

Міністерство освіти і науки України
Чернівецький національний університет
імені Юрія Федьковича



Інноваційні технології перекладу

Навчальний посібник

Чернівці
ЧНУ
2023

УДК 81'25=112.2(075.8)
К 937

Рекомендовано до друку ухвалою кафедри
германського, загального і порівняльного мовознавства
(протокол № __ від _____ 20 __ року)

К 937 Інноваційні технології перекладу : навч. посібник для студентів першого (бакалаврського) рівня вищої освіти [Електронний ресурс] / уклад. А. А. Король. Чернівці : ЧНУ. 2023. 119 с.

Пропоноване навчальне видання рекомендовано для дисципліни «Інноваційні технології перекладу», що належить до обов'язкових компонентів ОПП «Німецько-український переклад та переклад з другої іноземної мови» для здобувачів першого (бакалаврського) рівня вищої освіти, спеціальності 035 Філологія, 035. 043 Філологія (Германські мови та літератури (переклад включно), перша – німецька)), галузі знань: 03 Гуманітарні науки. Навчальний посібник спрямований на підготовку здобувачів до здійснення якісного та швидкісного усного та письмового перекладів, використовуючи сучасні комп'ютерні технології.

УДК 81'25=112.2(075.8)
© Чернівецький національний
університет, 2023

ЗМІСТ

Вступні зауваження	5
--------------------------	---

ТЕМА 1. Типи, структура, характеристики комп'ютерних словників та інформаційних онлайн-ресурсів

9

1.1. Визначення терміна “слова” з погляду комп'ютерної лексикографії	10
1.2. Організація словника: макроструктура та мікроструктура ...	11
1.3. Принципові відмінності між електронними і паперовими словниками	12
1.4. Критерії класифікації електронних словників	19
1.4.1. Види двомовних перекладних словників	27
1.4.2. Види одномовних словників та їхнє значення для практики перекладу	30
1.4.3. Характеристики електронного конкордансу	34
1.5. Особливості структури і використання онемасіологічних словників	36
1.5.1. Тематичний словник	38
1.5.2. Атлас лексики	39
1.5.3. Тезаурус	40
1.5.4. Інверсійний словник	41
Використана література та інформаційні ресурси	43
Семінарське заняття 1	44
Теоретичні питання для обговорення	45
Практичні завдання	46
Лабораторна робота 1	46

ТЕМА 2. Корпусні ресурси в діяльності і підготовці перекладача. Перевірка якості перекладу та функції текстових редакторів

47

2.1. Електронні лінгвістичні корпуси як довідковий ресурс перекладача	47
2.2. Корпуси паралельних текстів	59

2.3. Основні принципи пошуку в корпусах текстів	64
Використана література та інформаційні ресурси	70
Семінарське заняття 2	71
Теоретичні питання для обговорення	71
Практичні завдання	72
Лабораторна робота 2	73

ТЕМА 3. Системи автоматизованого перекладу (CAT-системи)

3.1. Загальна характеристика систем машинного та автоматизованого перекладу	74
3.2. Системи автоматизованого перекладу (CAT-системи)	84
Використана література та інформаційні ресурси	90
Семінарське заняття 3	90
Теоретичні питання для обговорення	90
Практичні завдання	92
Лабораторна робота 3	92

ТЕМА 4. Локалізація перекладу

4.1. Переклад, локалізація, адаптація: основні відмінності	104
4.2. Методи локалізації веб-сайтів	108
4.3. Лінгвістичні та екстралінгвістичні рівні локалізації	110
Використана література та інформаційні ресурси	115
Семінарське заняття 4	115
Теоретичні питання для обговорення	115
Практичні завдання	115
Лабораторна робота 4	116
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	117

ВСТУПНІ ЗАУВАЖЕННЯ

При всьому різноманітті теоретичних і методичних підходів до підготовки майбутніх перекладачів проблема формування їхньої інформаційної компетентності наразі є надзвичайно актуальною.

Варто зазначити, що як у бакалаврських, так і у магістерських програмах переважної більшості університетів світу значне місце займають дисципліни, пов'язані з підготовкою майбутніх перекладачів до застосування інформаційних технологій у професійній діяльності. Тому перш за все цінним вважаємо вивчення досвіду зарубіжних університетів, які здійснюють підготовку перекладачів [Амеліна; Тарасеко 2017, с. 7].

Основну увагу ми звернули на підготовку перекладачів у Федеративній Республіці Німеччина, що здійснюється у 8 класичних університетах (Universität), 9 університетах прикладних наук (Fachhochschule) і 7 інститутах (Institut). Досліджуючи досвід навчальних закладів Німеччини у контексті формування основного набору компетентностей сучасного перекладача, К. Егеберг акцентує увагу не тільки на необхідності володіння на високому рівні рідною та іноземними мовами, а й особливо на тих вимогах, які пов'язані зі сформованістю інформаційної компетентності, а саме:

– грамотному використанні термінології у перекладі, володінні стратегіями пошуку в мережі Інтернет, що означає не тільки знайомство з існуючими банками даних, а й уміння створювати власний банк даних, який оптимізує роботу перекладача;

– володіння компетенцією, за якою перекладачі повинні мати уявлення про технологію здійснення перекладу програмами машинного перекладу, систему Translation Memory (TM), CAT-засоби (Computer Assisted Translation Tools) [Egeberg 2003].

Спрямовуючи процес підготовки майбутніх перекладачів на формування у них інформаційної компетентності, вищі навчальні заклади Німеччини враховують, що можливості застосування інформаційних технологій у професійній діяльності перекладачів варіюються від базової електронної підтримки (наприклад, робота з програмами обробки текстів чи електронними словниками) до повністю автоматизованого перекладу текстів за допомогою відповідного програмного забезпечення. Зокрема, у галузевому

перекладі (технічному, юридичному, медичному тощо) неможливо обійтися без використання відповідних програмних засобів [Амеліна; Тарасеко 2017, с. 8]. Тому програмні засоби, необхідні для забезпечення ефективної діяльності професійних перекладачів, стали невід'ємною частиною змісту дисциплін, які входять до навчальних планів і програм вищих навчальних закладів Німеччини [Амеліна 2014]. При цьому враховується, що знання програмних засобів має досить короткий період актуальності з огляду на швидку появу більш нових його версій з розширеними функціональними можливостями. Тому у німецьких університетах розглядається питання про те, як забезпечити сталий розвиток інформаційної компетентності професійного перекладача в галузі перекладу у зазначеному аспекті.

Навчальні заклади Німеччини здійснюють інформаційну підготовку майбутніх перекладачів різними шляхами, що передбачають відмінності як в організації опанування навчального матеріалу, так і у його змістовому наповненні. Зокрема, університет Хільдесхайм (Universität Hildesheim) у процесі підготовки бакалаврів з перекладу пропонує для вивчення такі дисципліни: «ІТ і використання засобів масової інформації»; «Основи фахової комунікації»; «Менеджмент перекладів» тощо.

Натомість підготовка перекладачів-бакалаврів у Лейпцигському університеті (Universität Leipzig) передбачає вивчення дисципліни «Термінознавство і мовна технологія». При цьому пізнавальний процес студентів спрямовується на вивчення технічних і програмних засобів, засвоєння конкретних прийомів роботи з термінологією, які, у першу чергу, зорієнтовані на застосування термінологічних баз при здійсненні перекладу з використанням систем автоматизованого перекладу [Studienführer Translation Bachelor of Arts 2014, с. 4].

У Заарландському університеті (Universität des Saarlandes) нормативною для майбутніх перекладачів є дисципліна «Термінологія, обробка мовних даних, автоматизований переклад», у курсі якої вони вивчають теоретичні та практичні аспекти перекладу, пов'язані із застосуванням спеціалізованого програмного забезпечення, зокрема: інструменти для автоматизованої роботи з термінологією, для машинного і автоматизованого перекладу та управління проектами в перекладі [Амеліна; Тарасеко 2017, с. 9].

У деяких німецьких університетах не передбачені окремі дисципліни, пов'язані з інформаційною підготовкою перекладачів. Однак, до змісту навчальних планів підготовки перекладачів обов'язково входять відповідні модулі або теми. Так, в університеті імені Йоганнеса Гутенберга міста Майнц (Johannes Gutenberg-Universität Mainz) із загального обсягу 360 годин нормативної дисципліни «Перекладознавство» на вивчення тем з володіння інформаційними технологіями в перекладі відводиться 90 годин. Зокрема, до змісту зазначеної дисципліни, крім огляду актуального стану досліджень в сучасному перекладознавстві як науці про продукти і процеси письмового та усного перекладу, входить поглиблене вивчення обраних студентом тем як із загального, так і прикладного перекладознавства, шляхом дослідження, наприклад, теорії перекладу, історії перекладу, міждисциплінарних підходів до перекладознавства, автоматизованого перекладу, систем Translation Memory тощо [Modulhandbuch des Bachelorstudiengangs B.A. Sprache, Kultur, Translation 2013].

Послідовність інформаційної підготовки майбутніх перекладачів найбільш яскраво прослідковується в університеті прикладних наук Кельна. Бакалаври-перекладачі, які навчаються в цьому університеті, у процесі вивчення різних дисциплін навчального плану шляхом засвоєння відповідних тем набувають теоретичних знань і практичних навичок щодо обробки мовних даних, особливо щодо управління термінологією, технологій перекладу з урахуванням галузевої специфіки тощо. У подальшому магістранти-перекладачі цього навчального закладу вивчають цілу низку окремих дисциплін, спрямованих на формування умінь у руслі інформаційної підготовки [Амеліна; Тарасеко 2017, с. 11].

Розглянемо до прикладу дисципліну «Перекладацький проект з використанням інструментів перекладу», яка призначена для формування таких здатностей: самостійно здійснювати перекладацькі проекти (управління проектом і переклад) на основі інструментів перекладу; перекладати з іноземної мови на німецьку та навпаки фахові тексти з різних предметних сфер, а саме технічного, юридичного, економічного спрямування. Дисципліна орієнтована на практичну реалізацію перекладацьких проектів, починаючи від розгляду вихідного матеріалу (наприклад, інструкції клієнта) до створення проектних структур відповідно до стандарту

«DIN EN 15038 Translation services – Service Requirements» і власне до перекладу з використанням спеціалізованих програмних засобів (переважно SDL Trados Studio, MultiTerm, WinAlign). Передбачається практико-орієнтований переклад спеціалізованих текстів з урахуванням специфічної фахової термінології й специфіки відповідного типу тексту; розробка та впровадження стратегій дослідження за допомогою сучасних інструментів, таких як електронні бази даних, сервіси мережі Інтернет, інтерактивні системи автоматизованого перекладу з доступом до відповідних інформаційних ресурсів; оформлення перекладів у різних форматах, з використанням таких програмних продуктів, як MS Word, Tag Editor, MS Excel, MS PowerPoint тощо. Корисною для початку вивчення цієї дисципліни є наявність базових знань з програмного забезпечення із категорії CAT-систем [Амеліна; Тарасеко 2017, с. 13].

Розробляючи плани підготовки майбутніх перекладачів з урахуванням вивчення ними новітніх інформаційних технологій, зарубіжні університети спрямовують процес їхньої підготовки на різні можливості застосування інформаційних технологій у професійній перекладацькій діяльності, які варіюються від базової електронної підтримки (наприклад, використання програм обробки текстів чи електронних словників) до автоматизованого перекладу текстів шляхом застосування відповідного програмного забезпечення.

На основі розгляду навчальних програм і вивчення досвіду підготовки перекладачів у провідних закладах вищої освіти Європи до переліку обов'язкових компонентів освітньо-професійної програми «Німецько-український переклад та переклад з другої іноземної мови» першого (бакалаврського) рівня вищої освіти за спеціальністю 035 Філологія, спеціалізації 035.043 Філологія (Германські мови та літератури (переклад включно), перша – німецька) галузі знань 03 Гуманітарні науки включено навчальну дисципліну «Інноваційні технології перекладу».

ТЕМА 1

Типи, структура, характеристики комп'ютерних словників та інформаційних онлайн-ресурсів

План лекції:

- 1.1. Визначення терміна “слова” з погляду комп'ютерної лексикографії.
- 1.2. Організація словника: макроструктура та мікроструктура.
- 1.3. Принципові відмінності між електронними і паперовими словниками.
- 1.4. Критерії класифікації електронних словників.
 - 1.4.1. Види двомовних перекладних словників.
 - 1.4.2. Види одномовних словників та їхнє значення для практики перекладу.
 - 1.4.3. Загальні характеристики електронного конкордансу.
- 1.5. Особливості структури і використання ономасіологічних словників: тематичний словник, атлас лексики, тезаурус, інверсійний словник.

Чому ж перекладачі звертаються по допомогу до словників? Типовими є такі 3 основні сценарії з безлічі можливих:

1. Ми зустріли у вихідному тексті взагалі невідоме нам слово, не знаємо його значення і відповідно не можемо його перекласти в певному контексті.
2. Ми приблизно розуміємо, що означає це слово, але хочемо перевірити, уточнити нюанси вживання.
3. Ми чудово розуміємо контекст вживання слова, який фрагмент дійсності за ним стоїть, але не знаємо, як це висловити іншою мовою.

Для розв'язання конкретного перекладацького завдання існує певний тип словника та інформаційних онлайн-ресурсів в допомогу як усному, так і письмовому перекладачеві. У цьому підрозділі пропонується розглянути принципові відмінності між електронними і паперовими словниками; критерії класифікації електронних словників; види двомовних та одномовних перекладних словників; характеристики електронного конкордансу; особливості структури і використання ономасіологічних словників (тематичний словник, атлас лексики, тезаурус, інверсійний словник).

1.1. Визначення терміна “слова” з погляду комп’ютерної лексикографії

Значення “слова” з термінологічної точки зору не є чітко окресленим. З метою автоматичної обробки тексту, укладання комп’ютерних словників та будь-якого іншого технічного опрацювання текстового матеріалу все ж таки необхідно встановити чітке термінологічне розмежування між різними значеннями цього поняття. 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 разом. Аналогічно фонетичними називаються ті слова, що вимовляються безперервно [Фокін 2019, с. 11].

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

Лексема – це слово як частина лексичної системи певної мови. Ключовою характеристикою лексеми є наявність спільного лексичного значення для всіх форм.

Зрештою, терміни “слово”, “лексема”, “вокабули” не мають взаємовинятковий характер. 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 твердження можна розтлумачити по-різному: 50 000 реєстрових слів (лем), словникових статей чи навіть графічних слів [Фокін 2019, с. 12].

Ще два важливих для опису лексики терміни – це **словоформа** та **слововживання**. Під час підрахунку обсягу тексту може виникнути непорозуміння. Тому введено поняття словоформ і слововживань. Кількість “словоформ” – це кількість графічних слів у тексті без урахування повторів; “слововживання” (інколи позичається англійський термін “токен” (англ.: “token”), що, однак, не зареєстрований у провідних словниках української мови) – усі графічні слова з урахуванням повторів.

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 [Фокін 2019, с. 13].

1.2. Організація словника: макроструктура та мікроструктура

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 правої частини (**глоси**), що містить пояснення леми і, часом, додаткові характеристики [Фокін 2019, с. 14].

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 леми. У такому випадку лема і глоса мають настільки симетричний характер, що виявляється можливою їхня взаємна перестановка (якщо подається тільки один переклад леми). Такого роду словники зі значно обмеженою глосною 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 глосною часто взаємно симетричні, у той час як словники в повному сенсі цього слова характеризуються асиметрією між лемою та глосною: 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 частиномовну 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 [Фокін 2019, с. 15].

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 [Перебийніс 2009, с. 188-190].

Порівняймо до прикладу обсяг значення іменника «середовище» на прикладі паперового «Великого українсько-німецького словника» на 480 000 одиниць за ред. Михайла Смоля та онлайн-словника **multitran.com**.

<p><i>m</i> -n, -n, <i>Mittelstandsbauer</i> <i>m</i> -n, -n; 2. (людина посередніх здібностей) <i>mittelmäßiger Mensch</i> середовище <i>с</i> 1. <i>Umwelt</i> <i>f</i> =; (соціально-побутові умови) <i>Milieu</i> [mil'jə:] <i>n</i> -s, -s; (оточення, коло людей, довколишня обстановка) <i>Umgebung</i> <i>f</i> =, -en; в літературному ~і in literarischen Kreisen; 2. (сукупність природних умов) <i>Umwelt</i> <i>f</i> =, <i>Umgebung</i> <i>f</i> =, <i>Milieu</i> <i>n</i> -s, -s; охорона навколишнього ~а <i>Umweltschutz</i> <i>m</i> -es; 3. (речовини, тіла тощо, що заповнюють простір) <i>Medium</i> <i>n</i> -s, -di;en</p>	<p><i>lich</i>; книжн. <i>seriös</i>; 3. (важливий) <i>wichtig</i>, <i>gewichtig</i>; ~ недолік <i>schwerwiegender Mangel</i> серйозність <i>ж</i> <i>Ernst</i> <i>m</i> -es, <i>Ernsthaftigkeit</i> <i>f</i> = серйозно <i>ernst</i>; im <i>Ernst</i>; <i>ernstlich</i>; ~ хворий <i>ernstlich krank</i>; зовсім ~ <i>ganz ernst</i>, in <i>vollem Ernst</i>; ти ~? <i>ist das dein Ernst?</i>; <i>meinst du das ernst?</i> серп <i>ч</i> <i>Sichel</i> <i>f</i> =, -n; ~ місяця <i>Mondsichel</i> <i>f</i> =, -n серпанок <i>ч</i> 1. (пелена) <i>Dunst</i> <i>m</i> -es, <i>Dünste</i>, <i>Schleier</i> <i>m</i> -s, =; (туман) <i>Nebelschleier</i> <i>m</i> -s, =; 2. (прозора тканина) <i>Mull</i> <i>m</i> -(e)s</p>
--	---

Рис.1.1. Результат пошуку лемі «середовище» в паперовому двомовному словнику

Як бачимо з Рис. 1.1., паперовий словник на пошук за реєстровим словом «середовище» містить 3 позиції у той час, як в українській версії онлайн-словника **multitran.com** результат пошуку видає 46 фраз у 6 тематиках:

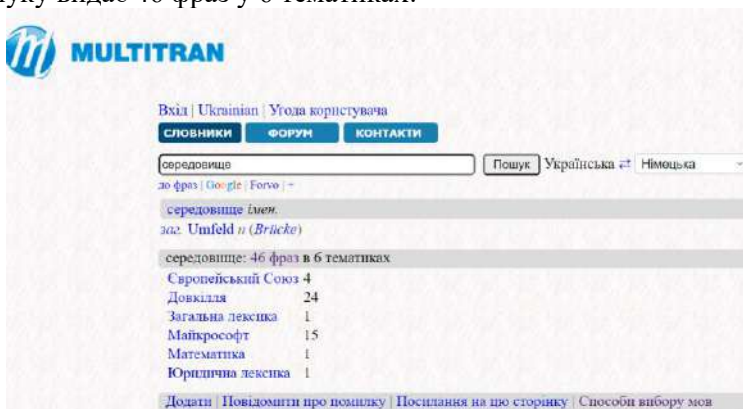


Рис.1.2. Результат пошуку лемі «середовище» в онлайн-словнику

Перейшовши за гіперпосиланням, отримуємо повну статтю до теми СЕРЕДОВИЩЕ у всіх тематиках одночасно:

Вхід Українська Угода користувача		
СЛОВНИКИ ФОРУМ КОНТАКТИ		
середовище		Пошук Українська ↔ Німецька
Терміни що містять середовище усі форми лише в заданій формі		
Тематика	Українська	Німецька
довк.	абіотичне середовище	abiotische Umwelt
кат., Майкр.	бібліотека шаблонів C++ для середовища виконання Windows	C++-Vorlagenbibliothek für Windows-Runtime
довк.	випробування у штучному середовищі	in vitro-Test (Дослідження, проведене у штучному середовищі)
довк.	водне середовище	aquatische Umwelt
довк.	геофізичне середовище	Geophysikalische Umwelt (Фізична земля і навколишні умови, що складаються із океанів і внутрішніх вод, нижчих і верхніх шарів атмосфери, вільного простору, мас і форми суші)
кат., Майкр.	датчик середовища	Umgebungssensor
довк.	екологія водного середовища	Wasserökologie (Вивчення взаємин між живими водними організмами, а також стосунків між ними організмами і їх довкіллям)
ЄС	етнічне середовище	ethnischer Hintergrund (Brücke)
довк.	живітне середовище	lebende Umwelt
довк.	нестійке середовище	Empfindliche Umwelt (Певна ділянка суходолу, що перебуває під державним чи приватним контролем, який має, або при реалізації відбудовних заходів може мати бажані екологічні умови. Ці умови сприяють збереженню і/або створенню умов для життя дикої природи, стабільній ґрунті, збереженню або вторинному використанню водних ресурсів, рослинного покриву і аналогічних важливих екологічних функцій. Райони з нестійким середовищем можуть бути різними за своїми розмірами, можуть охоплювати річки чи розповсюджені природні умови, рослини й тварини)
кат., Майкр.	об'єктна модель середовища виконання	Laufzeitobjektmodell
кат., Майкр.	об'єктна модель середовища виконання ПБД	BDC-Laufzeitobjektmodell
кат., Майкр.	об'єктна модель середовища виконання складових застосункових середовищ виконання	Composites-Laufzeitobjektmodell
кат., Майкр.	застосункові середовища виконання	Common Language Runtime
довк.	зайнятість і навколишнє середовище	Arbeit und Umwelt (Питання чи ініціативи, що відносять до взаємин екологічних проблем та проблем зайнятості, включно стійкої екологічно безпечної розвитку, вплив екологічного руху на процес створення нових робочих місць, створення робочих місць, пов'язаних з екологічною проблематикою)
юр.	закон про охорону навколишнього середовища від забруднення	Immissionsschutzgesetz (Brücke)
кат., Майкр.	клієнт групової політики для безпроводного середовища	Gruppenrichtlinienclient für Funkverbindungen
кат., Майкр.	клієнт групової політики для проводового середовища	Gruppenrichtlinienclient für Kabelverbindungen
кат., Майкр.	клієнтське середовище виконання служб ПБД	BDC-Client-Laufzeitumgebung
ЄС	культурне середовище	kultureller Hintergrund (Brücke)
ЄС	мовосприятливе середовище	sprachenfreundliche Umwelt (Brücke)

Рис. 1.3. Перехід за гіперпосиланням до повної статті в multitrans.com за всіма тематиками вживання слова «середовище»

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 [Фокін 2019, с. 16]. 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).

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 гласами, такі зв'язки допоможуть у розробці алгоритму автоматичного пошуку синонімів [Фокін 2019, с. 17].

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;
- 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.

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 конструювання; так само одним натисканням можна згенерувати, наприклад, словник німецьких дієслів, а, за потреби – німецьких перехідних дієслів, німецьких неперехідних дієслів. 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.

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) [Фокін 2019, с. 18-19].

Електронний словник має ряд можливостей, відсутніх у традиційних лексикографічних працях, зокрема з його допомогою можна швидко отримати інформацію, сформульованому користувачем у зручній саме для нього формі. Для цього існують різні *види пошуку*, які можна класифікувати таким чином:

1. Алфавітний пошук, що має наступні різновиди: простий пошук за першими символами в заголовку словникової статті; простий пошук за цілим словом у заголовку словникової статті; розширений пошук у заголовку словникової статті з використанням логічних виразів; розширений пошук підрядка в словниковій статті з використанням логічних виразів.

2. Пошук за ідентифікаційним номером словникової статті.

3. Тематичний пошук.

4. Пошук за власним ім'ям автора.

5. Міжсловниковий пошук.

6. Комбінований пошук.

8. Перекладний пошук.

