлення	Так	Так	Словоформа, лема, словосполучення, речення.	Так	Ні	Так	Так	Автоматично генерується частотний словник словформ.

		тощо.								
Іспанська, CORDE (Corpus Diacrónico del Español) http://corpus.rae.es/cordenet.html										
250		За датою, темою, назвою автором, регіоном, типом видань, типом мовлення тощо.	Ні	Ні	Словоформ а, словосполучення, речення.	Так	Ні	Ні	Так	
Іспанська, CDH (Corpus del Diccionario Histórico del Español) http://web.frl.es/CNDHE/view/inicioExterno.view										
353		Назва, автор, дата, регіон, тема, лексико-семантичний клас.	Так	Так	Словоформа, словосполучення, речення.	Так	Ні	Ні	Так	
Іспанська, COPRES XXI http://www.rae.es/recursos/banco-de-datos/corpes-xxi										
225		За датою, темою, назвою тексту, автором, регіоном, типом видань, типом мовлення, жанром, лексико-семантичним	Так	Так	Словоформ а, словосполучення, речення.	Так	Ні	Так	Так	

		класом, шрифтом.								
Італійська , PAISÀ, італійська http://www.corpusitaliano.it/it/contents/funding.html										
1 250		---	Так (частин омовна , члени реченн я)	Так	Словоформ а, словосполу чення, речення.	Так	Ні	Так	Ні	
Італійська , CORIS (Corpus di Riferimento dell'Italiano Scritto) http://corpora.dslo.unibo.it/coris_ita.html										
130 (разом з CODIS)		Жанрово-тематичні категорії	Так	Так	Словофор ома, лема	Так	Ні	Так	Ні	Запити можна здійснювати мовою CQL; база текстів поповнюється щотри роки.
Кількість мільйонів слово-	Охоплен ий часовий	Категорії підкорпусів, фільтри пошуку	Лінгвіст ична розмітк	Лемати затор	Можли- вість пошуку	Можли- вість пошуку	Наявність переклад них	Синхр онний зріз	Діахрон ічний зріз	Примітки

вживань	період, роки		а		виразів з кількох слів	за регуля рними вираза ми	текстів у корпусі	мови	мови	
Італійська , CODIS (Corpus di Riferimento dell'Italiano Scritto) http://corpora.dslo.unibo.it/CODIS/										
130 (разом з CORIS)		Жанрово-тематичні категорії	Так	Так	Словоформ а, лема	Так	Ні	Так	Ні	Запити можна здійснювати мовою SQL; база текстів поповнюється щотри роки.
Німецька , Проект Deutscher Wortschatz http://wortschatz.uni-leipzig.de/										
250		---	---	---	---	---	---	Так	Ні	На основі запиту генерується візуальний тезаурус.
Німецька , Limas-Korpus http://www.korpora.org/Limas/										
1	1970- 1971 pp.	33 тематичні категорії	Ні	Ні	Так	Так	Ні	Ні	Так	

Німецька , Deutsche Referenzkorpus (kurz DeReKo) http://www.ids-mannheim.de/cosmas2/web-app/hilfe/allgemein/dbanfrage.html										
1800	з 1964 р.	Класифікаований за періодами, історичний корпус, морфосинтаксичний розмічений корпус; нові надходження	Так	Так	Так	Так	Ні	Так	Так	
Польська , Zrównoważony tematycznie i gatunkowo Korpus Języka Polskiego PWN ("Państwowe Wydawnictwo Naukowe") https://sjp.pwn.pl/korpus										
100		---	Так	Так	Словоформа, словосполучення, речення.	Так	Ні	Так	Ні	
Польська , Narodowy Korpus Języka Polskiego http://nkjp.pl/										
1500		---	Так	Так	Словоформа, лема, словосполучення, речення	Так	Так	Так	Ні	
Російська , Национальный корпус русского языка										

http://www.ruscorpora.ru/										
> 600		Морфологічні, семантичні категорії слів; типи виразів; пунктуаційні знаки в контексті	Так	Так	Так	Так	Так	Так	Так	Залити мовою CQL; можна формулювати маски виразів до трьох слів.
Російська , Корпус літературного російського мови http://www.narusco.ru/										
> 1					Ні	Ні				Можливість пошуку за акцентуацією.
Французька , Frantext https://www.frantext.fr/										
251		За датою, темою, жанром	Так	Так	Словоформ а, сполучення двох, трьох слів.	Так	Ні	Так	Так	Ресурс платний.
Франтекст , Frantext, http://www.cnrtl.fr/corpus/frantext/frantext.php										
Приблизно 14.6 ²⁷	500 творів французької	За жанром, автором, часовим періодом.	Ні	Ні	—	Ні	Ні	Так	Так	Для здійснення пошуку необхідно завантажити корпус або підкорпус на

Обсяг зазначеного корпусу вказано у кількості символів, зокрема, понад 146 млн. Відповідно до даних Раві Паріха, середня довжина графічного слова у французькій мові становить понад 10 символів. Докладніше статистику для французької та інших мов можна переглянути за адресою: <http://www.ravi.io/language-word-lengths>

	літерату ри від XVIII до XX століття									комп'ютер.
--	--	--	--	--	--	--	--	--	--	------------