9. Інформаційний пошук у мережі Інтернет.

Останнім часом, поряд з традиційними електронними словниками, виникли і активно розвиваються інтерактивні словники [Долинський; Скиба 2016, с. 91].

Як зазначає О. Панченко, *інтерактивний словник* – це електронна (комп'ютерна) лінгвістична змінна конструкція, в яку будь-який користувач може внести свої зміни та доповнення до даного тексту. Сучасні інформаційні та комунікаційні технології дозволяють створювати новий формат словників – інтерактивні словники, визначальною особливістю яких є незалежність

словникових об'єктів від просторових і тимчасових обмежень, використання гіпермедіальних об'єктів, динамічність словникових об'єктів і можливість їх редагування. Зміст інтерактивних словників, представлених у мережі Інтернет, є досить значний і постійно поповнюється. Ці словники різноманітні за обсягом: від всеосяжної Вікіпедії до Джинсового словника, що містить всього декілька слів. Низка словників є електронною обробкою вже наявних видань, інші ж створюються особисто тим або іншим авторами. Інтерактивний словник має такі переваги:

- ✓ можливість додавання користувачем коментарів або нових слів у персональний словник;
- ✓ система авторизації доступу до інформації, яку контролює адміністратор словника;
- ✓ захист бази даних від несанкціонованого використання (копіювання, зміни або видалення інформації) [Долинський; Скиба 2016, с. 92].

Вікісловник – яскравий зразок інтерактивного словника, що надає користувачеві широкі можливості для отримання необхідної інформації. Однак, він має низку недоліків, зокрема деякі користувачі навмисне спотворюють зміст статті чи створюють неправдиві статті. Електронні словники дають можливість створити «портрет слова», відмінний від того, який можна знайти у паперовому словнику, цей портрет є комплексним і багатограним, та складається з п'яти-дев'яти компонентів і різною кількістю зон (граматичною, етимологічною) [Долинський; Скиба 2016, с. 92].

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 доречно 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 [Фокін 2019, с. 20].

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-11” (<http://sum.in.ua>), “*duden.online*” (<https://www.duden.de/>) та інші. 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 [Фокін 2019, с. 21]. 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.

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). 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.

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 [Фокін 2019, с. 22]. 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. Якщо, наприклад, відкрити словникову статтю в *duden.online* леми „*Geige*“, окрім звичних граматичних характеристик і тлумачення значення слова можна прослухати його вимову, клацнувши *Aussprache*, та побачити зображення предмету на малюнку:

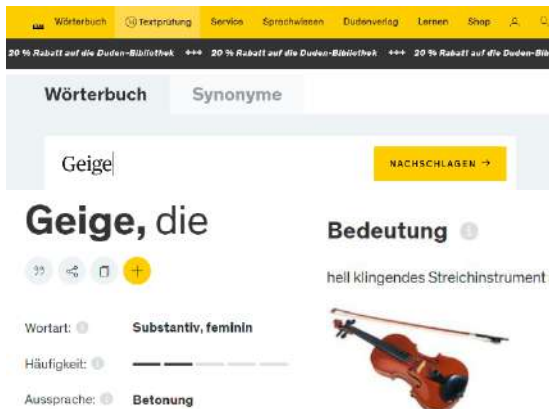


Рис. 1.4. Результат пошуку тлумачення лема „*Geige*“ в онлайн-словнику DUDEN

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 ще немає частотних словників сполучуваності, хоча близькими до них можна вважати словники біграм; для більшості мов ще не створено словників граматичної сполучуваності (для їхнього автоматичного створення потрібно мати корпус текстів з граматичною розміткою).

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.

Розглянемо деякі з Інтернет-словників, які використовуються найчастіше в професійній діяльності перекладача, які працюють з мовними парами «німецька-українська».

Мультитран (Multitran/Multytran) – словник-інтернет-система двомовних словників з підтримкою багатьох мов світу, один із найповніших і найпопулярніших автоматичних онлайн-словників. Сайт: <http://www.multitran.com>.



Це редагований багатомовний онлайн-словник, запущений 1 квітня 2001 року. Активно працюють форуми «Мультитран». Будь-який користувач може поставити питання і поспілкуватися з колегами-перекладачами. Важлива особливість – інтерактивне поповнення словників.

На даний момент дозволяє перекладати з/на 13 мов.

Містить:

- ✓ понад 5 мільйонів термінів;
- ✓ терміни кожної мовної пари словника доступні в двох напрямках перекладу;
- ✓ більше 800 предметних тематик;
- ✓ можливість активного самостійного поповнення словника користувачами;
- ✓ більше 1000 перекладачів, які поповнюють онлайн-версію МТ прямо на сайті;
- ✓ алфавітний, морфологічний, фразовий (по словосполученням) пошук;
- ✓ автоматичний пошук стійких словосполучень в тексті запиту;
- ✓ одночасний пошук в словниках, форумах, а також в базі англійських і російських пропозицій і базі посилань.

Словником можна користуватися через інтернет в онлайн-режимі; в автономному режимі на персональних комп'ютерах під Microsoft Windows; на телефонах через WAP-сайт; на персональних комп'ютерах під GNU / Linux.

Важливо зазначити, що Мультитран перекладає лише слова та фрази, а текст, навіть невеликий шматок, що складається буквально з двох-трьох пропозицій, перекласти за допомогою цього сервісу не вдасться.

Glosbe є багатомовним онлайн-словником, який створюється носіями мов. Він містить безкоштовні словники для майже кожної існуючої мови та пам'яті перекладів з 1 013 284 995 речень, а також приклади перекладу речень, що містять перекладені фрази («база пам'яті перекладів»). Сайт: <https://uk.glosbe.com>



Рис. 1.5. Результат пошуку тлумачення лема „Überfall“ в онлайн-словнику Glosbe

У Glosbe можна знайти переклади з німецького на українську з різних джерел. Переклади відсортовані від найпоширеніших до менш популярних. Можна перевірити не лише переклади німецькою або українською, а також проаналізувати приклади використання, які показують десятки перекладених речень. Так, можна побачити не лише переклад фрази, яку шукаєте, але й те, як вона перекладається залежно від контексту. Перекладені речення, знайдені в Glosbe, походять із паралельних корпусів (великих баз даних із перекладеними текстами).

У Glosbe є можливість прослухати аудіозаписи та якісні комп’ютерні читалки. Окрім перекладу тексту, у Glosbe містяться зображення, які представляють пошукові терміни.

Lingea на порталі dict.com – ще один корисний сервіс, що використовує сучасні технології та методи при створенні онлайн-словників, в яких усі словникові статті оброблені традиційно й так само, як у книжкових словниках.

На **Lingea** користувачі можуть оцінити такі переваги:

- ✓ надзвичайна якість даних можна порівняти з найліпшими книжковими словниками;
- ✓ детальні словники для 32 мов: у кожному словнику в одному напрямку понад 45 000 словникових стетей;

- ✓ словникові статті завжди пристосовані до рідної мови користувача: транскрипція вимови, контекстуальні примітки та морфологічні інформації;
- ✓ посередині головної панелі можна зобразити для головних європейських мов ще наступні пропозиції: розмовні практика, перегляд граматики, вживання слів та різні цікаві факти про мову;
- ✓ швидкий інтелегентний пошук: знайде слово наведену у будь-якій формі, запропонує правильну словникову статтю вже після наведення 3-ох букв, вмiє помітити друкарські помилки, знайде фрази та приклади з допомогою повнотекстового пошуку, зручний у використанні
- ✓ зрозуміло зобразить словникову статтю, синоніми, теми та результати повнотекстового пошуку
- ✓ містить в собі віртуальну клавіатуру зі знаками всіх потрібних алфавітів
- ✓ пропонує історію пошуку.

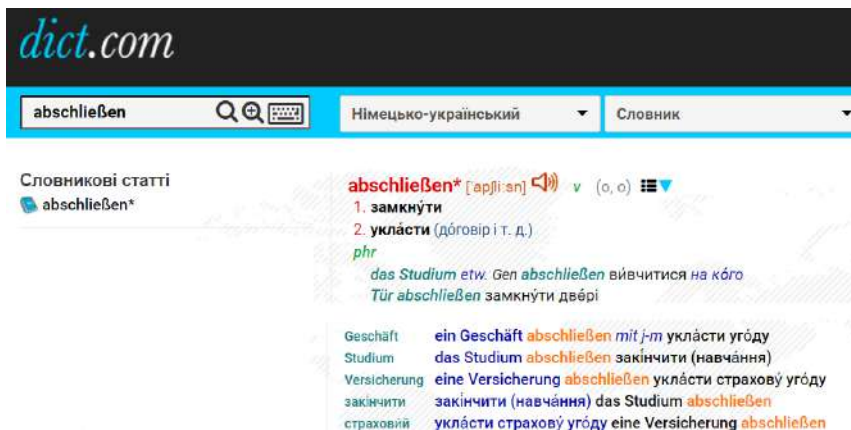


Рис. 1.6. Результат пошуку тлумачення лема *abschließen*“ в онлайн-словнику dict.com

Розглянемо можливості пошуку, які пропонує онлайн-словник:

1) пошук в обох напрямках: не треба вибрати напрямок перекладу, аплікація сама розпізнає мову, з якої наведене слово

походить;

2) морфологічний пошук: можна навести слова у будь-якої формі (рід, число, відмінок, час...) – це зручно передусім для початківців;

3) повнотекстовий пошук: під основною словниковою статтею зображаються наступні наявності наведеного слова, а саме у прикладах або словосполученнях;

4) підказувач: достатньо навести перші 3 букви слова та зобразиться пропозиція словникових статей, які ви, можливо, шукаєте;

5) виправлення друкарських помилок: якщо напишете слово неправильно, запропонує вам словник правильну можливість.

Для здійснення якісного перекладу з іноземної мови українською варто послуговуватися різними видами онлайн-словників сучасної української мови (тлумачним, орфографічним, фразеологічним, словником синонімів, антонімів, словозмін). Розглянемо, до прикладу кілька сучасних вітчизняних пропозицій.

Словник української мови.

Одинадцятитомний «Словник української мови» (або СУМ-11),

що його було видано у 1970-1980 роках, став першим у історії тлумачним словником української мови. На ньому базуються всі сучасні українські тлумачні словники, зокрема «Великий тлумачний словник сучасної української мови» (ВТССУМ) та оновлений «Словник української мови у 20 томах», публікацію якого розпочала Академія наук України у 2010 році. Сайт: <http://sum.in.ua>



Після більшості тлумачень у словнику наведено справжні випадки вживання слова, що ілюструють його значення.

Словник іншомовних слів містить:

- ✓ алфавітний покажчик;
- ✓ ілюстрації словникових ресурсів;
- ✓ тематичні словники;
- ✓ відомості про правопис іншомовних слів.

Можна переглянути статті словника в алфавітному порядку або знайти тлумачення певного терміну через форму запити.

Тлумачний словник іншомовних слів містить науково-популярне пояснення слів чужомовного походження з наведенням їх українського відповідника і синонімічного ряду.

Наразі у «Словнику» 20119 статей, які охоплюють термінологію зі сфер культури, науки, політики, економіки тощо. Усі словникові статті містять етимологічну довідку (див. список скорочень) і вказівку на сферу застосування чи вживання терміну [Амеліна; Тарасенко 2017, с. 164].

Суттєву допомогу перекладачеві під час здійснення перекладу можуть надати довідкові інтернет-ресурси, як-от **Glossary**, який надає можливість отримати тлумачення не тільки одного терміну, але й цілої предметної області, містить глосарій за темами. Представлено алфавітний перелік глосаріїв, тематичний перелік статей та тематичний навігатор. До основних тем цього довідкового ресурсу належать: Економіка; Життя; Управління; Суспільство; Власність; Фізичні величини; Документи; Хімічні з'єднання; Інформація; Планета Земля; Право [Амеліна; Тарасенко 2017, с. 165]. Сайт: <https://glossary.ua/ua/>

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 [Фокін 2019, с. 23].

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. 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. Гемінгвея, 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 украй необхідним. Немає потреби доводити, що в технічному перекладі наявність довідкового джерела з візуальними компонентами стане тільки плюсом для перекладача [Фокін 2019, с. 24].

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 *DeepL*, *Glosbe*, *Linguee*, *Reverso Context* та інші.

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 [Фокін 2019, с. 25]. Словники, побудовані на матеріалі двомовних паралельних текстів, такі як *DeepL*, *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 А перекладено як В, це не означає автоматично, що В буде завжди перекладатися як А, 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 леми через глосу. У випадку двомовних перекладних словників ідеться про пошук леми 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.

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 загальноживаної лексичної одиниці цільовою мовою, інколи справді доречно користуватися словниками, побудованими на основі паралельних текстів, таких як *DeepL*, *Glosbe*, *Linguee*, *Reverso Context* та інш., не говорячи вже про широке застосування одномовних словників (ономасіологічних, тезаурусних та інших) [Фокін 2019, с. 28].

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.

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 [Фокін 2019, с. 29].

На практиці двомовні перекладні словники часто й не розв'язують труднощів перекладу. Тому професіоналам перекладу у своїй діяльності частіше доводиться користуватися одномовними лексикографічними ресурсами.

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 мови, навряд чи збагатиметься об'єм семантичних ознак поняття (інтенціонал) та сукупності позначуваних денотатів (екстенціонал) вихідної та цільової лексем [Фокін 2019, с. 30]. До прикладу, 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 [Фокін 2019, с. 31].

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.

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 [Фокін 2019, с. 31].

Семасіологічні словники становлять переважну більшість лексикографічних ресурсів, тому, відповідно, визначник “семасіологічний” навіть не фігурує у їхніх назвах. Насамперед семасіологічні словники представлені тлумачними словниками.

Форма	Зміст
Лема	Глоса
adoptieren	als eigenes Kind annehmen BEISPIEL: <i>ein Kind adoptieren</i>

Рис 1.7. Структура семасіологічного словника

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”, станом на 2019 р. видано дев’ять томів, доступні також онлайн, в яких статті розташовані за алфавітним порядком до прикметника “мішурний” включно.

На сайті Українського мовно-інформаційного фонду міститься ресурс “Словники України”. Словник охоплює значну кількість лексем української мови, у тому числі власних назв (також 32 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.

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-довідник з культури української мови" [Гринчишин 1996], упорядкований Д. Гринчишином, А. Капелюшним, О. Сербенською, З. Терлак.

Подібним застосуванням характеризуються також *орфографічні* та *орфоепічні словники*. Безумовно, кожен словник має містити орфографічно правильні словоформи, що супроводжуються коментарями орфоепічних своєрідностей, але найчастіше це стосується початкової словникової форми, зазначеної у лемі, тоді як спеціалізований орфоепічний словник містить усі словоформи різних відмінків, відмін та інших граматичних категорій та їхніх форм [Фокін 2019, с. 35]. 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.

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; на лінгвістичному порталі *mona.info* є можливість згенерувати частотний словник слів 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 [Фокін 2019, с. 36].

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 Стронгом (грецькою мовою та івритом).

Конкорданс зазвичай складався зі списку слів, наявних у певного автора, що супроводжувалися посиланням на контекст вживання цього слова. У комп'ютерну еру обсяги опрацьовуваних текстів збільшилися, тому створення списку всіх слів, словосполучень було б непомірно великим, та й у цьому зазвичай немає потреби: лише для однієї лексики 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 (слів, 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 [Фокін 2019, с. 37]. Кожне слово у них представлено з мінімальним попереднім та подальшим контекстом.

Недолік конкордансу, у порівнянні зі словниками сполучуваності, в тому, що на запит користувача в ньому може з'явитися забагато зайвих даних, які користувачеві доводиться аналітично опрацювати 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 заледве одна словоформа з контекстом. При натисненні на гіперпосилання відкривається ширший контекст вживання заданого слова.

Konkordanz x +

Handbuch für Biblische Seelsorge : Wort in Gesamtes Buch

VERFEINERN

Häufigkeit 1 bis 1.367

Einträge

- ✓ Alle Wortformen kombinie...
- ✓ Ohne häufige Stoppwörter
- ✓ Ohne Zahlen

Sprache

- Deutsch 7.635
- Englisch 25
- Transkribiert 2
- Latein 1

Textebene

- Fließtext 6.732
- Fußnotentext 1.115
- Überschrift-Text 648
- Großer Text 188
- Titel 2
- Tag 1

Wort 53.932 Eintrag Häufigkeit

- * Bedeutsamkeit bedeutsam; bedeutsamer bedeutsam; bedeutsamer 23
- * Bereich Bereiche; Bereichen; Bereiches Bereiche; Bereichen; Bereiches 23
- * Bitterkeit 23**

S 22 vergibst. Eine Haltung der Bitterkeit würde nicht nur dir selber...

S 26 führt zu einer Haltung der Bitterkeit, die der Beziehung zu Gott...

S 26 nicht irgendeine Wurzel der Bitterkeit aufspresse und euch zur...

S 26 führt zu Depressionen und Bitterkeit. Wird dieser Zorn nicht...

S 81 an Selbstmitleid, Zorn und Bitterkeit festhält, wenn dir...

S 83 die ihm zugrunde liegende Bitterkeit aufgedeckt werden kann...

S 83 aufgedeckt werden kann. Bitterkeit ist eine Hauptursache für...

S 83 wir uns im Zustand der Bitterkeit alleine und verlassen...

S 83 Bibel: EPHESER 4,31 Alle Bitterkeit und Wut und Zorn und...

S 84 ansonsten werden Zorn und Bitterkeit uns und den Menschen, die...

S 84 nicht irgendeine Wurzel der Bitterkeit aufspresse und euch zur...

S 92 bedeutet, sich ohne Bitterkeit, Hass oder Groll an...

S 163 wir unsere Bitterkeit durch Hoffnung überwinden können...

Рис. 1.8. Приклад оформлення конкордансу, згенерованого комп'ютером

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 [Фокін 2019, с. 38].

1.5. Особливості структури і використання ономазіологічних словників

Переважає 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. “σημασία” – “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 думці. 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.

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

Рис.1.9. Структура ономасіологічного словника

Проблеми пошуку відповідних засобів вираження трапляються не лише в перекладі: носії певної мови час від часу почуваються 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 пошук вдалого рішення може стати справді виснажливим [Фокін 2019, с. 45].

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.

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 [Фокін 2019, с. 46]. Саме це і зумовлює відсутність єдиної універсальної схеми для їхнього впорядкування, а також складність їхнього пошуку у схемі.

Розуміння поняття ономазіологічності 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 ономазіологічного словника. У зв'язку з цим вважаємо за необхідне пояснити наявні між ними розбіжності у наступних підпунктах [Фокін 2019, с. 47].

1.5.1. Тематичний словник

Найпростішим і найяскравішим прикладом словника, що за суттю можна вважати ономазіологічним, 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 укладати, адже створена основа для однієї пари мов може слугувати й для інших.

Однак практика показує, що професіонали такими словниками не користуються, а початківцям, які охоче 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 би 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 лексему [Фокін 2019, с. 48].

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.

1.5.2. 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.

1.5.3. 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. Інколи тезауруси також іменують словниками споріднених ідей [Фокін 2019, с. 49-50].

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 величину. До прикладу, саме так укладається тезаурус у Лейпцизькій колекції корпусів:

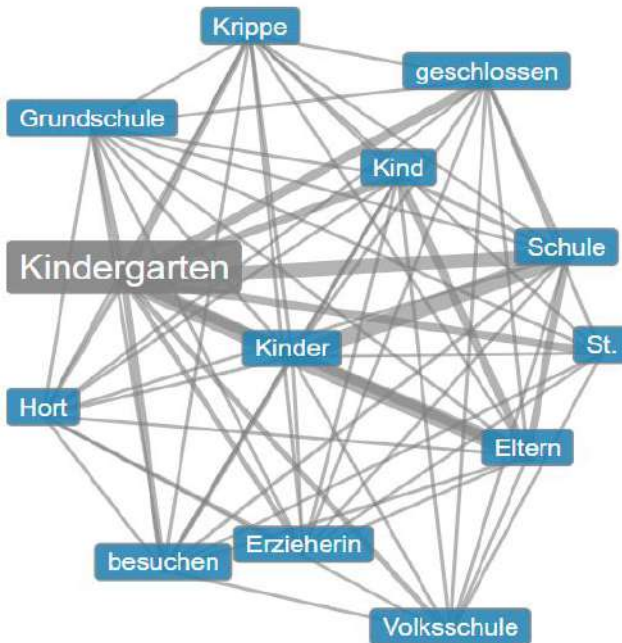


Рис. 1.10. Приклад візуального тезаурусу до реєстрового слова „Kindergarten“ (<http://corpora.uni-leipzig.de>)

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-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. 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 теґу семантичного відношення – операція вкрай делікатна; окремі типи відношень (квазісинонімія, на відміну від синонімії та антонімії) виявляти важко”;

– 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. [Фокін 2019, с. 50-53].

1.5.4. 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 уможлиблює 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].

Зміст	Форма
Глоса	Лема
männliche Person, die [berufsmäßig] mündlich übersetzt (Berufsbezeichnung)	Dolmetscher, der

Рис. 1.11. Структура інверсійного словника

Серед доступних та популярних серед користувачів є, 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) Сайт: <https://www.onelook.com/reverse-dictionary.shtml>;

2) “Reverse Dictionary” характеризується надзвичайно простим інтерфейсом без додаткових опцій, однак він є доволі популярним серед користувачів [Reverse] Сайт: <https://reversedictionary.org>.

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

Такі словники теж слід вважати ономасіологічними, 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 пов’язані з дефініціями [Фокін 2019, с. 54].

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 [Фокін 2019, с. 55-56].

Використана література та інформаційні ресурси:

1. Амеліна С. М., Тарасенко Р. О. Підготовка перекладачів у зарубіжних університетах до застосування інформаційних технологій: монографія. Київ : «Центр учбової літератури», 2017. 256 с.

2. Амеліна С.М. Особливості інформаційно-технологічної підготовки майбутніх перекладачів у вищих навчальних закладах Німеччини / С.М.Амеліна, Р.О.Тарасенко // Вища школа. 2014. №8. С. 94-100.

3. Долинський Є.В., Скиба К.М. Інформаційні технології у професійній діяльності перекладача : навчальний посібник з інформаційних технологій для студентів 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. 195 с.

4. 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.

5. Фокін С.Б. Комп’ютерні інформаційні технології в перекладацькій та перекладознавчій діяльності. К.: Київ. нац. ун-т ім. Т.Шевченка, 2019. 254 с.

6. Словник української мови (СУМ). Академічний тлумачний словник. – [Електронний документ]. Режим доступу: <http://sum.in.ua/>

7. Словник фразеологічних синонімів. – [Електронний документ]. Режим доступу: <http://www.rozum.org.ua/index.php?a=term&d=24&t=546>

8. 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
9. British Parliament “Hansard” Corpus – [Електронний ресурс]. Режим доступу: <https://www.hansard-corpus.org/semTags1.asp>
10. Corpora Collection of Leipzig University. – [Електронний документ]. Режим доступу: <http://corpora.uni-leipzig.de>
11. Merriam-Webster Thesaurus. – [Електронний документ]. Режим доступу: <https://www.merriam-webster.com/thesaurus>
12. One Look Reverse Dictionary. – [Електронний документ]. Режим доступу: <https://www.onelook.com/reverse-dictionary.shtml>
13. Oxford English Dictionary. – [Електронний документ]. Режим доступу: <http://www.oed.com/>
14. Reverse Dictionary – [Електронний документ]. Режим доступу: <http://reversedictionary.org/>
15. Wordnet. A lexical database for English. – [Електронний документ]. Режим доступу: <https://wordnet.princeton.edu>

СЕМІНАРСЬКЕ ЗАНЯТТЯ 1

Теоретичні питання для обговорення:

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.
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, 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.

10. Чим відрізняється макроструктура від мікроструктури словника?

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

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

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

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

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

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

17. Які існують різновиди двомовних перекладних словників?

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

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

20. Назвіть приклади семасіологічних словників.

21. Назвіть щонайменше 7 критеріїв класифікації словників.

22. У чому відмінності між ономасіологічним і семасіологічним словником?

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

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

25. Як організована макроструктура ономасіологічного словника?

26. У чому причини недоліків двомовних перекладних словників?

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

28. За яким принципом розташовуються лексеми у тезаурусі?

29. В яких значеннях вживається термін “атлас лексики”?
30. Як називається словник, в якому можна проводити пошук леми за дефініцією?
31. Назвіть приклади словників, автоматично згенерованих комп’ютером. Яке можливе їхнє застосування?
32. Які функції словників, недоречні чи неможливі в паперовому форматі, цілком прийнятні в електронних словниках?

Практичні завдання:

1. Підготуйте доповідь з презентацією до теми *«Пошук інформації в мережі Інтернет. Інформаційно-пошукові системи»*. Продемонструйте можливості інформаційно-пошукових систем на конкретних прикладах шляхи вирішення перекладацьких проблем у професійній діяльності перекладача.

Рекомендована література: Долинський Є.В., Скиба К.М. Інформаційні технології у професійній діяльності перекладача : навчальний посібник з інформаційних технологій для студентів 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. 195 с.

2. Підготуйте доповідь з презентацією до теми *«Характеристики термінологічних словників. Методи пошуку перекладу терміна за допомогою комп’ютерних технологій»*. Продемонструйте покроково процес пошуку перекладу терміна з німецької мови українською на власному прикладі.

Рекомендована література: Фокін С.Б. Комп’ютерні інформаційні технології в перекладацькій та перекладознавчій діяльності. К.: Київ. нац. ун-т ім. Т.Шевченка, 2019. 254 с.

Лабораторна робота 1:

Для засвоєння навичок роботи із системою електронних словників, зробіть переклад німецького тексту українською мовою (за власним вибором), використовуючі популярні інтернет-словники. *Форма роботи:* командна.

Інструкція до виконання: Кожна група студентів працює над одним і тим самим текстом певного жанру, але за допомогою різних онлайн-ресурсів. Після виконання завдання – обговорення та

порівняння якості перекладу.

Група студентів А: готує переклад наукового тексту обсягом 150-200 слів (1000-1300 знаків з пробілами);

Група студентів Б: переклад художнього тексту обсягом 150-200 слів (1000-1300 знаків з пробілами);

Група студентів В: переклад публіцистичного тексту обсягом 150-200 слів (1000-1300 знаків з пробілами).

ТЕМА 2

Корпусні ресурси в діяльності і підготовці перекладача. Перевірка якості перекладу та функції текстових редакторів

План лекції:

- 2.1. Електронні лінгвістичні корпуси як довідковий ресурс перекладача
- 2.2. Корпуси паралельних текстів
- 2.3. Основні принципи пошуку в корпусах текстів

2.1. Електронні лінгвістичні корпуси як довідковий ресурс перекладача

На окрему увагу серед довідкових ресурсів сучасного перекладача заслуговують електронні лінгвістичні корпуси.

З розвитком комп'ютерних технологій розвивається така наука, як корпусна лінгвістика. Корпуси є, власне, інструментом дослідження мовного матеріалу й моніторингу мовних явищ. Сьогодні корпусна лінгвістика й корпуси мовних даних відіграють важливу роль у навчанні та вивченні мов [Амеліна, Тарасенко 2017, с. 166].

Електронні лінгвістичні корпуси – це зібрання машиночитаних текстів, відібраних таким чином, щоб максимально репрезентувати мову та її різноманіття; велика кількість текстів природної мови, що мають комп'ютерну форму і є об'єктом певного лінгвістичного дослідження, де під «природним» розуміють усе, що фактично було висловлено в усній або писемній формі; зібрання відібраних і впорядкованих текстових уривків, що їх використовують як взірць мови.

Незважаючи на існування різних варіантів визначення,

більшістю корпусних лінгвістів за визначальні ознаки корпусу приймаються такі характеристики:

- ✓ машиночитаність,
- ✓ автентичність текстів (включаючи усні транскрибовані),
- ✓ добірність,
- ✓ репрезентативність [Амеліна, Тарасенко 2017, с. 167].

У вузькому значенні під корпусом розуміють зібрання текстів в електронній формі, що презентує певну мову. При такому розумінні апріорною детермінативною ознакою корпусу текстів є машиночитаність.

Першим з електронних корпусів текстів є *Браунівський корпус* (1962–1963 рр.), створений У.Френсисом і Г. Куцурою. Перша версія Браунівського корпусу була представлена простим текстовим форматом (із невеликою кількістю структурної розмітки для виділення абзаців, заголовків, цитованих фрагментів тощо). Пізніше корпус був доповнений розміткою частин мови й морфологічних ознак слів. Загальний обсяг корпусу – 1 млн. слів. Згодом це число й загалом структура побудови корпусу стала стандартом для створення аналогічних корпусів текстів. Багатьма лінгвістами визнається тільки електронна форма існування корпусу, стандартність та можливість його багатократного використання.

Отже, лінгвістичний корпус – це певним чином організоване комп'ютерне зібрання писемних і/або усних текстів будь-якої національної мови, якому притаманні обов'язкові ознаки і який призначений для наукового та практичного вивчення цієї мови [Амеліна, Тарасенко 2017, с. 167-168].

Значення лінгвістичних корпусів полягає в тому, що:

- одного разу зроблений корпус може багато разів використовуватися;
- корпус показує мовні дані в їх реальному оточенні;
- корпус має важливе значення для перекладу, викладання, дослідження і живання мови;
- корпус характеризується показністю, або збалансованим складом текстів, що дозволяє використовувати його для тестування пошукових машин, машинних морфологій, систем перекладу, а також використовувати його в різних лінгвістичних дослідженнях.

Корпуси можна використовувати для отримання різноманітних довідок і статистичних даних про мовні й мовленнєві

одиниці. Прикладами найбільш відомих лінгвістичних корпусів є:

- Brown Corpus;
- Ланкастерський корпус англійської мови (Lancaster Oslo-Bergen Corpus, LOB);
- British National Corpus;
- International Corpus of English;
- Bank of English;
- Cobuild Corpus;
- *Мангеймський корпус німецької мови*;
- Чеський національний корпус;
- корпуси китайської, турецької, естонської, албанської і багатьох інших мов.

За обсягом ці корпуси є різними. Наприклад, *Британський національний корпус* нараховує 100 млн. слововживань, Корпус австралійської періодики – 300 млн., Корпус чеської мови – 300 млн., Корпус німецької мови – 778 млн. слововживань [Амеліна, Тарасенко 2017, с. 169].

Призначення корпусу – показати функціонування лінгвістичних одиниць в їх природньому контекстному оточенні. На основі корпусу можна одержати дані про:

- частоту словоформ, лексем, граматичних категорій;
- зміну частотності вживання конкретного слова;
- зміни контекстів у різні періоди часу;
- поведінку мовних одиниць різних авторів;
- особливості їх сполучуваності, керування тощо.

Щоб довільне зібрання текстів могло називатися корпусом, воно повинно мати певні важливі характеристики. По-перше, мати електронну форму і бути стандартно організованим, тобто відповідати вимогам міжнародного стандарту кодування корпусу. По-друге, бути репрезентативним і збалансованим. По-третє, бути вичерпним; у корпусі ідеально має бути не забагато і не замало певних текстів або їх уривків.

Найчастіше основними, визначальними ознаками текстового корпусу вважають машиночитаність чи наявність, по-перше, електронної форми подання і, по-друге, спеціальної системи кодування текстових даних, та репрезентативність, чи здатність корпусу правильно відображати мову або її частину [Амеліна, Тарасенко 2017, с. 171].

Електронні корпуси текстів української мови – безкоштовні, можна заходити на їхній сайт і здійснювати пошук без реєстрації.

1. Корпус української мови – це електронне зібрання текстів, упорядковане, організоване й оформлене так, щоб ним мали змогу користуватися як науковці-професіонали, так і пересічні користувачі, які намагаються вдосконалити своє знання української мови (<http://korpus.org.ua>). Тексти призначені не для читання, а для з'ясування питань «Яке слово краще вжити у тому чи іншому випадку?», «Яке словосполучення буде більш доречним?», «Як часто у текстах використовується та чи інша словоформа?».

Корпус української мови *

<input type="text"/>	Шукати		
5	цитат на текст по	450	символів

Про пошук

- Пошук здійснюється дослівно (без лематизації, стемінгу та виправлення помилок).
- Для пошуку словоформ та гнізд можна використовувати шаблонн * (замість будь-якої комбінації літер) та ? (замість однієї літери). Наприклад, запит *гідн?ст** відповідатиме як слову *гідність*, так і слову *гідностей*.

* Загальномовний (або національний) неанотований та несистематизований корпус української мови. Містить 6,6 Гб україномовних текстів з електронної бібліотеки [Чтиво](#). Відповідно, не всі тексти є достатньо добре вичитаними та якісними, щоб відповідати вимогам корпусної лінгвістики. Радимо ставитися з пильністю та критичним мисленням до результатів пошуку, а надто — у спірних ситуаціях. Додаткове консультування із паперовим першоджерелом чи сканом (у багатьох випадках наявного через посилання) буде необхідним для будь-якого серйозного дослідження.

Рис. 2.1. Домашня сторінка корпусу української мови

Корпус містить близько 50 млн. слововживань. Підкорпуси становлять: публіцистики – 17 млн. слововживань; художньої прози – 24 млн.; поезії – 1 млн.; наукового стилю – 3 млн.; офіційно-ділового – 2 млн., фольклору – 32 тис.

Кожен уведений текст маркується за жанровими особливостями й типом тексту.

Корпус текстів може бути джерелом для укладання різних словників, зокрема: алфавітно-частотних – слів і словоформ; граматичних – з морфологічною або синтаксичною інформацією, напр., про синтаксичні моделі – дієслівні, субстантивні, атрибутивні; морфемних і словотвірних – із частотними

характеристиками, за якими можна вивчати комбінаторно-дистрибутивну будову слів або словотвірне значення кожної афіксальної морфеми в текстах; тезаурусів; словників неолексем [Амеліна, Тарасенко 2017, с. 174].

Це можливо, оскільки у базі для цього передбачено відповідні параметри: граматичні (частина мови і категоріальні значення, напр., рід, число, відмінок, особа тощо); структурні (моделі морфної структури слів різних частин мови; моделі керування – дієслівні, іменні, атрибутивні тощо); лексико-семантичні (системні відношення: синонімія, антонімія, омонімія, ідеографія); статистичні.

2. Корпус української мови на лінгвістичному порталі MOVA.info - тестова версія корпусу текстів української мови, розроблена співробітниками лабораторії комп'ютерної лінгвістики Інституту філології Київського національного університету імені Тараса Шевченка Н.П.Дарчук (керівник проекту), В.М.Сорокіним (програміст), О.Б.Сірук, Я.В.Ходаківською, Н.Г.Чейлитко, М.О.Лангенбах. Сайт: <http://www.mova.info>

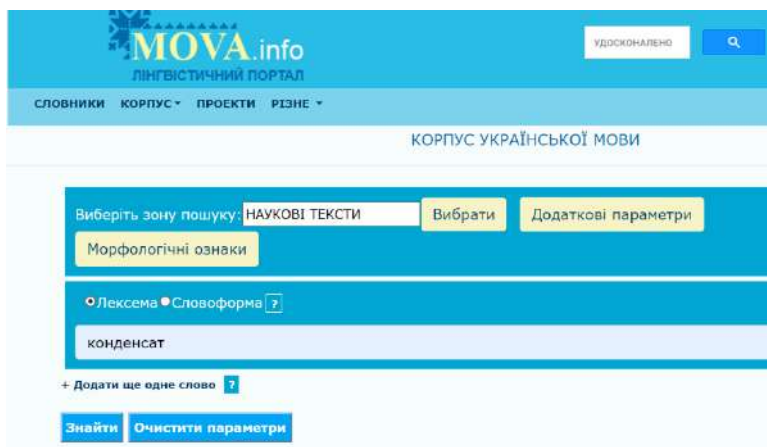


Рис. 2.2. Домашня сторінка корпусу української мови MOVA.info

Корпус містить млн. слововживань з різних галузей: законодавчі тексти (2 737 766), наукові тексти (8 737 464), поетична мова (801 267), публіцистика (46 966 935), фольклорні тексти (1377

912), художня проза (40 748 676). Переваги: простий інтерфейс; Недоліки: низька глибина контексту (максимально 7-20 слів); невеликий вибір параметрів додаткового пошуку.

Існує можливість здійснювати пошук у синтаксичному підкорпусі на цьому ж ресурсі.

СИНТАКСИЧНИЙ ПІДКОРПУС ?

Виберіть зону пошуку:

1. Лексична модель словосполучення

Головне слово *	Залежне слово
(початкова форма)	(початкова форма)
<input type="text"/>	<input type="text"/>

* у цьому полі необхідно записати те слово, яке виступає у слові-головині, незалежно від його позиції

2. Морфологічні ознаки

Частина мови (голове)	Частина мови (залежна)
Морфологічна характеристика	Морфологічна характеристика
<input type="text"/>	<input type="text"/>

3. Синтаксичні ознаки

Тип зв'язку

Рис. 2.3. Синтаксичний підкорпус корпусу української мови MOVA.info

Генеральний регіонально анотований корпус української мови (ГРАК) – велика, репрезентативна, структурована колекція текстів українською мовою у супроводі програми, яка дозволяє будувати на базі корпусу власні підкорпуси, шукати слова, граматичні форми та їх сполучення, а також обробляти результати пошуку, сортувати, робити збалансовані вибірки і одержувати різну статистичну інформацію. Корпус охоплює період з 1816 по 2020 р. і містить понад 90 тисяч текстів різних жанрів, близько 26 тисяч авторів. Розробники: М. Шведова, Р. фон Вальденфельс, С. Яригін, А. Рисін, В. Старко, Т. Ніколаєнко та ін. Київ, Львів, Єна, 2017-2022. uacorpus.org. Сайт: <http://uacorpus.org/Kyiv/ua>



ІНФОРМАЦІЯ ПРО КОРПУС - РОЗМІТКА ТЕКСТІВ - ПОШУК У ГРАКУ - ДОГЛЯД ЗА КОРПУСОМ - ГРАКУ НА FACEBOOK

ДОСЛІДЖЕННЯ НА ОСНОВІ ГРАКУ - ІНШІ КОРПУСИ УКРАЇНСЬКОЇ МОВИ ТА СЛОВ'ЯНСЬКИХ МОВ

ВІДКРИТІ РЕСУРСИ ТА ІНСТРУМЕНТИ ДЛЯ РОБОТИ З УКРАЇНСЬКОЮ МОВОЮ - СЛОВНИКИ - ГЛОСАРІЙ ТЕРМІНІВ - ГАЗЕТИ 1945-2000

Українська -

ГРАК

Генеральний регіонально анотований корпус української мови (ГРАК) – це велика, репрезентативна, структурована колекція текстів українською мовою у супроводі програми, яка дозволяє будувати на базі корпусу власні підкорпуси, шукати слова, граматичні форми та їх сполучення, а також обробляти результати пошуку, сортувати, робити збалансовані вибірки і одержувати рану статистичну інформацію.

Рис. 2.4. Домашня сторінка корпусу української мови ГРАК

Корпус призначений для здійснення лінгвістичних досліджень з граматики, лексики, історії української літературної мови, а також для використання під час укладання словників та граматик. Корпус може бути використаний для вивчення мови на високому рівні, також при підготовці навчальних матеріалів, підручників, навчальних словників та вправ з використанням прикладів з реальних текстів, з урахуванням частотної сполучуваності тощо. Корпус не є взірцем нормативної української мови, в ньому можуть трапитись слова і сполучення, які не відповідають сучасним нормам літературної мови.

Лексикографічні проекти європейських університетів. Цінним ресурсом для перекладачів є лексикографічні бази, створені провідними європейськими університетами. Одним з таких ресурсів є так званий корпус Коллінз, який містить, зокрема: словники; тезауруси; програму-перекладач. Слід зазначити, що у цьому проекті представлена українська мова.

Академія наук Берлін-Бранденбург (Die Berlin-Brandenburgische Akademie der Wissenschaften) розробила цифровий словник німецької мови, який є інформаційною словниковою системою німецької мови, що відображає її історію й сучасний стан.

Для створення цієї бази використано понад 410 000 лексикографічних статей з п'яти словникових джерел. Її обсяг є досить великим – 1,8 млрд. корпусних документів з 15 корпусів, які

охоплюють статистичні профілі слів і траєкторії їх розвитку. Зазначена база надає можливість безкоштовної перевірки правопису (можна надіслати до 50 000 знаків). Сайт: <https://www.zdl.org>

Словниковий портал Лейпцигського університету (Wortschatz-Portal der Universität Leipzig) містить такі дані:

- Слова дня: актуальні поняття. Обираються зі щоденних газет та інформаційних служб. Розміщуються щоденно о 7:00.

- Більше 230 корпусів 200 мовами. Портал постійно розвивається і розширюється, включаючи нові ресурси.

- Міжнародний портал лексики англійською мовою (дозволяє пошук слів у різних мовах у 229 корпусах).

- Більше 100.000 слів і фраз англійською та німецькою мовами (з частотою вживання).



Рис. 2.5. Домашня сторінка Словникового порталу Лейпцигського університету

Web-сервіси дозволяють прямий доступ до даних проекту з будь-якого програмного забезпечення. Сайт: https://corpora.uni-leipzig.de/de?corpusId=deu_newsrawl_2011

На особливу увагу заслуговують лексикографічні ресурси Інституту німецької мови (Institut für Deutsche Sprache), які об'єднують кілька лінгвістичних корпусів.

Одним із них є **Корпус німецької стандартної мови**, який містить 43 фрагменти звукозапису, починаючи з 1975 року, загальною тривалістю 5 годин 48 хвилин. Це записи 43 студентів і вчених (жінок і чоловіків) у віці від 17 до 27 років з 43 рівномірно розподілених місць у старих федеральних землях. Мовці народилися і виростили в цих обраних місцях. У записах, доступних для використання сторонніми особами, можна почути мову інформаторів, які читають розділ з Основного закону. Записи були оцифровані в Архіві розмовної німецької мови (Archiv für Gesprochenes Deutsch (AGD)).



Рис. 2.6. Домашня сторінка Архіву розмовної німецької мови

Записи були оцифровані в Архіві розмовної німецької мови, а на основі метаданих був створений список тем. Записи транскрибуються, транскрипти синхронізовані (вирівняні) зі звуком. У корпусі містяться також глосарії до них (слова і лемми) [Амеліна, Тарасенко 2017, с. 176]. Сайт: http://agd.ids-mannheim.de/korpus_index.shtml

Корисним для перекладача-практика може бути **Корпус німецьких діалектів** (Korpus Deutsche Mundarten), створений на основі даних дослідження діалектичного розшарування та розпаду діалекту. Сайт: http://agd.ids-mannheim.de/ZW--_extern.shtml

Корпус містить 73 звукозаписи з 1965 по 1967 роки загальною

тривалістю 42 годин 28 хвилин. Це запис історій, розмов та читання носіїв діалектів, жінок та чоловіків різного віку з округу Бьоблінген.

Корпус вітчизняних та зарубіжних діалектів складено із заархівованих записів зарубіжних учених в Архіві німецької мови. Доступними для використання є 183 звукозаписи з періоду 1965-1974 рр. загальною тривалістю 69 годин 25 хвилин. Це записи 184 мовців (жінок і чоловіків) з Австралії, Німеччини, Канади, Мексики, Австрії, Швейцарії та США в різних мовних подіях. Зокрема, це розповіді, переклади та стандартні тексти (контрольні речення, список слів, цифри, дні тижня, місяці). Записи були оцифровані в Архіві розмовної німецької мови (раніше – Німецький мовний архів). Транскрипти 14 записів з Австралії, Канади та США були опубліковані.

Інститут німецької мови (IDS) здійснює, зокрема, розширення та підтримку сучасної письмової німецької мови.

Німецьким довідковим корпусом є DeReKo, найбільше у світі зібрання німецької корпорації як емпірична основа для лінгвістичних досліджень. Сайт: <https://www.ids-mannheim.de/digspra/kl/projekte/korpora>

Корпуси сучасної письмової німецької мови Інституту німецької мови характеризуються таким чином:

- налічують 32,83 млрд. слів (станом на 10.10.2017 р.) і є найбільшим у світі лінгвістичним зібранням електронних корпусів з письмовими німецькими текстами сьогодення та недавнього минулого;

- доступні безкоштовно через COSMAS II;

- містять художні, наукові та науково-популярні тексти, велику кількість газетних текстів та широкий спектр інших типів текстів і постійно поповнюються;

- створені з урахуванням обсягу, мінливості, якості та своєчасності, а на етапі використання COSMAS II дозволяють формувати віртуальні корпуси, які є репрезентативними або пристосованими до конкретних завдань;

- містять лише захищені авторським правом матеріали [Амеліна, Тарасенко 2017, с. 177].

Корпуси текстів Інституту німецької мови (IDS) є цінною емпіричною базою для досліджень, і не тільки лінгвістичних, у самому Інституті німецької мови, а також для досліджень

германістів у національному й міжнародному масштабі, що підтверджує постійно зростаюча кількість онлайн-користувачів.

Вони частіше використовуються для міждисциплінарної тематики, наприклад, у сферах психології, неврології, когнітивної науки, логотерапії, комунікації, медіа та статистики.

Найголовнішим принципом вибору нових текстів для розширення DeReKo, відповідно до його концепції як оригінального зразка використання письмової мови та дуже великого корпусу загального призначення, є максимізація масштабу та розподілу. На практиці, однак, до вибору були включені інші критерії, а саме, зокрема, витрати на отримання необхідних прав використання (авторське право) та витрати на обробку вихідних даних (конвертація).

Aktuelles Forschung Onlineangebote Service Publikationen Organisation Über uns IDS LEHRE-INSITUT FÜR DEUTSCHE SPRACHE

Suchbegriff eingeben

- ▶ Akquisition
- ▶ IDS-Textmodell
- ▶ Archiv
- ▶ Ausgaben
- ▶ Verfügbarkeit
- ▶ Textgeber
- ▶ Ehemalige Mitarbeiter
- ▶ Kooperationen
- ▶ Bibliographie IDS
- ▶ Bibliographie

Die weltweit größte Sammlung deutschsprachiger Korpora als empirische Basis für die linguistische Forschung

Das Deutsche Referenzkorpus – DeReKo

Die Korpora geschriebener Gegenwartssprache des IDS

- ▶ bilden mit 50,6 Milliarden Wörtern (Stand 02.02.2021) die weltweit größte linguistisch motivierte Sammlung elektronischer Korpora mit geschriebenen deutschsprachigen Texten aus der Gegenwart und der neueren Vergangenheit.
- ▶ sind über *COSMAS II* und *KorAP* kostenlos abfragbar
- ▶ enthalten belletristische, wissenschaftliche und populärwissenschaftliche Texte, eine große Zahl von Zeitungstexten sowie eine breite Palette weiterer Textarten und werden kontinuierlich weiterentwickelt.
- ▶ werden im Hinblick auf Umfang, Variabilität, Qualität und Aktualität akquiriert und erlauben in der Nutzungsphase über *COSMAS II* und v.a. *KorAP* die Komposition.

Рис. 2.7. Домашня сторінка Німецького довідкового корпусу DeReKo

Корпуси письмової сучасної мови повинні документувати фактичне вживання німецької мови і продовжувати вести це документування постійно, бажано щоденно. Це означає, що штучно створені тексти і веб-сторінки не розглядаються як можливі джерела, оскільки вони є дуже специфічним зрізом мови. Завданням проекту натомість є отримання підходящих, по можливості електронних матеріалів, які є автентичними документами вживання німецької мови. Германістам, дослідникам та перекладачам ці

матеріали потрібні для вивчення мови емпіричним шляхом і, відповідно, використання. У зв'язку з цим вважається само собою зрозумілим, навіть чесно, що автори надають свої матеріали для використання [Амеліна, Тарасенко 2017, с. 178]. Проте власників авторських прав часто відлякує побоювання втратити контроль щодо протиправного розповсюдження тексту після передачі його для корпусів. Крім того, більшість документів є не дуже придатною для використання з технічних причин. Автори текстів рідко хочуть або можуть підготувати свої джерела у форматі, який би дозволяв легко конвертувати їх у формат корпусу. З іншого боку, Інститут німецької мови (IDS) не має таких потужностей, щоб можна було взяти на себе необхідну затратну підготовку текстів. Через ці умови юридичні та фінансові переговори, як правило, є дуже складними. Під час створення корпусів відбувається також фільтрування текстів.

Газети, як складають головним чином вміст корпусів, містять також багато матеріалів, не цікавих з точки зору лінгвістики – спортивні турнірні таблиці, біржові курси, оголошення про різні розважальні заходи. Щоб уникнути появи цих небажаних матеріалів, здійснюється відповідна специфікація даних. Документи, які потрапляють до цього виду, відсортовуються.

Що ж в цілому до наповнення і обсягу корпусів, то варто зазначити, що Інститут німецької мови розпочав створення електронних корпусів ще у 60-ті роки. За період з 1992 року по 2015 рік обсяг корпусів збільшився приблизно з 28 мільйонів до 28 мільярдів текстових слів, що відповідає близько 70 мільйонам книжкових сторінок, якщо рахувати у середньому 400 слів на сторінку. Зрозуміло, що у створенні такого великого зібрання було задіяно багато співробітників. Архів корпусу постійно розширюється. Крім того, ведеться робота щодо обробки наявного матеріалу під кутом зору менеджменту якості. Однак, варто також зауважити, що з міркувань авторського і ліцензійного права певна частина заархівованих корпусів є недоступною ззовні, тобто за межами Інституту німецької мови. В цілому задекларовано, що корпуси IDS можуть використовуватись тільки з некомерційними, зокрема, науковими цілями. З цією метою на сайті розміщено відповідні інструкції [Амеліна, Тарасенко 2017, с. 184].

2.2. Корпуси паралельних текстів

У своїй більшості сучасні корпуси текстів одномовні, адже в цифровому форматі перекладів існує значно менше, ніж оригінальних текстів. Переклади, виконані на особисті замовлення, не публікуються і не підлягають розголошенню. Отже, далеко не завжди існує можливість укласти корпус паралельних текстів двома або більше мовами. Тим не менш, корпуси паралельних текстів теж активно розробляються, і варто розглянути їхні характеристики та специфіку використання [Фокін 2019].

Про успіхи в галузі укладання корпусів перекладів текстів свідчать такі популярні ресурси як **Glosbe** [Glosbe] або **Linguee** [Contextual Search Engine Linguee], **Reverso** [Reverso], відомі усім сучасним перекладачам. Звичайно, у цих корпусах містяться мовні пари, у рамках яких здійснюється велика кількість перекладів. Продемонструємо на Рис. 13 зразок пошуку німецького слова «*Impfstoff*» у двомовних паралельних текстах за допомогою онлайн-ресурсу Glosbe



Рис. 2.8. Результат пошуку лєми «*Impfstoff*» у паралельних корпусах GLOSBE

Зазначені корпуси складаються з двох колонок і пошукового рядка. Під час пошуку програма знаходить повні та часткові збіги з запитом користувача у лівій колонці, а також імовірні відповідники у правій колонці іншою мовою.

Завдяки такому пошуку можна знайти переклад тих слів (а що ще корисніше – словосполучень), які часто зустрічаються і перекладаються. У цих корпусах можна знаходити як фразеологічні одиниці, усталені та термінологічні сполучення, так і просто

механічне поєднання слів, що ідуть поспіль і можуть мати відповідник в іншій мові. Найціннішим у таких корпусах є можливість пошуку варіантів перекладу для лексем і словосполучень не лише в початковій словниковій формі, а в особовій/відмінковій формах (так званій “інтегрованої формі”), у складі конструкцій, в оточенні інших лексем. Такий підхід став справжнім проривом в укладанні довідкових перекладацьких ресурсів: класичні словники не здатні вмістити усі можливі сполучення слів, до того ж, у різних парадигматичних формах. Тим паче, що на вибір перекладацького відповідника впливає не лише семантика лексем, а й їхні граматичні характеристики.

Кількість словосполучень певної мови практично неосяжна. Окрім того, що теоретично можлива кількість сполучень слів рахується астрономічно невідомими обсягами, які обчислюються функцією факторіалу, про вичерпне додавання різних парадигматичних форм до двомовних словників не може бути і мови. На практиці ж сполучаються між собою далеко не всі лексеми й не всі парадигматичні форми, і саме в цьому сенсі величезна перевага корпусів з перекладеними текстами – наявність багатьох або навіть більшості узуальних сполучень слів та варіантів їхнього перекладу. Механічні поєднання тих чи інших графічних слів (що іменують біграмами, триграмами, n-грамами) програма виокремлює з текстів автоматично й так само автоматично проводить пошук сполучень слів у правій колонці, що, імовірно, відповідно до розташування у контексті відповідають за значенням сполученню слів у лівій колонці.

Описані вище корпуси, в яких можна шукати перекладні відповідники для слів і словосполучень, слід назвати перекладацькими паралельними корпусами. Вони зазвичай менш відомі, ніж перекладацькі корпуси, далеко не всі з них є у відкритому доступі [Фокін 2019].

На особливу увагу заслуговують бази інформаційних ресурсів, які за способом, формою та форматами представлення і зберігання текстових матеріалів максимально адаптовані для безпосереднього їх використання у системах автоматизованого перекладу [Амеліна; Тарасеко 2017, с. 153]. Останнім часом на створення комплексу таких баз спрямовує свої зусилля низка структур Єврокомісії (ЕК), зокрема:

- ✓ Європейський парламент (European Parliament);

✓ Об'єднаний дослідницький центр Генерального директорату Європейської комісії (Joint Research Centre of Directorate General of the European Commission (JRC));

✓ Генеральний директорат з перекладу (Directorate-General for Translation (DGT));

✓ агенція «Європейський центр з профілактики і контролю захворювань» (agency «European Centre for Disease Prevention and Control» (ECDC));

✓ Генеральний директорат з питань освіти та культури (Directorate General for Education and Culture (DG EAC)).

Кожна із цих структур започаткувала власну базу перекладених офіційних матеріалів майже на усіх мовах країн-членів Євросоюзу (ЄС), які позиціонуються як паралельні корпуси або бази паралельних текстів у відповідних файлових форматах. Такі ресурси, як правило, використовуються професійними перекладачами у поєднанні з системами автоматизованого перекладу (CAT-системи) для підвищення швидкості та якості перекладів. При цьому важливо відмітити багатоаспектність їх застосування, що може охоплювати створення баз пам'яті перекладів, екстрагування термінів до термінологічних баз, створення спеціальних словників для реалізації функцій автоматичного введення термінів у процесі перекладу тощо.

Незважаючи на різні підходи до побудови таких баз, перекладачі мають можливість, залежно від потреби, компілювати на їх основі власні двомовні чи багатомовні ресурси, пов'язуючи різні мовні версії одного і того ж документу [Амеліна; Тарасеко 2017, с. 153].

Підтвердженням значущості зазначених баз для підтримки процесу перекладу може слугувати їх характеристика, зокрема:

JRC-Acquis – багатомовний паралельний корпус на 22 мовах ЄС, що містить в цілому більше 1 мільярда слів. Цей корпус був започаткований Єврокомісією у 2006 році і став першим корпусом такого спрямування.

У ньому сконцентровані повнотекстові документи ЄС (директиви, адміністративні документи, декларації, резолюції, угоди, акти тощо) в TEI форматі, які після певних перетворень можуть бути використані у системах автоматизованого перекладу. JRC-Acquis є єдиним з корпусів ЄС, де вирівнювання текстів було здійснено із використанням двох різних типів програмного забезпечення.

Це дозволило провести порівняння ефективності вирівнювання і надавати користувачам можливість вибору результатів, які краще відповідають їхнім потребам. Завдяки використанню спеціалізованих програмних засобів користувачі мають змогу вилучати із цього корпусу двомовні вирівняні тексти для будь-якої мовної пари. Сайт: <https://ec.europa.eu/jrc/en/language-technologies/jrc-acquis>

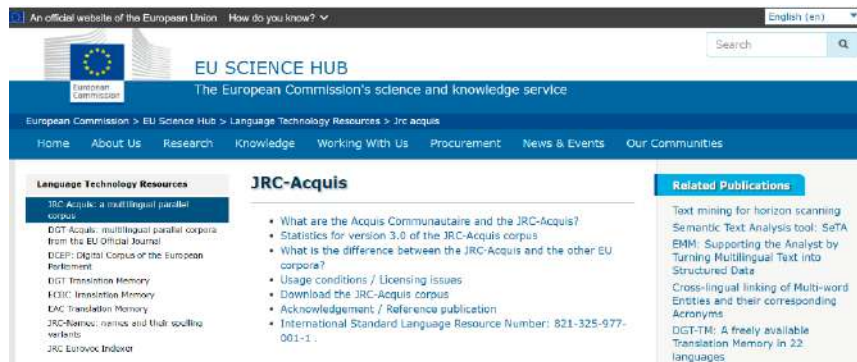


Рис. 2.9. Домашня сторінка багатомовного паралельного корпусу JRC-Acquis

DGT-Acquis – багатомовний паралельний корпус на 23 мовах ЄС. Він створений у 2012 році і включає в себе матеріали усіх серій «Офіційного журналу Європейського Союзу» (Official Journal), починаючи з 2004 року, які охоплюють публікації на теми, пов’язані з суспільним життям, політикою, економікою, охороною здоров’я, інформаційними технологіями, юриспруденцією, сільським господарством, продовольчою політикою, соціальними питаннями. Сайт: <https://ec.europa.eu/jrc/en/language-technologies/dgt-acquis>. Записи бази подані у форматі XML Formex-4, і їх кількість тільки на англійській мові становить 98 мільйонів слів [Амеліна; Тарасеко 2017, с. 154].

DGT-TM – база пам’яті перекладів, яка містить законодавчі матеріали на 24 мовах ЄС, у тому числі всі договори, правила і директиви, прийняті ЄС, а також постанови Європейського Суду. Ця база започаткована у 2011 році, але включає зазначені матеріали з 2004 року. Обсяг записів бази перевищує 1 мільярд слів, які

представлені у форматі TMX. Такий формат представлення даних забезпечує максимальну ефективність його використання у системах автоматизованого перекладу. При вирівнюванні паралельних тестів, призначених для розміщення у базі DGT-TM, здійснюються спеціальні заходи, що передбачають ігнорування як коротких несуттєвих речень, так і занадто довгих, а також усунення явних невідповідностей. Метою таких заходів є створення TMX файлів, які будуть максимально зручними для перекладачів.

Сайт: <https://ec.europa.eu/jrc/en/language-technologies/dgt-translation-memory>.

При цьому здійснюється багаторівнева лінгвістична перевірка всіх документів висококваліфікованими перекладачами, що спеціалізуються в конкретних предметних галузях. Крім того, якість перекладів контролюється юридичними службами та видавничим управлінням ЄК. Багато уваги приділяється забезпеченню узгодженості термінології, яка здійснюється із залученням представників адміністрацій держав-членів ЄС. При цьому перекладачі ЄС працюють у сучасному IT-середовищі, яке надає їм значну кількість спеціалізованих інструментів для оптимізації роботи і забезпечення якості та узгодженості перекладів. Незважаючи на існування певної суб'єктивності у перекладі в цілому, якість перекладених матеріалів у базі DGT-TM відповідає високому рівню. Результатом такого підходу до супроводу бази DGT-TM є найвищий показник завантажень ресурсів цієї бази у порівнянні з іншими корпусами та базами ЄС [Амеліна; Тарасеко 2017, с. 155].

ECDC-TM – база пам'яті перекладів, яка містить матеріали на 23 мовах ЄС, а також норвезькій та ісландській, основна частина яких стосується тем, що пов'язані з питаннями охорони здоров'я, профілактики та контролю захворювань. База створена у 2012 році і містить 320 тисяч записів, що набагато менше, ніж обсяг вище згаданих ресурсів. Але перевага бази ECDC-TM полягає у тому, що вона охоплює такі предметні сфери, матеріали з яких відсутні в інших базах. Записи бази представлені у форматі TMX і супроводжуються програмним забезпеченням, що дозволяє формувати двомовні бази пам'яті перекладів для будь-якої мовної пари. Сайт: <https://ec.europa.eu/jrc/en/language-technologies/ecdc-translation-memory>

ЕАС-ТМ – база пам'яті перекладів, яка створена з файлів, що містять переклади електронних форм проектів або заявок на фінансування програм, а також форми звітності за двома програмами ЄС: «Life-long Learning» та «Youth in Action Programme». База створена у 2013 році і містить понад 540 тисяч записів, які охоплюють кілька предметних сфер, а саме: освіта, професійна підготовка, культура, молодь та спорт. Особливістю цієї бази є те, що вона містить матеріали на більшій кількості мов, ніж будь-який інший корпус чи база Єврокомісії. Крім прийнятих 22 мов, вона включає в себе документи на хорватській, ісландській, норвезькій та турецькій мовах, оскільки ці країни мають право на участь у програмах ЕАС. Сайт: <https://ec.europa.eu/jrc/en/language-technologies/eac-translation-memory>. Як і у попередній базі, записи представлені у форматі ТМХ, а наявність спеціального ПЗ дозволяє формувати двомовні бази пам'яті перекладів для будь-якої мовної пари [Амеліна; Тарасеко 2017, с. 156].

2.3. Основні принципи пошуку в корпусах текстів

З огляду на широке розмаїття корпусних знарядь правила пошуку в різних корпусах можуть суттєво варіюватися, і все ж можна узагальнити його більш-менш універсальні принципи. Окрім простих елементарних запитів слова або сполучення слів, що можна виконувати у пошуковому рядку комп'ютерних оглядачів (браузерів) таких систем пошуку як Google, у корпусних ресурсах існують додаткові можливості, на яких зупинимось детальніше [Фокін 2019].

Однослівні запити пошуку. Можливо, найпоширенішим запитом пошуку є однослівні фрагменти, що їх користувачі вводять у пошуковий рядок. Під час такого пошуку необхідно чітко розмежовувати пошук словоформи та пошук лексеми. Пошук лексики передбачає вибірку словоформ з усієї парадигми заданої лексики. Скажімо, у корпусі **movia.info** відповідні опції називаються “словоформа/лексема” [Корпус текстів української мови]. Сайт: <http://www.movia.info/corpus.aspx?l1=209>.

Алгоритм пошуку за лексемою:

1. Заходимо на домашню сторінку MOVA.info, обираємо рубрику «Корпус української мови».
2. В поле вводимо слово, яке ми хочемо пошукати.
3. Обов'язково вибираємо зону пошуку – тематичну

спрямованість текстів.

4. Є додаткові параметри пошуку:

- глибина контексту (виставляти максимально),
- виведення результату – табличний (за вподобаннями),
- обрати лексема чи словоформа,
- морфологічні ознаки, де обирають частини мови (спрацьовує

нечисто!)

The screenshot shows the MOVA.info website interface. At the top, there is a navigation bar with the logo 'MOVA.info ЛІНГВІСТИЧНИЙ ПОРТАЛ' and a search bar. Below the navigation bar, there is a breadcrumb trail: 'Головна / Проекти / Корпус української мови'. The main heading is 'КОРПУС УКРАЇНСЬКОЇ МОВИ'. The search interface includes a dropdown menu for 'Виберіть зону пошуку' with 'НАУКОВІ ТЕКСТИ' selected, and buttons for 'Вибрати' and 'Додаткові параметри'. There is a section for 'Морфологічні ознаки' with radio buttons for 'Лексема' (selected) and 'Словоформа'. The search input field contains the word 'конденсат'. Below the search field, there is a button '+ Додати ще одне слово' and two buttons: 'Знайти' and 'Очистити параметри'. The search results are displayed under the heading 'НАУКОВІ ТЕКСТИ' in a table format.

<< у себе "	конденсат	образу ", в >>	Джерело
<< Тут фільтрується	конденсат	атмосферних опадів, насичений >>	Джерело
<< газ і газовий	конденсат	, Про державне регулювання >>	Джерело
<< Пара конденсується,	конденсат	змішується з водою і >>	Джерело
<< конденсації в дефлегматорі	конденсат	розподіляється на два шари >>	Джерело
<< При рекуперації виділяється	конденсат	, який надходить у >>	Джерело

Рис. 2.10. Результат пошуку лексеми «конденсат» у корпусі MOVA.info

Пошук у корпусі української мови ГРАК здійснюється через інтерфейс NoSketch Engine. Для пошуку в корпусі треба перейти на сторінку пошуку (https://parasol.vmguest.uni-jena.de/grac_crystal/#dashboard?corpname=grac15), обрати в меню

потрібний корпус (наприклад, Grac v. 15) і натиснути "Concordance".



Рис. 2.11. Сторінка пошуку у ГРАКу через інтерфейс NoSketch Engine

Для пошуку важливо знати, що ГРАК має морфологічну розмітку. Кожному токenu в корпусі приписано лему і теги.

У ГРАКу можна шукати за:

- ✓ за словоформою (word) – для пошуку слова в точній граматичній формі або для пошуку нелематизованого слова;
- ✓ за лемою (lemma) – для пошуку слова в усіх граматичних формах. Неюхідно обрати потрібний атрибут в меню:

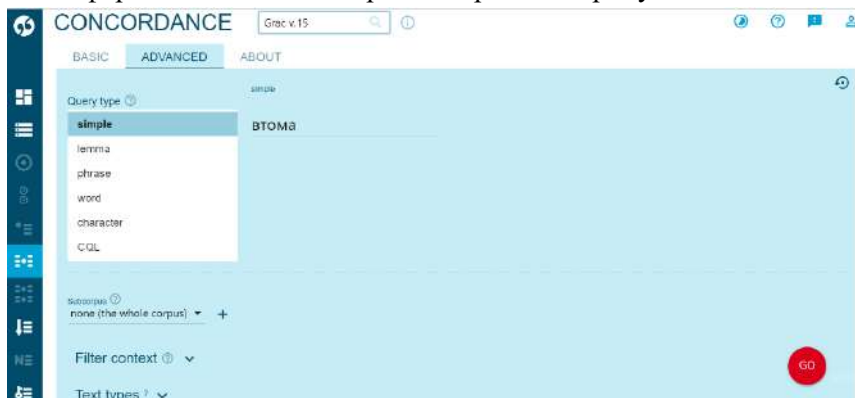


Рис. 2.12. Приклад простого пошуку у ГРАКу-15 лексеми «ВТОМА»

Простий запит (simple) працює не зовсім прозоро: для лематизованих слів він спрацьовує як пошук за лемою, а для нелематизованих – як пошук за словоформою. Сполучення кількох лематизованих слів опрацьовує як послідовність лем.

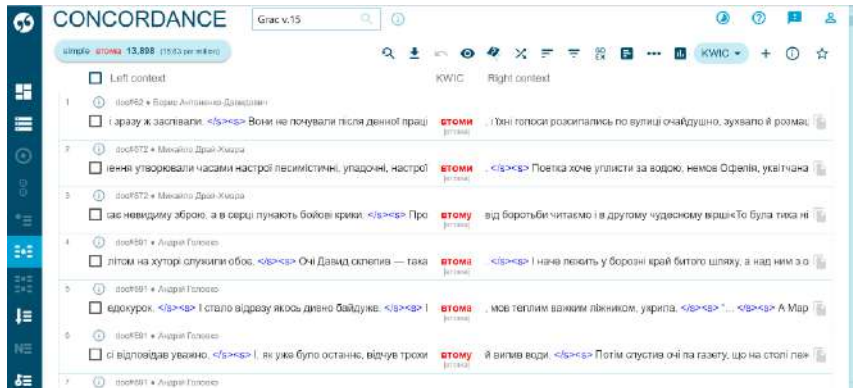


Рис. 2.13. Результат простого пошуку у ГРАКу-15 лексеми «ВТОМА»

У рядку пошуку *simple* працюють шаблони пошуку (англ. *wildcard*). Зірочку (*), знак питання (?) та подвійний дефіс (--) можна використовувати таким чином:

1) зірочка (*) для 0 або більшої кількості символів, наприклад, *шука** дасть результат *шука, шукати, шукають, шукання...*

*я*к* дасть результат *як, язик, ярмарок...*

2) знак питання (?) означає лише один символ, наприклад, *сто?* дасть результат *стою, стоп*, але не знайде *сто*

к?т знайде такі лемі *кіт, кат, кут*.

Щоб шукати зірочку і знак питання, використовуйте знак (\): наприклад, ** та \?*

3) подвійний дефіс (--) позначає дефіс, пробіл або відсутність символу, наприклад, *пів--села* дасть результат *пів села, півсела, пів-села*.

Багатослівні запити пошуку. Користувачів корпусу може цікавити не лише пошук оточення певної словоформи чи лексеми, а й ширші парадигматичні аспекти. Для перевірки сполучуваності слова, варіантності тих чи інших виразів, фразеологізмів, нюансів

дієслівного, іменникового, прикметникового керування доцільно вводити декілька слів. одразу Більшість сучасних корпусів такий пошук підтримують. У рамках багатослівного пошуку можна використовувати “байдужі символи”, що можуть заміщувати будь-яку літеру, сполучення літер, або регулярні вирази. Однак зі збільшенням обсягу корпусу запити з узагальненими виразами, масками слів, а надто – словосполучень, суттєво уповільнюють роботу програм.

Звичайно, ми не маємо змоги проілюструвати навіть мінімальну базову частину таких запитів, інструкції зі здійснення таких запитів займають сотні сторінок. CQL первинно розроблено з метою пошуку видань в електронному каталозі Бібліотеки Конгресу США. Чіткої уніфікації оформлення цих запитів у різних системах пошуку немає. З метою адаптації CQL до застосування у конкордансах розроблено систему пошуку Word sketch engine2.

Розглянемо багатослівний запит пошуку у корпусі української мови ГРАК, де, окрім простого, можна здійснювати пошук за фразою (точною послідовністю символів з пробілами; phrase), щоб знайти сполучення з кількох слів у точній граматичній формі. Також є можливість використовувати теги під час складнішого пошуку за допомогою CQL.

Використовуючи CQL, можна задати комплексні критерії для пошуку одного чи кількох токенів. Критерії для кожного токена мають бути у квадратних дужках [], у форматі: [атрибут="значення"].

Атрибут – це тип завдання для пошукової системи, а значення – його конкретна реалізація в цьому запиті.

Наприклад, щоб знайти лему бабак (різні граматичні форми слова бабак), використовуйте: [lemma="бабак"].

Щоб знайти лему майоріти, використовується такий запит: [lemma="майоріти"].

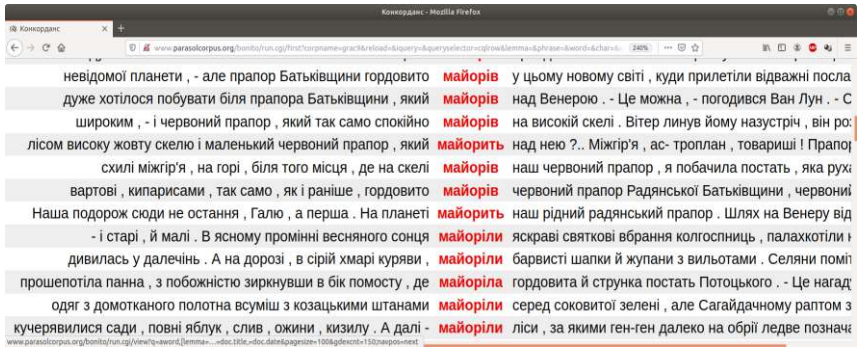


Рис. 2.14. Результат багатослівного пошуку у ГРАКу-15 лем «майоріти»

Запит для кожного токена має міститися в окремій парі квадратних дужок. Для пошуку фрази співати веселих пісень з урахуванням можливих граматичних варіантів головного та залежних слів введіть:

[lemma="співати"] [lemma="веселий"] [lemma="пісня"]

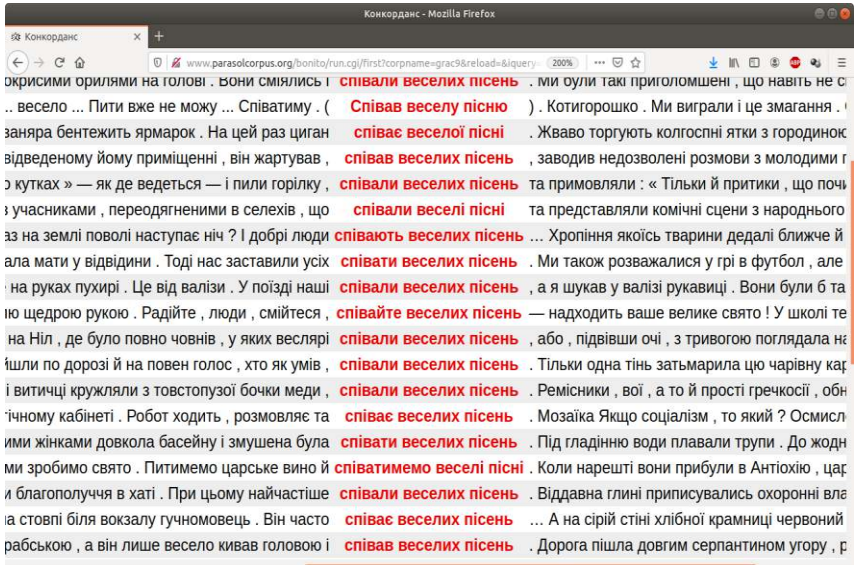


Рис. 2.15. Результат багатослівного пошуку фрази у ГРАКу-15

Пробіли – не виконують жодної функції в CQL. Можна використовувати пробіли, щоб зробити текст зручнішим для читання. Такий запит нічим не відрізняється від попереднього:

[lemma="співати"][lemma="веселий"][lemma="пісня"]

Однак вираз не спрацює, якщо в ньому буде декілька пробілів поспіль. У лапках не повинно бути жодних пробілів. Наприклад, такий пошук не дасть результатів, оскільки лема ніколи не може починатися з пробілу: [lemma=" хто"]

Ретельніше про мову корпусних запитів – CQL (Corpus Query Language) можна ознайомитись в інструкції за покликанням <https://drive.google.com/file/d/1Ti3VGpFYbzXRYO0tPVZoyGT343e9jiWI/view>.

Отже, у сучасних корпусах пошук можна здійснювати за допомогою різноманітних запитів: простих запитів, що складаються з одного або декількох графічних слів, а також метамовних засобів, з використанням спеціальних дескрипторів, тегів граматичної, частиномовної, семантичної, екстралінгвістичної характеристики слів, виразів, що підлягають пошуку; широкі можливості пошуку надає використання регулярних виразів, які в узагальненому вигляді описують певну множину текстових реалізацій, які підходять під певну маску і можуть бути знайдені за одну операцію. За окремими формальними показниками можна робити висновки й щодо семантики, функцій словоформ.

Перевага корпусів текстів над пошуком у загальному масиві інтернетних текстів полягає у тому, що корпуси достовірніше репрезентують як мову в цілому (нормалізовані корпуси), так і окрему підмову, стиль, жанр, дискурс [Фокін 2019].

Використана література та інформаційні ресурси:

1. Амеліна С. М., Тарасенко Р. О. Підготовка перекладачів у зарубіжних університетах до застосування інформаційних технологій: монографія. Київ : «Центр учбової літератури», 2017. 256 с.

2. Долинський Є.В., Скиба К.М. Інформаційні технології у професійній діяльності перекладача : навчальний посібник з інформаційних технологій для студентів першого (бакалаврського) рівня, а також другого (магістерського) рівня вищої освіти, галузі знань «Гуманітарні науки», спеціальності «Філологія», «Переклад». Хмельницький : ХНУ, 2016. 195 с.

3. Жуковська В.В. Вступ до корпусної лінгвістики: навчальний посібник. Житомир: Вид-во ЖДУ ім. І. Франка, 2013. 142 с.
http://eprints.zu.edu.ua/18909/1/korpusna_lingv.pdf

4. Фокін С.Б. Комп'ютерні інформаційні технології в перекладацькій та перекладознавчій діяльності. К.: Київ. нац. ун-т ім. Т.Шевченка, 2019. 254 с.

<https://www.youtube.com/watch?v=y2GYGaQ8TDo>

<https://www.sketchengine.eu/user-guide/usermanual/corpora/by->

<http://www.mova.info/corpus.aspx?l1=209>

<https://ec.europa.eu/jrc/en/language-technologies/eac-translation-memory>

<https://ec.europa.eu/jrc/en/language-technologies/ecdc-translation-memory>

<https://www.ids-mannheim.de/digspra/kl/projekte/korpora>

http://agd.ids-mannheim.de/ZW--_extern.shtml

http://agd.ids-mannheim.de/korpus_index.shtml

https://corpora.uni-leipzig.de/de?corpusId=deu_newscrawl_2011

<https://www.zdl.org>

<http://korpus.org.ua>

СЕМІНАРСЬКЕ ЗАНЯТТЯ 2

Теоретичні питання для обговорення:

1. Яку роль відіграють корпусна лінгвістика й корпуси мовних даних у вивченні мов та перекладі?

2. Дайте визначення електронним лінгвістичним корпусам!

3. Як називають зібрання текстів в електронній формі, що презентує певну мову?

4. Назвіть основні ознаки корпусу!

5. Який з електронних корпусів текстів був першим?

6. Які дані можна одержати на основі корпусу як довідкового ресурсу перекладача?

7. Які важливі характеристики повинно мати довільне зібрання текстів, щоб називатися корпусом?

8. Для чого призначені тексти в електронному корпусі?

9. Як здійснювати пошук у Корпусі української мови? Які підкорпуси є у розпорядженні перекладача/користувача?

10. Чи може корпус текстів бути джерелом для укладання різних словників?

11. Проілюструйте можливості цифровий словник німецької мови, який розробила Академія наук Берлін-Бранденбург (Die Berlin-Brandenburgische Akademie der Wissenschaften)!

12. Які дані містить Словниковий портал Лейпцигського університету (Wortschatz-Portal der Universität Leipzig)?

13. Чим може бути корисним Корпус німецьких діалектів (Korpus Deutsche Mundarten) у діяльності перекладача?

14. Назвіть найбільший довідковий корпус Німеччини!

15. Чи розглядаються штучно створені тексти і веб-сторінки як можливі джерела електронних корпусів? Поясніть свою відповідь!

16. Коли розпочав створення електронних корпусів Інститут німецької мови (IDS)?

17. Чому у своїй більшості сучасні корпуси текстів одномовні?

18. Опишіть процес здійснення пошуку в корпусі паралельних текстів на прикладі онлайн-ресурсу Glosbe!

19. Назвіть популярні онлайн-ресурси корпусів паралельних текстів, що містять мовні пари, у рамках яких здійснюється велика кількість перекладів!

20. Дайте короткий огляд основних баз перекладених офіційних матеріалів майже на усіх мовах країн-членів Євросоюзу, які позиціонуються як паралельні корпуси!

21. Як такі ресурси використовуються професійними перекладачами?

22. У чому полягає перевага корпусів текстів над пошуком у загальному масиві інтернетних текстів?

Практичні завдання:

1. Підготуйте доповідь з презентацією до теми *«Неконвенційні види виконання перекладацьких замовлень: Crowdsourcing»*. Продемонструйте можливості виконання перекладацьких проєктів за допомогою методу crowdsourcing.

Рекомендовані інтернет-ресурси:

https://ela.kpi.ua/bitstream/123456789/45821/3/Debela_magistr.pdf

<https://www.dropbox.com/s/b1sl8c96hu0w6xn/%D0%9A%D1%80%D0%B0%D1%83%D0%B4%D1%81%D0%BE%D1%80%D1%81%D0%B8%D0%BD%D0%B3%20%D0%BD%D0%B0%20%D0%A4%D0%98%D0%AF.pptx?dl=0>

2. Підготуйте доповідь з презентацією до теми *«Функції текстових редакторів, корисні у практиці перекладу»*,

прокоментуйте важливість функції пошуку і глобальної заміни у текстових редакторах, автоматичного розрахунку обсягу тексту, контролю якості перекладу за допомогою текстових редакторів та інші.

Лабораторна робота 2:

Підготовка матеріалів для укладання міні-корпусу інструкцій з експлуатації кухонного приладдя відомих німецьких виробників (міксер, блендер, кухонний комбайн, м'ясорубка, кавоварка, мультіварка, хлібопіч тощо). Для цього: з сайтів відомих німецьких виробників завантажте інструкцію з експлуатації приладів, яка містять український переклад. Далі виконайте такі кроки:

Крок 1. Завантажте PDF-файл інструкції на комп'ютер, За потреби розпізнайте текст за допомогою безкоштовної онлайн-програми, або ж конвертуйте файл з PDF у Word за допомогою Acrobat Adobe Pro.

Крок 2. Перевірте тексти обох файлів на наявність помилок, невідповідностей, щоб уникнути проблем при подальшому вирівнюванні паралельних текстів. Часом трапляються такі огріхи, як наприклад, неперекладений сегмент тексту:

Увага

- Ножи очень острые! Во избежание травм, пожалуйста, обращайтесь с ножами с особой осторожностью.
- Пристрій дозволяється використовувати особам з обмеженими фізичними й сенсорними можливостями чи розумовими здібностями або особам, що не мають достатнього досвіду та знань, якщо вони перебувають під наглядом або отримали інструкції щодо використання приладу й усвідомлюють небезпеку, що може виникнути через неправильну експлуатацію.
- Дітям заборонено бавитись з даним приладом.
- Пристрій заборонено використовувати дітям.
- Не слід допускати дітей до приладу

• Коли прилад працює від електромережі, зокрема під час роботи приводу, для накладання продуктів користуйтеся лише штовхачем.

• Деталі пристрою не призначені для використання у мікрохвильових пічках.

Помилка в інструкції:
залишили російський варіант!!!

Деталі

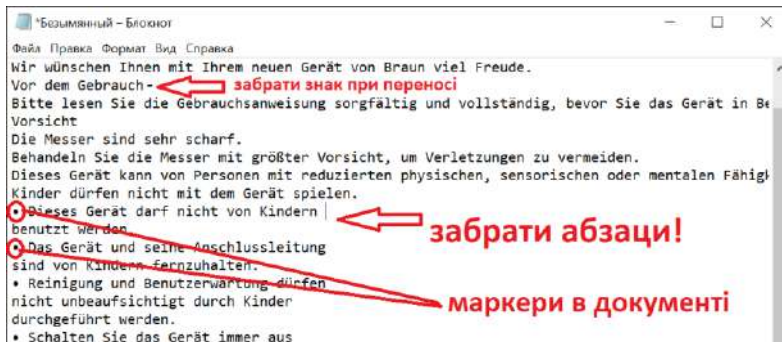
- 1 Сигнальна лампочка
- 2 Кнопка безпечки
- 3 «Розумний» перемикач швидкостей Smart Speed / зміна частоти обертання „all
- 4 Моторний блок
- 5 Кнопки розблокування EasyClick («легке натискання»)
- 6 Стрижень ActiveBlade («активний ніж»)
- 7 Стакан
- 8 Насадка-віничок з коробки швидкостей б віничок
- 9 Насадка для пюре

Крок 3. Переформатуйте файли у формат *.txt. У меню "Зберегти як" виберіть тип файла – *.txt (замість зірочки наберіть назву документа), кодування – UTF-8.

Крок 4. Розділіть німецький текст та відповідний український переклад інструкції та збережіть в окремі файли (за мовами). Обсяг

кожної інструкції – приблизно 2000 слів / 15000 знаків з пробілами).

Крок 5. Здійсніть рецензування текстових документів. Видаліть всю інформацію, що не є релевантною для вирівнювання тексту (малюнки, таблиці, зноски, колонтитули), при цьому залишаючи всі композиційні частини інструкції.



Крок 6. Збережіть файлові документи (наприклад, STABMIXER_DEU.txt; STABMIXER_UKR.txt) у відповідні папки на своєму комп'ютері.

ТЕМА 3

Системи машинного та автоматизованого перекладу (CAT-системи)

План лекції:

- 3.1. Загальна характеристика систем машинного та автоматизованого перекладу
- 3.2. Системи автоматизованого перекладу (CAT-системи)

3.1. Загальна характеристика систем машинного та автоматизованого перекладу

Майбутнім перекладачам доведеться працювати з великими обсягами різної іншомовної літератури та документації, що потребує вирішення проблеми автоматизації перекладу. Таку проблему дозволяють вирішити *системи машинного перекладу* (МП).

Особливу, і в деякій мірі специфічну, групу лінгвістичного програмного забезпечення складають системи машинного

перекладу. Системи МП виконують переклад, а людина при цьому необхідна для організації роботи комп'ютера. У системі «людина-комп'ютер» відбувається протилежний розподіл функцій. Аналіз літературних джерел дозволяє стверджувати, що «машинний», а точніше, комп'ютерний переклад – це письмовий переклад, оскільки в результаті ми одержуємо письмовий текст. Однак здійснює його не перекладач, а особлива комп'ютерна програма [Долинський; Скиба 2016, с. 96].

Сучасні комп'ютерні програми перекладу досить довершені, але вони дотепер не можуть вирішити найскладніше завдання процесу перекладу: вибір контекстуально необхідного варіанта, який у кожному тексті обумовлений багатьма причинами. При цьому, результат цього виду перекладу може бути використаний як чорновий варіант майбутнього тексту, що буде редагувати перекладач, а також як засіб, отримання загального уявлення про тему і зміст тексту в ситуації, коли відсутній перекладач.

Ще більш складне завдання являє собою переклад усного тексту за допомогою комп'ютерних програм, тому що проблема розпізнавання усного мовлення перебуває лише на початковому етапі свого рішення. Дотепер непереборною перешкодою є індивідуальне забарвлення звучання сегмента мовлення на будь-якій мові таке мовлення погано формалізується.

Інформація – абстрактне поняття, що має різні значення залежно від контексту. Тому необхідне обов'язкове редагування тексту науковим редактором, який володіє не лише мовою, якою виконаний переклад, але й мовою оригіналу. До перекладів, у галузі інформаційних технологій, залучаються лише ті перекладачі, які професійно володіють іноземною мовою та спеціалізуються на перекладах саме у сфері інформаційних технологій. Глибоке розуміння предметної галузі в перекладачів вдається розвинути завдяки їх постійній роботі з науковими редакторами. Паралельно відбувається виявлення важких для розуміння та перекладу тематик, яким варто приділити особливу увагу. Ідея використання комп'ютера для автоматичного перекладу текстів виникла ще на початку появи обчислювальної техніки. Для автоматичного перекладу документів з однієї мови на іншу розроблено багато різних програм. Однак через складнощі опису семантики природничих мов до цього часу остаточно проблему перекладу ще

не вирішено. Проте сучасні засоби автоматизації перекладу досягли такого рівня, який дає змогу ефективно використовувати їх на практиці [Долинський; Скиба 2016, с. 96].

Програми автоматичного перекладу документів доцільно використовувати:

- при абсолютному незнанні іноземних мов;
- у разі необхідності одержати переклад швидко, наприклад, при перекладі Web-сторінок;
- для створення підрядкового перекладу – чернетки, що використовується для повноцінного перекладу;
- у разі пересилання документів іноземним партнерам.

Вивчаючи проблему машинного перекладу, слід розглянути і окремо різні її аспекти. Ларі Чайлдс запропонував ще у рамках міжнародної конференції з технічної комунікації 1990 року такий поділ МП:

- ✓ повністю автоматизований (автоматичний) переклад;
- ✓ автоматизований машинний переклад за участю людини;
- ✓ переклад, здійснений людиною з використанням комп'ютера.

Повністю автоматизований (автоматичний) машинний переклад. Цей вид машинного перекладу і мається на увазі більшістю людей, коли вони говорять про машинний переклад. Сенс тут простий: у комп'ютер вводиться текст однією мовою, цей текст обробляється і комп'ютер виводить той самий текст іншою мовою.

Автоматизований машинний переклад за участю людини. Цей вид машинного перекладу тепер є можливим. Кажучи про машинний переклад за участю людини, зазвичай мають на увазі редагування текстів як до, так і після їхньої обробки комп'ютером. Люди-перекладачі змінюють тексти так, щоб вони були зрозумілі машинам. Після того, як комп'ютер зробив переклад, люди знову-таки редагують грубий машинний переклад, роблячи текст на вихідній мові правильним. Крім такого порядку роботи, існують системи МП, що вимагають постійної присутності людини-перекладача, і це допомагає комп'ютеру виконувати переклад особливо складних або неоднозначних конструкцій. Машинний переклад за допомогою людини застосовний більшою мірою до текстів з обмеженим вокабуляром вузько обмеженої тематики.

Сучасні системи МП працюють приблизно за таким самим алгоритмом, за яким загалом працює звичайний перекладач. Зазвичай система виконує переклад по реченнях. Вже на цьому етапі можливі потенційні помилки. Перекладач бачить весь текст або значний його фрагмент і розуміє контекст. Відсутність контексту породжує помилки. Потім система МП виконує синтаксичний аналіз для знаходження найбільш вірогідного підмета, присудка та інших членів речення. На жаль, далеко не всі реальні речення можуть співпадати з моделями, які існують у блоці синтаксичного аналізу системи. Підмет, виражений інфінітивом, дуже часто інтерпретується системою МП як присудок чи його частина, тоді переклад виходить неякісним. Після синтаксичного аналізу речення на мові оригіналу слідує конструювання граматичної конструкції для мови перекладу і заміна цієї конструкції слів однієї мови на еквіваленти в іншій мові. Але в умовах багатозначності помилки просто неминучі [Долинський; Скиба 2016, с. 97].

У цій коротко представленій технології машинного перекладу можна відмітити критичні точки, в яких неважко побачити резерви підвищення його якості. Якщо текст оригіналу написаний чітким логічним стилем, якщо речення повні і короткі, то синтаксис тексту перекладу буде цілком задовільним. Як ми вже відмітили, у наш час вже є солідні фірми, документацію яких готують «сумісно» не розробники і творці продукції, а кваліфіковані лінгвісти.

Якщо задуматися над лексичними проблемами систем МП, то і тут існують резерви для підвищення якості перекладу. Переглянувши текст, який направляється на переклад, лінгвіст цієї системи може вказати пріоритетний список словників, якими повинна користуватись система при перекладі. Звичайно, такі словники повинні бути тематичними. Вони потребують постійного корегування і доповнень, але при сучасному розвитку лексикографії і оснащення її потужними комп'ютерними засобами – це вже не є проблемою.

Переклад, здійснюваний людиною з використанням комп'ютера. При цьому підході людина-перекладач ставиться в центр процесу перекладу, тоді як програма комп'ютера розцінюється як інструмент, що робить процес перекладу ефективнішим, а переклад – точним. Це звичайні електронні словники, які забезпечують переклад необхідного слова,

покладаючи на людину відповідальність за вибір потрібного варіанта і сенс перекладеного тексту. Такі словники значно полегшують процес перекладу, але вимагають від користувача певного знання мови і витрат часу на його здійснення. Та все ж сам процес перекладу значно прискорюється і полегшується.

Системи МП не дають прийнятної якості вихідного тексту. Вищої якості можна досягти за допомогою попереднього налаштування системи (продукти серії PROMT XT надають користувачеві безліч можливостей для цього), що абсолютно неприйнятно при невеликих об'ємах тексту, що перекладаються, або шляхом подальшого редагування, а це тільки уповільнює роботу, якщо перекладач використовує сліпий метод друку.

Системи МП не гарантують дотримання єдності термінології, особливо при роботі колективу перекладачів над великим проектом. Точніше, вони можуть гарантувати єдність термінології, але за умови дуже уважного поводження з призначеними для користувача словниками, а на це не завжди варто розраховувати.

Програми перекладу забезпечують повний цикл перекладу всього документа: введення початкового тексту, переклад на іншу мову, редагування, форматування і збереження перекладеного тексту [Долинський; Скиба 2016, с. 98].

Отже, розглянемо кілька найпоширеніших у користуванні програм-перекладачів, які дозволяють вирішити проблему перекладу та покращити ефективність праці перекладача.

Перекладач Google, Google Translate, або Google Translator – безкоштовний сервіс компанії Google, що дозволяє автоматично перекладати слова, фрази та web-сторінки з однієї мови на іншу. Сайт: <https://translate.google.com/?hl=uk>.

Google використовує власне програмне забезпечення для перекладу. Використовується підхід статистичного машинного перекладу. Вважається найпопулярнішим ресурсом для миттєвого перекладу з української мови на понад 100 мов світу і навпаки.

Google Translate була запущена в квітні 2006 року як статистична служба машинного перекладу, яка використовувала документи організації Об'єднаних Націй та Європейського парламенту для збору лінгвістичних даних. Під час перекладу вона шукала закономірності в цих документах, щоб допомогти вирішити,

який з них найкраще підійде для перекладу. Спочатку сервіс підтримував всього 4 мови (англійську та арабську, китайську і російську). Зараз додаток включає більше 100 мов і здатний перекладати на 37 мов за допомогою фото, 32 мови – за допомогою голосу і 27 мов – за допомогою відео в режимі реального часу. Вражас, якби не його недоліки. З 2008 року підтримується й переклад на українську мову. Станом на вересень 2016 року, Google-перекладач підтримує майже всі функції для української мови, включно з вимовлянням слів вголос, візуальним перекладом за допомогою офіційного застосунку для Android тощо. Функція миттєвого перекладу за допомогою камери смартфона доступна для україномовних користувачів з липня 2019 року.

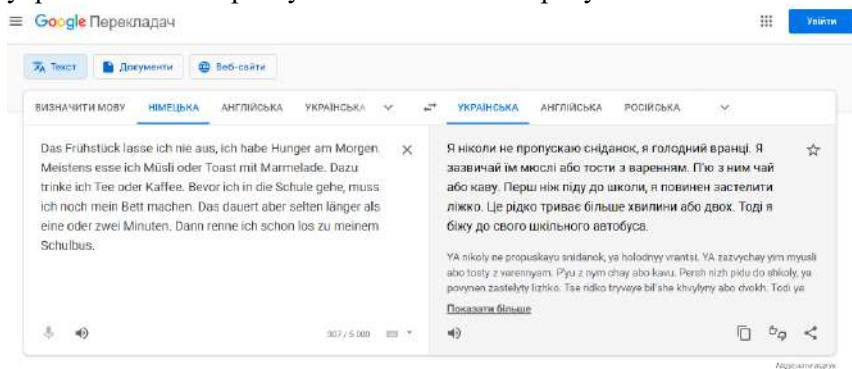


Рис. 3.1. Приклад перекладу німецького сегменту українською мовою в системі машинного перекладу Google Translate

Проте існують певні недоліки використання сервісу Google Translate для перекладу:

1. Машинний переклад економить час і скорочує витрати, але безкоштовні платформи перекладу не ідеальні, коли мова йде про важливі/конфіденційні/творчі документи. Клієнти можуть вибрати бюро перекладів і бути впевнені, що їх документи будуть у безпеці та захищені.

2. Машинне навчання та інструменти машинного перекладу, як відомо, мають помилки. Зараз Google Translate використовує комбінацію машинного навчання та допомоги волонтерів, щоб переконатися, що переклади є більш точними, але вони все ще далекі від досконалості. У бюро перекладів вже склалася практика

забезпечення якості, що дозволяє гарантувати кращі результати.

3. Google Translate досі не розуміє літературу, музику та поезію через нюанси, необхідні для передачі оригінальних ідей при перекладі.

4. Зворотні переклади з використанням Google Translate не гарантують, що контент був перекладений точно.

5. Для використання Google Translate необхідно постійне підключення до Інтернету.

6. Перекладачі не можуть використовувати Google Translate, тому що їм довелося б вводити в програму все, що говорили люди на зборах і конференціях, щоб забезпечити переклад. Це було б трудомісткою, неефективною і неточною справою.

DeepL Translator – сервіс нейронного машинного перекладу, також високо оцінюється користувачами, особливо за точність перекладу. Вважається, що він подекуди навіть кращий за Google Translate. Сайт: <https://www.deepl.com/uk/translator>.

DeepL Translator був запущений 2017 року компанією Linguee і пропонував переклади між англійською, німецькою, французькою, іспанською, італійською, нідерландською та польською мовами. Після запуску DeepL заявив, що він перевершив своїх конкурентів в самостійно проведених сліпих тестах і за показниками BLEU, включаючи Google Translate, Amazon Translate, Microsoft Translator і функцію перекладу Facebook. У 2020 року також було заявлено, що сервіс перевершив вищезгаданих конкурентів.

Служба була створена німецькою компанією Linguee GmbH, заснованою у 2008 році Героном Фралінгом, який раніше працював у Google, та Леонардом Фінком. Вони створили словник для 25 мов, який можна використовувати для перекладу через мережеву службу. З випуском DeepL у 2017 році назву компанії було змінено на DeepL GmbH. У березні 2020 року DeepL оголосила, що досягла значного збільшення продуктивності в порівнянні зі своїми конкурентами, змінивши свою систему штучного інтелекту. З 2022 року DeepL Translator став доступним для української мови.

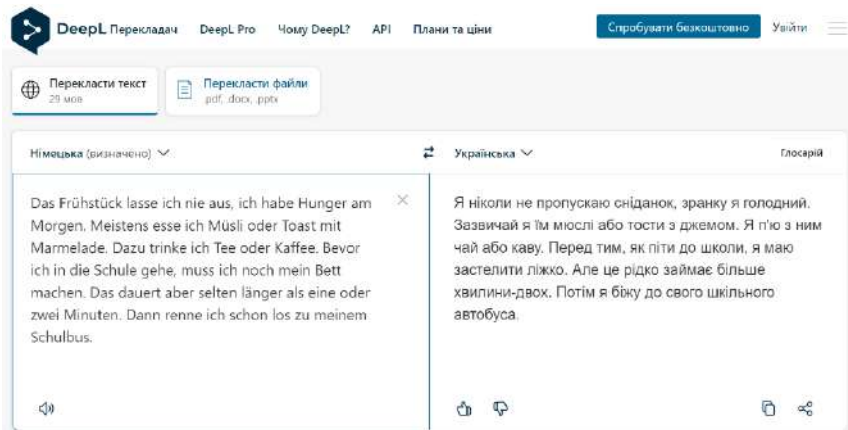


Рис. 3.2. Приклад перекладу німецького сегменту українською мовою в системі машинного перекладу DeepL

Перекладачем можна користуватися безкоштовно, з обмеженням 5000 символів на переклад. Клієнти комерційного програмного забезпечення можуть використовувати платний API для вбудовування DeepL у своє програмне забезпечення — від 5,99 євро на місяць для особистого користування.

Також можна перекладати Документи Microsoft Word у форматах файлів Office Open XML (.docx), а також презентації PowerPoint (.pptx). Зноски, форматування та вбудовані зображення зберігаються.

Перекладач DeepL отримав премію Webby Award 2020 за найкращі практики за технічні досягнення (застосунки, мобільні пристрої та функції), обидві в категорії «Застосунки, мобільні пристрої та голос».

Система машинного перекладу **PROMT XT**. В основу програмних продуктів компанії PROMT поставлено вирішення наступних фундаментальних проблем:

- 1) зрозуміло, що чим більший словник, тим кращий переклад, перша проблема – проблема створення великих словників для систем;
- 2) система повинна розпізнавати фразові звороти;
- 3) речення для перекладу пишуться за визначеними

правилами, а значить є ще одна проблема: записати всі ці правила у вигляді програми.

Можливості програми PROMT:

- забезпечення перекладу документів з англійської, німецької і французької мов на російську/українську та навпаки;
- до неї можна підключати кілька десятків спеціалізованих словників, що забезпечує правильний переклад термінів, які стосуються певної області знань;
- динамічне відстеження напрямку перекладу, тобто визначення мови оригіналу і перекладу;
- переклад вмісту буфера обміну, поточного параграфа, виділеного фрагмента тексту або всього тексту;
- забезпечення будь-якого з можливих напрямків перекладу, підключення і відключення словників, доповнення та виправлення їх, складання списку зарезервованих слів, які не перекладаються;
- робота безпосередньо з програмами розпізнавання текстів, наприклад, *FineReader*;
- не виходячи з програми, можна використовувати відомі способи редагування та форматування оригіналу і перекладу;
- забезпечення перевірки орфографії оригіналу і перекладу після встановлення прикладних програм для перевірки правопису (*Lingvo Corrector*, *Pronus*, *Орфо*, *Hugo*).

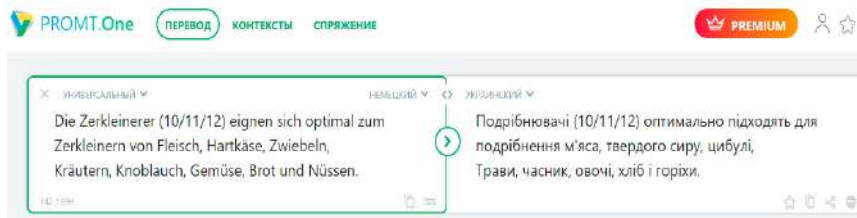


Рис. 3.3. Приклад перекладу німецького сегменту українською мовою за допомогою системи машинного перекладу PROMT

Програмне забезпечення **Pragma** включає два незалежні компоненти Монітор і Коректор. Вони дозволяють перекладати текстові документи з однієї мови на іншу у вказаних напрямках, а

також створювати і коректувати призначені для користувача словники.

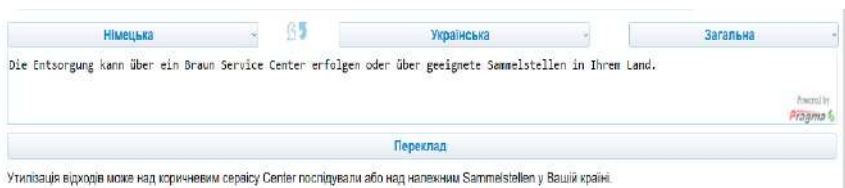


Рис. 3.4. Приклад перекладу німецького сегменту за допомогою системи машинного перекладу PRAGMA

Після інсталяції *Pragma Monitor* програма буде постійно активною і в операційній системі, іконка цієї програми, буде знаходитись на панелі задач, біля годинника. Коректор – окремий компонент, що входить до складу Pragma, який дозволяє створювати і коректувати словники користувача і проглядати системний словник.

На сьогоднішній день до системи включені англійська, російська, німецька, латиська та українська мови з можливістю перекладу в будь-якому напрямі. Переклад можливий між будь-якою парою з цих мов (всього 20 напрямів) [Долинський; Скиба 2016, с. 99]. Також програма дозволяє перекладати документи в різних офіс- та інтернет-додатках з однієї мови на іншу. Програма може інтегруватися в Notepad, WordPad, Word, MS Outlook, Outlook Express, Internet Explorer, Netscape, Mozilla, Open Office. Без реєстрації переклад здійснюється тільки через Інтернет. Після реєстрації з'явиться можливість перекладу в автономному режимі. Програма встановлює всі необхідні словники автоматично.

Уже зараз існують багато фірм, які готують свою документацію з розрахунком на те, що вона буде перекладатися за допомогою систем МП. У таких фірмах випрацюваний власний стиль з обмеженою кількістю граматичних конструкцій і словником, який перешкоджає виникненню проблем багатозначності. При підготовці документа важливим критерієм оцінки мовного стиля являється оцінка перекладу тексту системою МП, якою володіє певна фірма. Враховуються можливі проблеми і передбачаються помилки, які можуть бути усунуті в процесі постредагування.

Якість машинного перекладу і рентабельність його систем значно підвищують дві додаткові технологічні операції – редагування до і після. В процесі попереднього редагування відбувається доопрацювання тексту, який підлягає перекладу. Цей етап нагадує операцію перефразування, якою часто користуються перекладачі. Подумки перекладач перебудовує фразу в більш зручну для перекладу форму: спрощує граматичні конструкції, уточнює багатозначні лексеми. Як правило, великі фірми, які випускають доволі великі об'єми супровідної документації, зацікавлені в перекладі своєї документації одразу на декілька мов. Ретельне лінгвістичне опрацювання і попереднього редагування вихідного тексту економічно доцільні, оскільки підвищують якість одразу декількох перекладів.

Не можна не враховувати тенденцію до формалізації мови оперативної комунікації. В електронній пошті, у коротких повідомленнях у мобільному телефонному зв'язку, в автоматичних підтвердженнях на запити частіше всього використовуються мовні «штампи», що створює вигідну ситуацію для використання систем МП. Додатковий шанс для подібних систем надається у зв'язку з новими тенденціями в діловодстві. Комп'ютеризація текстової діяльності підвищує технологічність текстів. Комп'ютер спрощує редагування тексту під час його підготовки, сприяє його подальшій формалізації, робить його більш технологічним. Усе це розширює сферу застосування систем МП.

Не можна не помітити, що постійно, можна сказати щоденно, в Інтернеті з'являється все більше сайтів, які вміщують систему МП, якою можна скористуватися в онлайн-режимі. Зараз уже не стоїть питання про можливість або неможливість звернення до такого перекладу. Існують лише дві взаємопов'язані проблеми – визначення доцільності використання тієї чи іншої системи МП і пошук засобів підвищення якості отриманого перекладу [Долинський; Скиба 2016, с. 100].

3.2. Системи автоматизованого перекладу (САТ-системи)

Серед класу програмного забезпечення, які суттєво впливають на підвищення ефективності перекладацької діяльності, окремої уваги заслуговують САТ-системи. САТ (computer assisted translation) – це комп'ютерні програми, які автоматизують виконання цілого ряду технічних операцій перекладача, що у

переважній більшості пов'язані із пошуком, впорядкуванням та співставленням інформаційних ресурсів для перекладу.

Завдяки автоматизації таких операцій САТ-системи скорочують трудовитрати на перекладацький проект, пришвидшуючи його виконання та зменшуючи при цьому вартість робіт, але головним виконавцем перекладів при цьому залишається перекладач. Засоби автоматизації дають змогу уникнути багаторазового перекладу фрагментів тексту, які вже були перекладені раніше та зберігаються в так званій базі пам'яті перекладів – Translation Memory, або ТМ. Наприклад, якщо раніше вже був перекладений якийсь сегмент тексту і він був занесений до бази пам'яті перекладів, то наступного разу, коли такий фрагмент зустрінеться у вихідному тексті, САТ-система запропонує його переклад автоматично. САТ-системи дають перекладачеві можливість накопичувати в базах пам'яті перекладів двомовні текстові пари на вихідній і цільовій мовах. Під час роботи з текстами, близькими за жанром і тематикою, переклад пришвидшується завдяки наповненню бази пам'яті перекладів.

Застосування САТ-систем не лише пришвидшує і здешевлює переклад, а й допомагає витримувати термінологічну та стилістичну одноманітність у перекладених текстах, що безумовно підвищує його якість. Підвищенню якості перекладу сприяє також те, що більшість САТ-систем забезпечують узгоджену роботу із термінологічними базами у відповідних форматах, а багато з них містять також засоби формальної перевірки якості перекладу, зокрема відповідності глосаріїв і одноманітності перекладів повторюваних сегментів [Амеліна; Тарасенко 2017, с. 191].

Використання баз пам'яті перекладів під час роботи у середовищі САТ-систем на сьогодні фактично є єдиним способом забезпечення однотипності перекладу за умови командної роботи над великими проектами, коли одночасно працює велика кількість перекладачів і редакторів.

Принцип роботи сучасних САТ-систем полягає в тому, що програма ділить вихідний текст на сегменти (як правило, це речення або частини речень). Розглянемо процес сегментації тексту, який треба перекласти, у найпоширеніших системах автоматизованого перекладу, що доступні онлайн на безоплатній основі з / без реєстрації, або ж оффлайн після інсталяції програми на комп'ютер:

MateCat (<https://www.matecat.com/>), Wordfast Anywhere (<https://www.freetm.com/anywhere.html>), Smartcat та OmegaT.



Рис. 3.5. Принцип сегментації вихідного тексту в системі автоматизованого перекладу MateCat

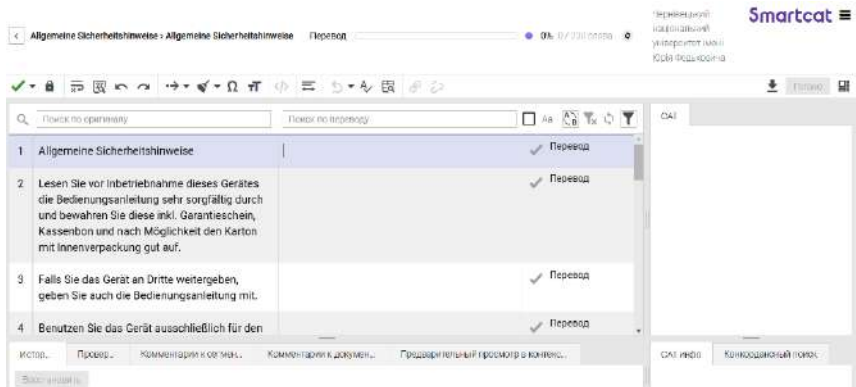


Рис. 3.6. Принцип сегментації вихідного тексту в системі автоматизованого перекладу Smartcat

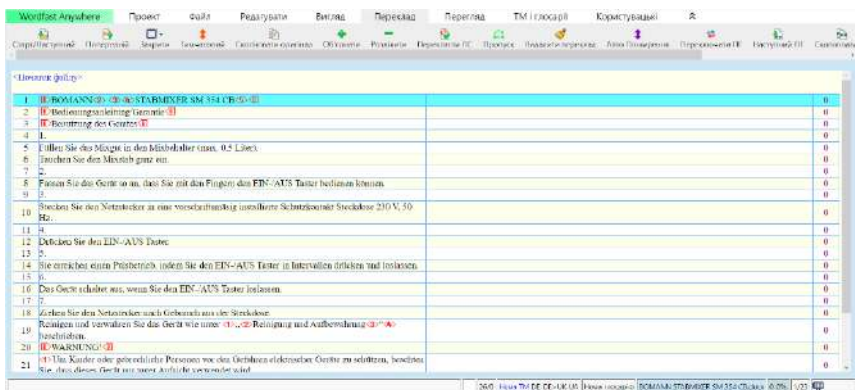


Рис. 3.7. Принцип сегментації вихідного тексту в системі автоматизованого перекладу Wordfast Anywhere



Рис. 3.8. Принцип сегментації вихідного тексту в системі автоматизованого перекладу OmegaT

Після автоматичної сегментації вихідного тексту системою перекладач може розпочинати працювати над текстом перекладу, використовуючи при цьому низку корисних інструментів. При цьому перекладач вписує переклад кожного сегмента у відповідному місці робочого вікна, якщо текст представлений у вигляді таблиці, праворуч від нього. Переклад сегмента зберігається разом із вихідним текстом. Записуються також ім'я перекладача та дата перекладу (що важливо для командної роботи).

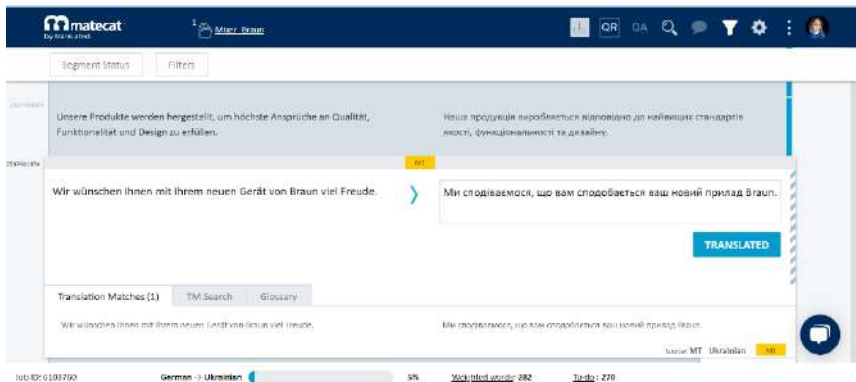


Рис. 3.9. Процес перекладу вихідного тексту в системі автоматизованого перекладу MateCat людиною

До сегмента можна повернутися будь-якої миті, щоб перевірити або змінити переклад. Програма заносить сегмент у пам'ять перекладів, тому якщо він знову трапиться в початковому тексті, його переклад буде підставлено з ТМ автоматично. Крім того, CAT-програма має функцію пошуку нечітких відповідностей: вона виявляє сегменти, що лише частково збігаються з уже перекладеним (наприклад, на 75%), а потім надає «поради» з їх перекладу.

Багато CAT-систем дають змогу підключати до процесу перекладу термінологічні бази для автоматичного пошуку та пропозицій перекладів термінів (іноді для цього застосовують додаткові модулі або програми, наприклад MultiTerm для SDL Trados).

Сучасні CAT-системи не лише працюють із базами пам'яті перекладів, глосаріями, термінологічними базами, а й підтримують модулі перевірки якості перекладу, роботу з метаданими (тегами розмітки html та інших форматів) і виконують інші важливі функції. Більшість CAT-систем дають змогу перевіряти орфографію в перекладених матеріалах (або вбудованими засобами, або підключаючи модуль перевірки правопису Word). Фактично усі потужні CAT-системи мають у своїй структурі окремі модулі для вирівнювання паралельних текстів, які розбивають їх на сегменти,

встановлюють зв'язки між відповідними сегментами у вихідній та цільовій мовах та автоматично формують на їх основі бази пам'яті перекладів [Амеліна; Тарасенко 2017, с. 192].

Для зручності роботи майже всі CAT-системи забезпечують можливість групувати файли із текстами на вихідній мові, відповідні бази пам'яті перекладів та термінологічні бази в окремі проекти. Спектр форматів файлів, у яких можуть бути представлені матеріали для перекладу досить широкий, і включає формати більшості поширених текстових редакторів, формати офісних програм та багато інших.

Переваги застосування CAT-систем полягають у наступному:

- поєднують комплекс інструментів для перекладу документації, ведення термінологічних глосаріїв, створення баз пам'яті перекладів;
- є ефективними при перекладі технічної, фінансової, юридичної та іншої документації з високим ступенем повторюваності текстів;
- економія часу на переклад та форматування текстів становить від 30 до 50%;
- актуальні під час колективної роботи, коли необхідно забезпечити узгоджений переклад групи перекладачів в рамках одного проекту;
- усі результати автоматично зберігаються в єдиній базі даних, доступній усім учасникам процесу перекладу;
- перекладачі можуть знаходитися в одній локальній мережі або підключатися віддалено.

До систем автоматизованого перекладу, які наявні на ринку спеціалізованого програмного забезпечення, можна віднести:

- | | |
|----------------|---------------------|
| - SDL Trados | - OmegaT |
| - STAR Transit | - Transolution |
| - WordFisher | - Translation |
| - MemoQ | - Workspace |
| - WordFast | - Fluency Now |
| - Déjà Vu | - XTM International |
| - WorldServer | - MemSource |
| - MultiTrans | - Anaphraseus |
| - LogiTerm | - MetaTaxis |

Найбільш поширеним програмним продуктом, який використовується професійними перекладачами, є пакет SDL Trados [Амеліна; Тарасенко 2017, с. 193], проте для роботи в системі необхідно придбати ліцензійне програмне забезпечення.

Використана література та інформаційні ресурси:

1. Амеліна С. М., Тарасенко Р. О. Підготовка перекладачів у зарубіжних університетах до застосування інформаційних технологій: монографія. Київ : «Центр учбової літератури», 2017. 256 с.

2. Долинський Є.В., Скиба К.М. Інформаційні технології у професійній діяльності перекладача : навчальний посібник з інформаційних технологій для студентів першого (бакалаврського) рівня, а також другого (магістерського) рівня вищої освіти, галузі знань «Гуманітарні науки», спеціальності «Філологія», «Переклад». Хмельницький : ХНУ, 2016. 195 с.

3. Фокін С.Б. Комп'ютерні інформаційні технології в перекладацькій та перекладознавчій діяльності. К.: Київ. нац. ун-т ім. Т.Шевченка, 2019. 254 с.

4. <https://www.youtube.com/watch?v=JIEpH39HVtM>

ПРОГРАМИ TRANSLATION MEMORY:

Trados: <http://www.translationzone.com/trados.html>

Déjà Vu: <http://www.atril.com/>

OmegaT: <http://www.omegat.org/>

Wordfast: <http://www.wordfast.com/>

MetaTaxis: <http://www.metataxis.com/>

MemoQ: <http://kilgray.com/>

Star Transit: <http://www.star-group.net/DEU/group-transit-nxt/transit.html>

WordFisher: <http://www.wordfisher.com/>

Across: <http://www.across.net/us/translation-memory.aspx>

MateCat: <https://www.matecat.com>

СЕМІНАРСЬКЕ ЗАНЯТТЯ 3

Теоретичні питання для обговорення:

1. Назвіть перекладацькі завдання, які сучасні комп'ютерні програми перекладу дотепер не можуть вирішити!

2. Опишіть реальні ситуації доцільності використання програм автоматичного перекладу документів!

3. Які види машинного перекладу Вам відомі?

4. Чим відрізняється повністю автоматичний машинний переклад від автоматизованого машинного перекладу за участі людини?

5. Опишіть процес перекладу, здійснюваного людиною з використанням комп'ютера!

6. Чи гарантують системи машинного перекладу дотримання єдності термінології, особливо при роботі колективу перекладачів над великим проектом?

7. Які фундаментальні проблеми вирішуються за допомогою система МП PROMT XT компанії PROMT?

8. Які можливості для перекладача дає ця програма МП (PROMT XT)?

9. Програмне забезпечення Pragma включає два незалежні компоненти Монітор і Коректор. Як їх можна використовувати на практиці?

10. Які дві додаткові технологічні операції значно підвищують якість машинного перекладу і рентабельність його систем?

11. Що означає "CAT-системи"?

12. Назвіть основні переваги CAT-систем!

13. Що таке пам'ять перекладів – Translation Memory?

14. За допомогою яких засобів автоматизації перекладач може уникнути багаторазового перекладу фрагментів схожого тексту?

15. Чому за умови командної роботи над великими перекладацькими проектами важливе використання баз пам'яті перекладів під час роботи у середовищі CAT-систем?

16. У чому полягає принцип роботи сучасних CAT-систем?

17. Чи можуть CAT-системи підключати до процесу перекладу термінологічні бази для автоматичного пошуку та пропозицій перекладів термінів?

18. Які окремі модулі мають сучасні потужні CAT-системи?

19. Перерахуйте популярні системи автоматизованого перекладу, які наявні сьогодні на ринку спеціалізованого програмного забезпечення?

20. У чому полягає сутність використання баз пам'яті перекладів, що є основою роботи цілої групи програмних продуктів, віднесених до CAT-систем?

Практичні завдання:

1. Опрацюйте самостійно матеріал до теми «Системи створення баз пам'яті перекладів (*Translation Memory* або скорочено *TM-tools*).
2. Підготуйте доповідь на тему «Системи термінологічного менеджменту».
3. Підготуйте презентацію про порівняльну характеристику перекладу за допомогою машинних та людських ресурсів.
4. Підготуйте презентацію про оцінку якості перекладу тексту з німецької мови українською за допомогою однієї з програм машинного перекладу.

Лабораторна робота 3:

Крок № 1. Підготовчий етап створення бази пам'яті перекладів інструкцій з експлуатації кухонного приладдя відомих німецьких виробників (міксер, блендер, кухонний комбайн, м'ясорубка, кавоварка, мультиварка, хлібопіч тощо).

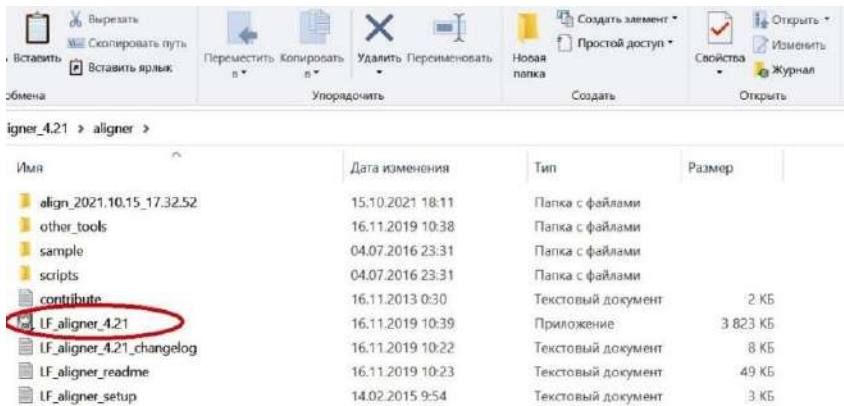
1.1. Підготуйте для вирівнювання паралельних текстів 2 файли інструкцій (німецький оригінал та український переклад) у форматі *.txt UTF-8 без зайвих елементів, як-от малюнків, схем, таблиць, колонтитулів тощо (Див. Лабораторна робота № 2).

Примітка: Для вирівнювання текстів використовують раніше перекладені матеріали, які можуть зберігатися у різних форматах: .txt, .rtf, .doc, .docx, .xls, .xlsx, .ppt, .pptx, .odt тощо.

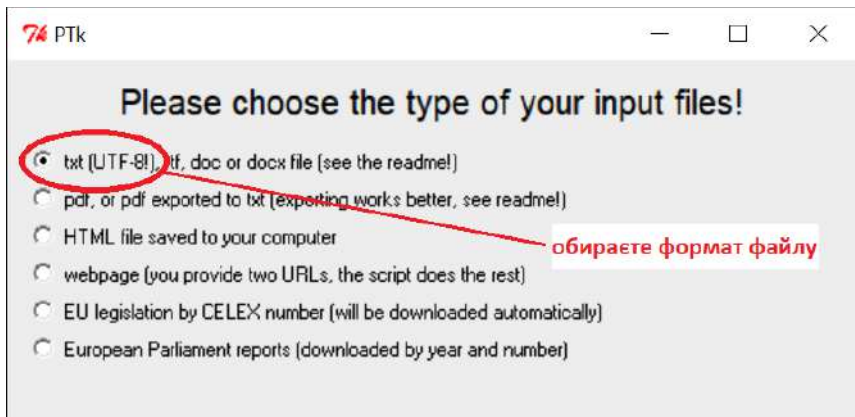
Крок № 2. Вирівнювання паралельних текстів: Основою бази пам'яті перекладів є вирівняні паралельні тексти, тобто тексти, у яких співставлені однакові сегменти тексту (як правило речення) у вихідній і цільовій мовах та збережені у форматі .tmx. Існує цілий клас програм для вирівнювання паралельних текстів, які називаються ALIGNER. Деякі з них потребують встановлення на комп'ютері, але є і такі, що працюють онлайн. Серед найбільш зручних в користуванні – програма *LF Aligner* (безкоштовна, простий інтерфейс).

2.1. Проінсталуйте на свій комп'ютер програму *LF Aligner* (завантажити інсталяційні файли можна безкоштовно з Інтернету).

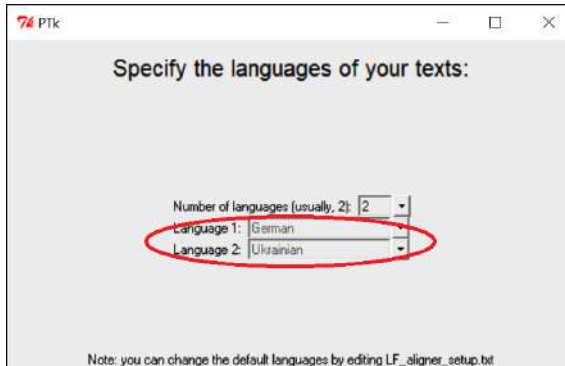
2.2. Розархівуйте папку з файлами програми *LF Aligner* та відкрийте файл з назвою `LF_aligner_4.21:`



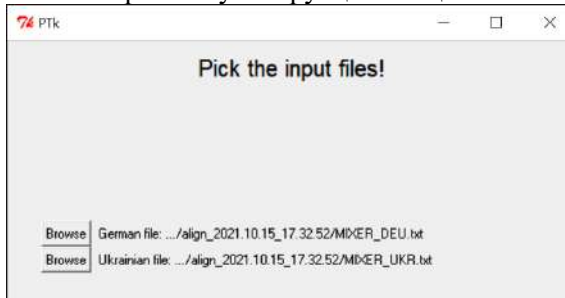
2.3. Оберіть тип формату підготовлених файлів, які Ви збираєтесь завантажувати, – .txt (UTF-8), натисніть Next:



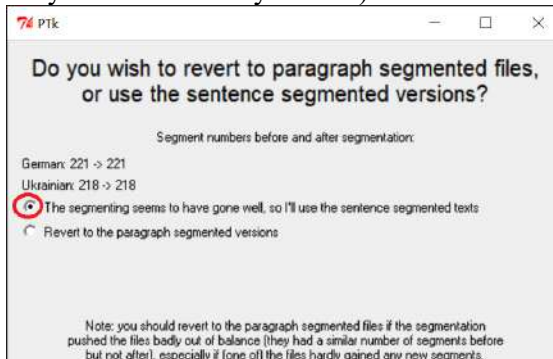
2.4. Вкажіть мови файлів: для цього оберіть зі списку мов спершу вихідну – «German», потім мову перекладу – «Ukrainian».



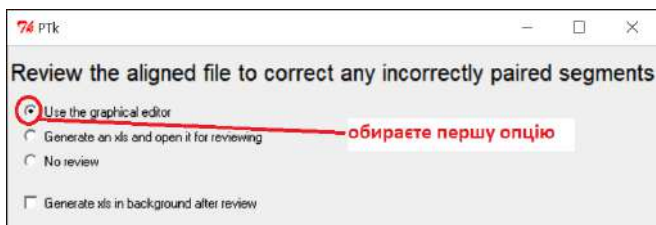
2.5. Завантажте Ваші підготовлені файли у форматі txt по черзі, починаючи з оригіналу інструкції німецькою мовою:



2.6. Оберіть першу відповідь на запитання про сегментацію тексту (пропонується за замовчуванням):



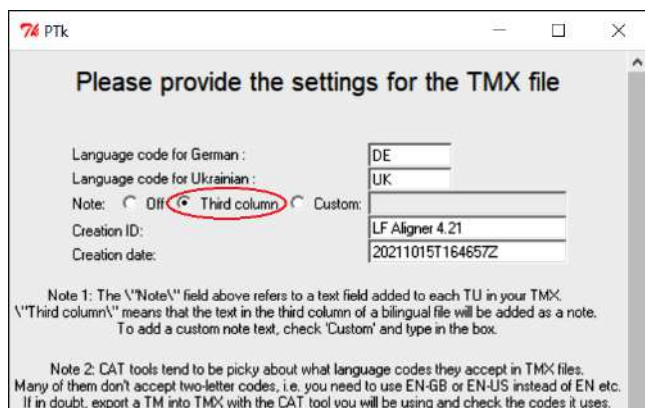
2.7. Оберіть можливість редагування у вирівняному паралельному тексті:



2.8. Дайте згоду на створення файлу з вирівняним паралельним текстом у форматі .tmx ():



2.9. Виберіть структуру сегментації за трьома стовбцями:



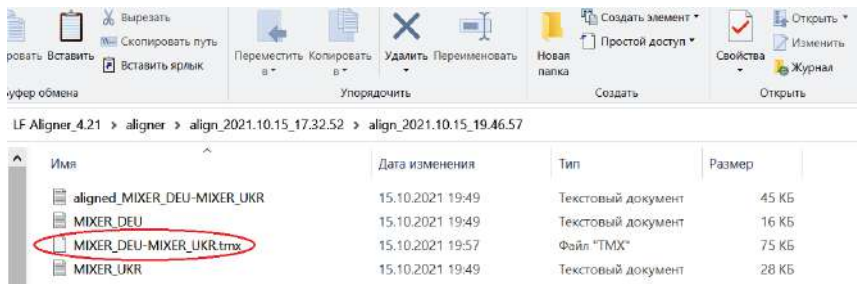
2.10. Після кількох хвилин обробки даних Ваші завантажені файли з інструкцією двома мовами згенеруються автоматично в один файл .tmx, про що Вас повідомить програма у віконечку наприкінці роботи. Натисніть ОК та вихід з програми.

2.11. У вирівняному паралельному тексті перевірте кожен сегмент на відповідність перекладу, оскільки можливі помилки. Виправіть неточності, пересуваючи або ж розділяючи сегменти за допомогою кнопок внизу документа і обов'язково збережіть усі зміни у файлі:

LF Alignment Editor 1.8 - aligned_MIXER_DEU-MIXER_UKR.txt

File	Edit	Help
53	Kontrollleuchte Gerät/Betriebszustand	Сигнальна лампочка MIXER_DEU
54	Rotes - Blinken Gerät ist betriebsbereit.	Пристрій/робочий стан MIXER_DEU
55	Grün - Gerät ist entriegelt (Sicherungstaste ist gedrückt) und kann eingeschaltet werden.	Влимає червоним - Пристрій готовий до використання. MIXER_DEU
56	???????	Світиться зеленим - Пристрій розблоковано (кнопку розблокування перемикача натиснуто) і може бути ввімкнено. MIXER_DEU
57	Rot - Frühwarnung vor Überhitzung.	Світиться червоним - Заздегледіне попередження про перегрівання. MIXER_DEU
58	Der Motor wird nach und nach langsamer und schaltet sich dann aus. Lassen Sie das Gerät abkühlen.	Мотор уповільнюється крок за кроком, до повного вимкнення. MIXER_DEU
59	Es kann wieder verwendet werden, wenn die Kontroll-Leuchte rot blinkt.	Дозвольте пристрою охолонути. MIXER_DEU
60	Wenn die maximale Geschwindigkeit bei hoher Last und langer	Пристрій знову готовий до роботи, коли сигнальна лампочка заблимає MIXER_DEU

2.12. Відкрийте папку, яку програма *LF Aligner* автоматично створила після завершення процесу вирівнювання паралельних текстів. У ній Ви побачите чотири документи, один з яких (tmx-файл) можна використовувати для створення бази пам'яті перекладу, або ж для роботи з корпусами текстів.



Крок № 3: Підключення створеної бази пам'яті в програмі автоматизованого перекладу *MateCat* до нового перекладацького проєкту.

Вирівняний паралельний текст у форматі tmx-файлу (що власне і є нашою базою пам'яті) можна завантажувати в будь-яку програму автоматизованого перекладу, щоб ефективніше і якісніше перекладати необхідний текст: Trados Studio, DéjàVuX, memoQ та ін. (на платній основі); або ж працювати з безкоштовними CAT-програмами онлайн: Wordfast Anywhere (Сайт: <https://www.freetm.com>), MateCat (Сайт: <https://www.matecat.com>). Для роботи в програмі MateCat онлайн навіть не обов'язково реєструватися.

3.1. Перейдіть за посиланням <https://www.matecat.com> на домашню сторінку MateCat (у Chrome, Edge або Safari). Можна використовувати платформу безкоштовно як анонімний користувач, однак реєстрація надає додаткові переваги, такі як збереження приватної пам'яті перекладів і глосаріїв на невизначений термін, а також отримання та керування проектами зі сторінки «Мої проекти». Щоб увійти, натисніть значок профілю у верхньому правому куті головної сторінки. З'явиться спливаюче вікно, у якому Ви можете увійти в Matecat за допомогою кнопки «Увійти через Google» або зареєструватися, використовуючи іншу адресу електронної пошти та пароль. Увійшовши в Google, ви миттєво отримаєте доступ до Matecat.



3.2. Створіть новий перекладацький проект на платформі MateCat. Для цього виконайте наступні пункти:

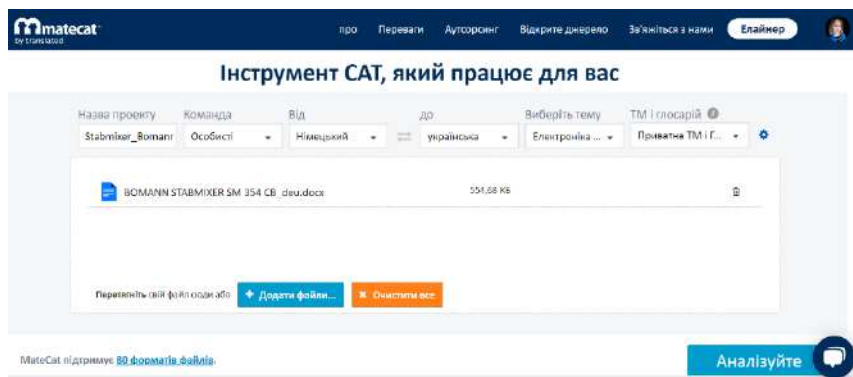
1) у полі «**Назва проекту**» введіть назву свого нового перекладацького проекту, наприклад «*Stabmixer_Bomann*»;

2) якщо Ви працюєте самостійно над перекладом, то далі у випадяючому меню **«Команда»** залиште за замовчуванням опцію **«Особисті»**, у випадку командної роботи над перекладацьким проектом можна в цьому ж полі задати пошук інших перекладачів і запропонувати спільно працювати над замовленням;

3) наступний крок – вибір мовної пари (мова вихідного тексту та мова перекладу). Ви також можете вибрати кілька цільових мов, натиснувши **«Кілька мов»** у спадному меню цільової мови;

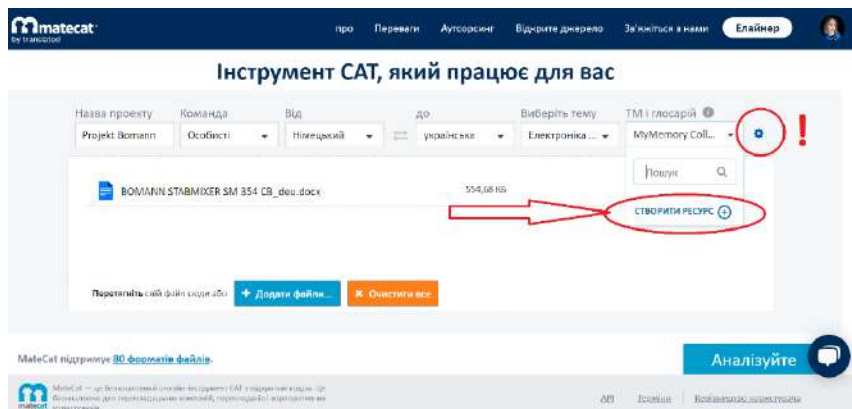
4) залежно від тематичного спрямування вихідного тексту в полі **«Виберіть тему»** знайдіть відповідну галузь, в якій здійснюватимете переклад, наприклад **«Електроніка/електротехніка»**;

5) завантажте файл(и) для перекладу в будь-якому з 78 доступних форматів файлів. Можна безпосередньо перетягнути файл(и) або знайти їх, натиснувши кнопку **«+ Додати файли»**. Ви можете створювати проекти з 150 файлами з обмеженням розміру 100 МБ кожен і перекладати їх без будь-яких затримок у програмі. Завантажувати можна файли, які не перевищують ліміт у 250 тисяч слів. Якщо файл більший, ви побачите таке повідомлення: **«Заборонено. Matecat не може створити проект»**. Якщо потрібно видалити завантажений файл, натисніть значок смітника. Якщо ви хочете видалити всі завантажені файли, ви можете натиснути кнопку **«Очистити все»**;

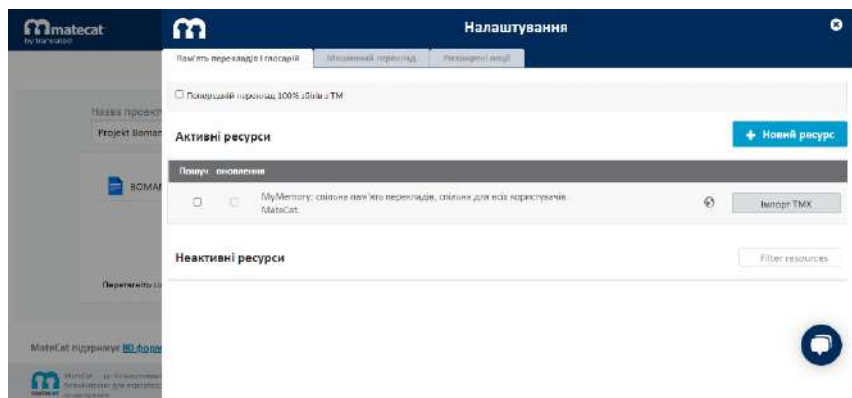


3.3. Створіть нову пам'ять перекладів (ТМ) та/або Глосарій або виберіть існуючу безпосередньо на вкладці **«ТМ і глосарій»**. 3

опції «**MyMemory Collection**» можна обирати приватну ТМ лише за умови існування особистої колекції пам'яті перекладів. У разі створення першого перекладацького проєкту на цій платформі оберіть опцію «**Створити ресурс**», або ж натиснути на значок шестірні і налаштувати параметри проєкту:



При цьому відкриється панель «**Налаштування**», де можна додавати або редагувати параметри проєкту.

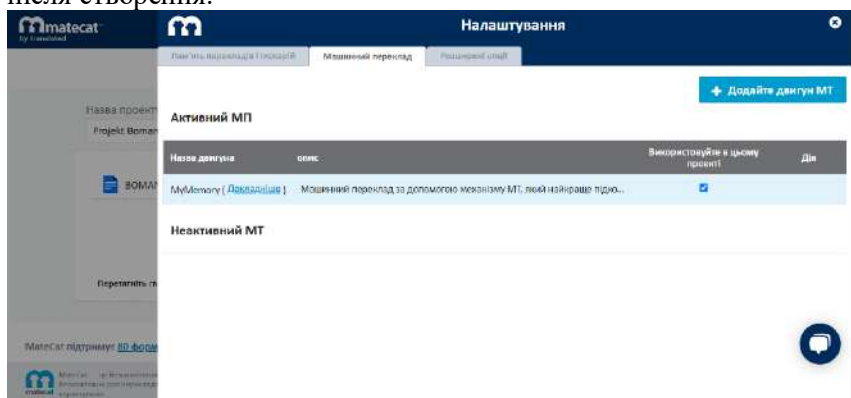


Панель налаштувань містить різні вкладки.

Вкладка «**Пам'ять перекладів і глосарій**»: тут ви можете створити одну або кілька приватних пам'яті перекладів, щоб

забезпечити конфіденційність своїх перекладів, або використовувати наявні. Також можна вибрати «**Попередній переклад 100% збігів із ГМ**», встановивши відповідний прапорець.

Вкладка «**Машинного перекладу**»: тут ви можете увімкнути або вимкнути механізм машинного перекладу, знявши відповідний прапорець. За замовчуванням МуMemory увімкнено. Ви можете вибрати іншу машину МТ зі своїх неактивних ресурсів або взагалі не використовувати МТ. Зверніть увагу, що МТ можна вимкнути лише під час створення проекту, його неможливо деактивувати після створення.



Вкладка «**Розширені опції**»: тут ви можете активувати або деактивувати наступні параметри.

Диктант: компонент перетворення мовлення в текст, який диктує ваші переклади, а не вводить їх.

Вгадати положення тега (мітки): зручний інструмент, який дозволяє Matecat автоматично розміщувати мітки там, де вони належать, і заощадити вам додаткові зусилля.

QA від lexiQA: увімкне автоматичні перевірки якості, надані lexiQA для мов, які наразі підтримуються.

Лічильник символів: ця функція дозволяє вставити лічильник символів, які використовуються в цільовому сегменті. Це може бути корисним, якщо під час перекладу потрібно дотримуватися обмеження кількості символів.

Міжмовні збіги: дозволяє пропонувати переклади іншими цільовими мовами, які ви знаєте, як посилання.

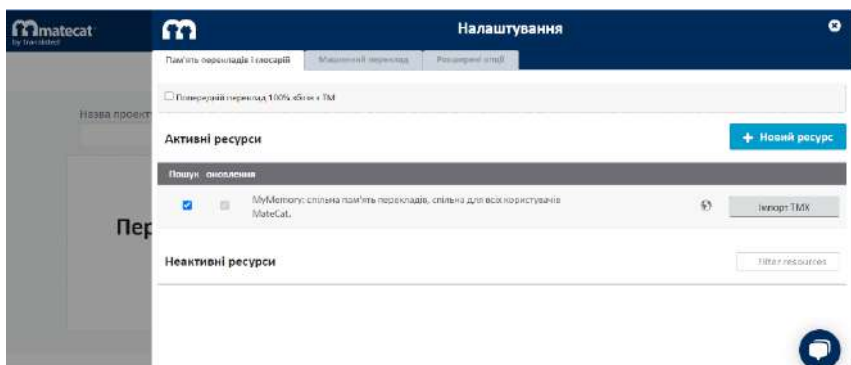
Правила сегментації: щоб вибрати спосіб поділу речень відповідно до певних типів вмісту. Зазвичай Matecat сегментує речення там, де він виявляє сильний пунктуаційний елемент (. ! ?). Це допомагає регулювати читабельність речення під час перекладу до певних типів тексту.

Повернемося до вкладки налаштування «**Пам'яті перекладів і глосаріїв**».

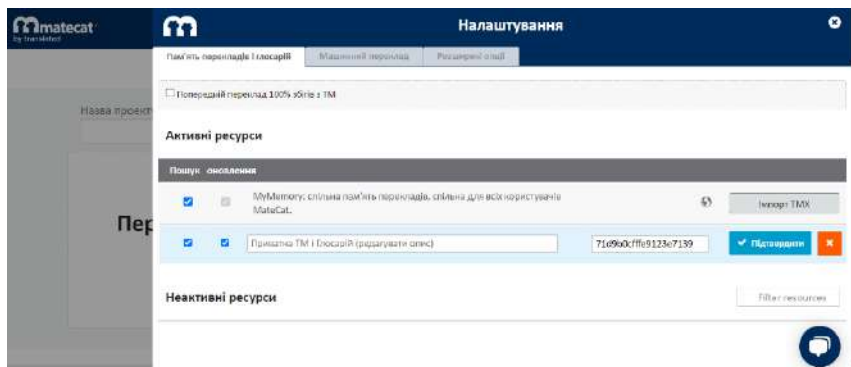
Під час перекладу ваші переклади зберігаються як двомовні речення в так званій пам'яті перекладів (TM) – це свого роду база даних, що містить двомовні сегменти (речення, набори слів або окремі слова, уже перекладені однією чи кількома мовними парами), які можна використовувати під час перекладу того самого проекту чи майбутніх проектів. Отже, під час перекладу, якщо ви знайдете сегмент, ідентичний або схожий на інший, який уже перекладено, пам'ять перекладів запропонує формулювання вже перекладеного сегмента, щоб ви могли використати його, частково чи повністю, для перекладу новий сегмент.

У Matecat ви можете зберігати свої переклади:

1) у **загальній загальнодоступній пам'яті перекладів**, до якої можуть отримати доступ через API усі користувачі Matecat. Якщо ви збережете свої переклади в загальнодоступній TM, усі користувачі Matecat зможуть переглядати їх як пропозиції. Це стане людським внеском у загальнодоступну TM і допоможе іншим користувачам Matecat під час перекладу в Matecat.



2) у **новій** або вже **існуючій приватній пам'яті перекладів**, призначеній одному чи декільком проектам або конкретному клієнту, що дає змогу зберегти ваш переклад у приватній пам'яті перекладів, до якої зовнішні користувачі не можуть отримати доступ. Щоб створити нову приватну пам'ять перекладів, натисніть «+Новий ресурс» в налаштуваннях проекту.



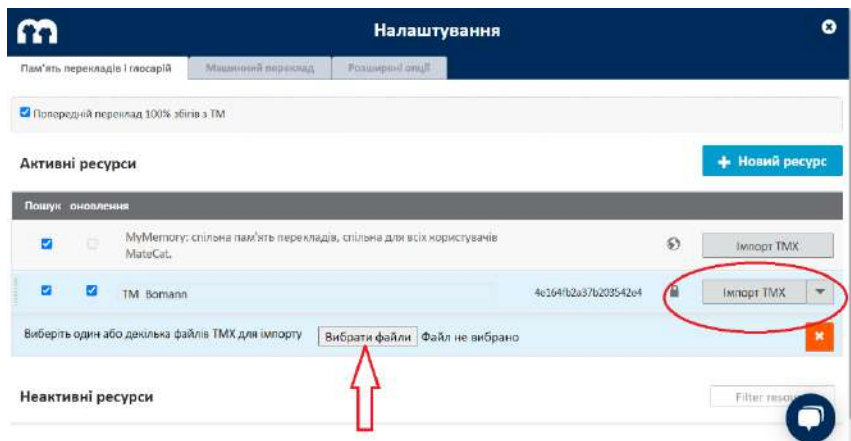
У Matecat кожна приватна ТМ ідентифікується унікальним буквено-цифровим значенням під назвою private TM Key. Це єдине значення, яке вам потрібне, щоб пов'язати вашу ТМ з вашим проектом. Він працює як пароль для доступу до вашої приватної ТМ у механізмі пам'яті перекладів Matecat.

Якщо ви зберігаєте свої переклади в приватній пам'яті перекладів, вони зберігатимуться лише у вашій приватній пам'яті перекладів, і тільки ви маєте ексклюзивне використання та доступ до своїх внесків до цієї приватної пам'яті перекладів.

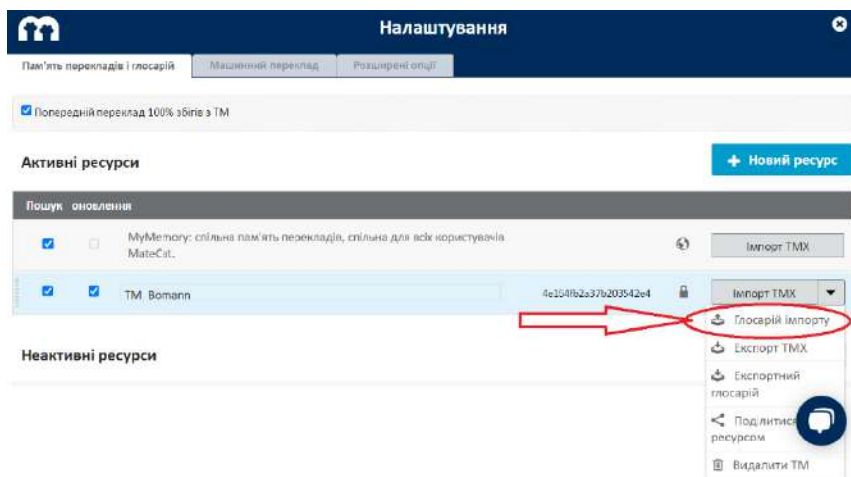
Якщо ви працюєте над проектом, створеним іншим користувачем (наприклад, вашим РМ), або ви працюєте з колегою, ви можете побачити пам'ять перекладів, яка вам не належить: вона виділена сірим кольором і покрита зірочками.

Далі натисніть «**Підтвердити**» створення приватної ТМ (перекладацької пам'яті і глосарію) з особистим кодом; назву нової ТМ рекомендовано перейменувати, наприклад *ТМ_Mixer_Vomann*, бо за замовчуванням система запропонує загальне формулювання «Приватна ТМ і глосарій».

3.4. Спробуйте імпотувати приватну пам'ять перекладу у форматі tmx-файлу з вирівняним паралельним текстом інструкції, яку Ви підготували на попередньому занятті. Для цього клацніть на «Імпорт ТМХ», виберіть підготовлений tmx-файл, завантажте власну ТМ зі свого комп'ютера і натисніть «Підтвердити».



Також можна завантажити власний глосарій у форматі xlsex для цього клацніть на випадаюче меню «Імпорт ТМХ», виберіть «Глосарій імпорту» та підтвердіть свої дії.



3.5. Закрийте налаштування проекту і поверніться на домашню сторінку Matecat зі створеним новим проектом і просто натисніть «Аналіз», щоб запустити аналіз обсягу для Вашого проекту.

3.6. Після завершення процесу аналізу ви можете натиснути кнопку «Перекласти», щоб перейти безпосередньо до редактора перекладу та перекласти самостійно.

Рекомендовані онлайн-ресурси для виконання лабораторної роботи:

1. Работа в MateCat:

<https://www.youtube.com/watch?v=J1EpH39HVtM>

<https://www.youtube.com/watch?v=B-wwDw62nEs>

<https://www.youtube.com/watch?v=A3qlaPwsQ4M>

Як завантажити глосарій: <https://www.youtube.com/watch?v=loILq0BYW7c>

Як створити ТМ: https://www.youtube.com/watch?v=dd3rEU_FUSw

2. Вирівнювання паралельних текстів. Паралельні корпуси. Пам'ять перекладів:

<https://www.youtube.com/watch?v=hXNxNK2nOo4&t=1260s>

ТЕМА 4 Локалізація перекладу

План лекції:

- 4.1. Переклад, локалізація, адаптація: основні відмінності
- 4.2. Методи локалізації веб-сайтів
- 4.3. Лінгвістичні та екстралінгвістичні рівні локалізації

4.1. Переклад, локалізація, адаптація: основні відмінності

Переклад і локалізація текстів для іншомовної аудиторії є однією з значущих проблем сучасного перекладознавства. Без інтернет-контенту вже не можливо уявити різні сфери діяльності людини, адже саме завдяки цьому відбувається діалог культур, ознайомлення одних народів з досягненнями, менталітетом і традиціями інших [Дольник; Галій 2021, с. 142].

Локалізація перекладу – поняття відносно нове, проте процес,

який вона передбачає, використовується перекладачами вже не перше сторіччя. Про значення терміна здогадатися нескладно. Локальний – значить місцевий, локалізація – процес, спрямований на те, щоб зробити що-небудь місцевим, в даному випадку – переклад.

Близьким за значенням до терміну «локалізація» можна назвати широко відомий термін «адаптація». Адаптація – поняття більш широке, воно передбачає пристосування тексту на всіх рівнях мови – включаючи граматичний, лексичний та стилістичний. А локалізація, в свою чергу – процес адаптації іноземного тексту до культурного контексту країни, мовою якої здійснюється переклад. Іншими словами, можна сказати, що локалізація – це один з видів адаптації.

Отже, локалізація – це динамічний процес адаптації програмного забезпечення, Веб-сайтів або супутньої технічної документації до стандартів і мовних традицій країни, для якої призначається згадане програмне забезпечення.

Характеризується глибинним аналізом оригіналу, високим рівнем перекладацької майстерності локалізатора, розумінням потреб замовника та цільового користувача локалізованого продукту, а також відрізняється наявністю низки екстралінгвістичних факторів, не притаманних класичному письмовому перекладу [Долинський; Скиба 2016, с. 22].

Термін «локалізація» став використовуватися в лінгвістиці порівняно нещодавно, прийшовши зі сфери бізнесу і міжнародного технічного співробітництва. Під локалізацією спочатку розумілася повна адаптація продукту до умов використання в конкретному регіоні, який розглядається в нерозривному зв'язку з мовою цього регіону. У лінгвоіндустрії переклад розглядається лише як окремий етап локалізації. Між іншим, в традиційній теорії перекладу, яка враховує фактор адресата, адаптація тексту до потреб і очікувань одержувача розглядається як одна з власне перекладацьких процедур [Шокина 2017, с. 86].

Якщо перекладацькі трансформації зачіпають конкретні фрагменти тексту, то локалізація є загальною перекладацькою процедурою, свого роду стратегією, яка визначає вид тексту в цілому. Тобто вона не обмежується перекладом окремих текстових складових та є формою міжкультурної комунікації, успішність якої можлива за умови врахування соціокультурних компонентів мови користувачів.

Створюючи іншомовну версію сайту, власники хочуть отримати: збільшення числа відвідувачів, що приходять з пошукових систем; нових клієнтів з інших країн; збільшення доходів. А основним завданням бюро перекладів є отримання коштів за свою роботу. Виходить, що у сторін різні цілі. Перекладачів за великим рахунком не обходить зростання прибутку сайту замовника, адже вони з неї не отримають додаткової оплати.

Щоб правильно виконати локалізацію сайту, кваліфіковані фахівці бюро перекладів повинні мати уявлення про основні принципи SEO. Перекладаючи інформацію іншою мовою, вони повинні також розуміти, що таке внутрішня оптимізація веб-сайту. Тобто основним завданням перекладача є гармонійне вписування перекладених ключових слів у певні частини тексту.

Коли користувач шукає потрібну йому інформацію, він вводить запит з одного або декількох слів в пошукову систему, яка видає найбільш релевантні сторінки сайтів, що входять в ТОП 10 Гугла. Але ж вони в цій видачі виявилися не просто так. Будь-який сайт створюється людьми, які знають, що таке оптимізація, просування і підвищення релевантності.

При оцінці релевантності сторінок сайту пошукові роботи виділяють певні елементи тексту, які є ключовими словами при видачі сайту за запитом користувача. Саме такими знаннями повинен володіти перекладач для локалізації сайту на будь-яку мову.

Локалізація актуальна для інформаційних, політичних сайтів, а також для різних компаній, оскільки це відкриває для них нові можливості, залучає нових клієнтів, особливо якщо переклад зроблено з урахуванням особливостей менталітету закордонних партнерів. 74,5 % користувачів Інтернету ніколи не здійснюють покупок на англomовних сайтах або роблять їх зрідка. З них 72,4 % зізнаються, що з більшою ймовірністю здійснять покупку, якщо інформація на сайті буде подана їхньою рідною мовою [Долинський; Скиба 2016, с. 23]. Локалізація сайту передбачає:

1. Збір та впорядкування наявного контенту (змісту сайту) оригінальною мовою, очищення його від тегів тощо.
2. Створення та затвердження єдиного глосарію (з метою ідентичного перекладу вживаних на різних сторінках сайту термінів).
3. Переклад усіх сторінок сайту цільовою мовою.

4. Технічне налаштування роботи іншомовної версії сайту або його копії на іншому домені (включно з підбором домену).
5. Розміщення текстів на сайті через його систему керування вмістом (CMS).
6. Перелінкування контенту.
7. Переклад окремих елементів і локальна пошукова оптимізація сайту (тайтли, зображення, ключові слова).
8. Повна вичитка та коректура новоствореної мовної версії.



Рис. 4.1. Схематичне зображення процесу локалізації програмного забезпечення

Локалізація потребує залучення висококваліфікованих спеціалістів з галузі лінгвістики і програмування: перекладачів, редакторів, веб-дизайнерів, веб-програмістів, тестувальників і керівників проекту. Локалізація програмного забезпечення – не просто переклад, це процес адаптації програмного забезпечення до культури певної країни, який може включати в себе:

- ✓ аналіз локалізованості;
- ✓ укладення багатомовного словника проекту;
- ✓ локалізацію інтерфейсу і системних повідомлень;
- ✓ локалізацію довідкових файлів;
- ✓ складання локалізованої версії;
- ✓ локалізацію супутніх маркетингових матеріалів;
- ✓ лінгвістичне і функціональне тестування.

Очевидно, що локалізація – дуже складний процес, що вимагає, крім знання мови ще і знання з різних сфер особливостей культури, традицій і звичаїв, політичного ладу інших країн. Тому, перш ніж починати будь-який переклад, необхідно зібрати як можна більше соціокультурної інформації і, бажано, з першоджерел [Долинський; Скиба 2016, с. 24].

4.2. Методи локалізації веб-сайтів

1. *Локалізація пошукової оптимізації*. Включати локалізацію пошукової оптимізації в процес локалізації веб-сайту – виправданий метод. Адже більшість компаній в першу чергу локалізують свій веб-сайт, а потім починають думати про пошукову оптимізацію і операції. Це займає більше часу і енергії, і такий підхід менш ефективний. Бажано планувати спочатку оптимізацію пошукових систем сайту, щоб процес локалізації був ефективнішим. Також необхідно створити свій початковий список ключових слів, які допоможуть оптимізувати зміст і мета-теги з найпоширенішими ключовими словами для місцевих ринків.

2. *Дослідження ринку*. Ще один цікавий метод – почати дізнаватись більше про конкретну місцевість, де зосереджені користувачі сайту. Після цього важливо врахувати цінність дослідження ринку для того, щоб дізнатися більше про ділові можливості. Звички споживачів змінюються відповідно до регіону, у яких функціонує спільна мова, тому дослідження місцевого ринку також потрібні.

3. *Графіка веб-сайту*. Також не слід ігнорувати в цьому випадку важливість дизайну. Змінна конструкція буде корисною, оскільки вона дозволяє наповнити сторінку динамічним контентом. У порівнянні з англійською мовою, деякі мови мають довші або коротші слова. Наприклад, більшість азійських мов мають слова, які могли б складати менше половини довжини слова з вихідної мови. Через численні труднощі, з якими стикаються азійські мови, можливо, буде потрібно збільшити шрифти до 120%, щоб переконатися, що люди можуть прочитати слова з легкістю. Так що, як правило, краще уникати використання фіксованого розміру комірки з текстом в ній. Варто переконатися, що простір можна змінити або переробити для того, щоб врахувати особливості шрифту.

4. *Протестуйте сайт на аудиторії*. Зі зрозумілих причин, важливо перевірити сайт в країні, для якої він призначений. Більшість престижних бюро перекладів роблять це для того, щоб переконатися, що характеристики веб-сайту насправді збігаються з перевагами і культурними особливостями цільової аудиторії. Таким чином можна бути впевненим у тому, що сайт стане популярним і його цінуватимуть за професіоналізм.

5. *Глосарії*. Глосарії існують для того, щоб забезпечити правильність і точність в процесі перекладу і локалізації, тому глосарії дійсно важливі і не повинні бути проігноровані. Перекладач повинен взяти до уваги той факт, що клієнт очікує, що він буде використовувати новітні галузеві словники, а також варто переконатися, що глосарії відповідають місцевості.

6. *Оптимізація під місцеві соціальні мережі*. Також рекомендується підключити сайт до місцевих соціальних мереж. Варто додати кнопки соціальних мереж, які широко використовуються серед цільової аудиторії, для того, щоб гарантувати зв'язок з місцевими соціальними мережами. Деякі країни, такі як Китай, мають різні соціальні мережі, і дуже важливо переконатися, що сайт підключений до цих конкретних мереж. Це допоможе отримати пошвавлення на сайті, оскільки користувачі зможуть легко обмінюватися контентом, пропозиціями та ідеями через їхні профілі.

7. *Системи управління контентом*. З поширенням перекладених і локалізованих сайтів для всіх цільових глобальних ринків, тепер практичніше використовувати системи управління контентом, які здатні обробляти локалізації та глобалізації кількох багатомовних веб-сайтів. Система управління контентом являє собою безліч процедур або сучасний веб-інструмент, який використовується для управління робочим процесом у середовищі спільної роботи. Вони допомагають розвивати і підтримувати загальну інформацію, як правило, не тільки для веб-сайтів, а й для документів і контролю вмісту. Системи управління контентом також використовуються для розширення можливостей декількох користувачів стосовно автора і редагувати та обмінюватися інформацією в різних форматах.

8. *Місцезнаходження та вибір мови*. Відвідувачі сайту повинні мати можливість легко вказати їх місце розташування і мову сайту. Тому спеціальні значки повинні бути включені в дизайн сайту, щоб вказати ці варіанти. Є багато макетів і форматів, тому завжди є з чого вибрати. Попередні дослідження на стадії відкриття проекту мають вказати, який тип індикатора мови буде найбільш підходящим.

Отже, очевидно, що локалізація веб-сайту має життєво важливе значення для будь-якого бізнесу, який прагне перейти на міжнародні

ринки. Стає все більше і більше компаній, які прагнуть вступити на ці прибуткові ринки, тому локалізація сайтів стала дуже популярним і прибутковим явищем. Ділова сторона включає глобальну і локальну координацію; виробництво і операції, включаючи фінансування; продажі та маркетинг; мовний переклад, в тому числі технології та лінгвістичну координацію; розробку програмного забезпечення, а також дизайн. Чим вище технологічні можливості цільової культури, тим більш імовірно, що локалізація веб-сайту буде ефективнішою.

4.3. Лінгвістичні та екстралінгвістичні рівні локалізації

Сюзанне Гьопферіх виокремлює два рівні локалізації на рівні тексту: поверхнева локалізація та глибинна локалізація [Görferich 2008, с. 335].

Перший рівень локалізації стосується мови й передбачає функціональну відповідність оригіналу й перекладу (еквівалентний переклад) та адаптацію перекладу для адресата відповідно до теорії скопосу [Reiß 1984]. Згідно цієї теорії переклад оцінюється залежно від ступеня досягнення перекладачем поставленої перед ним мети.

Прикладами поверхневої локалізації є адаптація одиниць чужої метричної системи до національної, подання дат, часу та чисел в тексті у формі, конвенціоналізований в межах національної культури.

Окрім приклади поверхневої локалізації ілюструють таблиці 1, 2.

Табл. 1

Національні стандарти для дати і часу

Міжнародний стандарт		ФРН	США
Дата	Рік/місяць/день – 2021-08-12 (ISO 8601)	день/місяць/рік 08.12.2021	місяць/день/рік 12.08.2021
Час	hh:mm:ss – 23:15:20 (ISO 8601)	15.10 Uhr або 15:10 Uhr	15.10 P.M.

Табл. 2

Національні стандарти для окремих чисел

ФРН	Million	Milliarde	Billion	Billiarde
США	million	Billion	trillion	quadrillion

Другий рівень локалізації скерований на адаптацію тексту до специфічної для певної культури ментальності, системи цінностей, конвенціоналізованих зразків поведінки у певній комунікативній ситуації тощо. Напр., інструкції з експлуатації японською мовою детально описують функціональність продукту та створюються у вигляді коміксів, що для користувачів з ФРН неприпустимо. Тобто, поверхнева локалізація відбувається на мовному рівні й граничить з перекладом, а глибинна – на рівні адаптації культурних диференцій, що передбачає інший спосіб подання інформації в тексті.

На *екстралінгвальному рівні* локалізація з використанням невербальних засобів широко використовується для адаптації програмного забезпечення, що передбачає застосування у ньому унормованих у певній країні піктограм та символів, типу та розміру шрифтів, вікон відповідного розміру та їхнього розташування у рядках меню. Для локалізації програмного забезпечення користуються спеціалізованими програмними продуктами, напр., Passolo, Corel CATALYST, KT International Superlinguist та ін., які дозволяють переклад діалогових повідомлень безпосередньо в програмних продуктах та тестування локалізованого продукту.

Для перекладу аудіовізуальних матеріалів використовують спеціалізовані інструменти, напр., Swift, яка об'єднує у собі окремі компоненти пам'яті перекладів, але додатково дозволяє створення субтитрів та їхнє форматування на екрані з дотриманням відео-стандартів.

Рівень локалізації визначають такі чинники як популярність продукту серед споживачів; наявність чи відсутність конкурентів, які пропонують подібну продукцію; дефіцит продукції.

Якщо продукція користується попитом серед споживачів і на ринку попит перевищує пропозицію, виробник не витратить значні кошти на його локалізацію. Крім того, популярність західної продукції у країнах третього світу та країнах, що розвиваються, часто спонукає виробників створювати технічну документацію й упаковку для продуктів таким чином, щоб забезпечити впізнаваність їх іноземного походження і, у такий спосіб, збільшити обсяг продажу. Якщо ж на ринку існують інші аналогічні товари чи послуги від компаній-конкурентів, які добре адаптовані до споживчого ринку, виробникові доведеться ретельно адаптувати та популяризувати власний продукт, щоб завоювати свою частку

ринку. Тому, перед створенням технічних текстів для представників іншої культури, технічний автор повинен ретельно проаналізувати культурні особливості країни-збуту, шкалу соціальних цінностей, мати відповідні фонові знання й володіти навичками міжкультурної комунікації.

Під терміном "інтернаціоналізація" розуміють процес створення фахового тексту незалежно від національної традиції. Основна ідея інтернаціоналізації тексту полягає в наступному:

1) заощадження коштів на локалізації технічної документації до національних стандартів ринків-збуту;

2) зрозумілість тексту за межами власної культури;

3) можливість повторного використання окремих сегментів тексту для публікації (DTP Desktop Publishing System), CMS (Content Management System) та перекладу (TMS Translation Memory System) різного типу документації на продукт.

Прикладом інтернаціоналізації різного типу текстів для полікультурної публіки є веб-сторінки, в процесі створення яких рекомендують дотримуватись наступного [Göpferich 2008, с. 340-341]:

1) для створення контенту використовувати сучасну контрольовану англійську мову, що базується на найбільш частотних лексичних одиницях, спрощених граматичних структурах для забезпечення його розуміння користувачам, які володіють англійською мовою як іноземною і не завжди на високому рівні;

2) уникати уживання модних слів, сленгу, неусталених скорочень або синтаксичних скорочень, які можуть бути незрозумілими для інтернаціональної публіки [Міщенко, с. 153];

3) подавати номери телефонів згідно міжнародних стандартів: з міжнародним кодом країни та регіону, напр.:

Україна: +38 044 5442312 (38 – міжнародний код України)

ФРН: +49 (0)721 9683733 (49 – міжнародний код ФРН)

4) уживати дати в усталеній для країни формі згідно норми ISO 8601.

5) подавати час у 24-годинному форматі згідно середньоєвропейського часу (UTC = Universal Time Coordinatet), норма ISO 8601, напр.: 23:12 UTC.

6) уживати одиниці виміру відповідно до усталеної метричної системи.

7) грошові суми вказувати із зазначення національної валюти згідно норм ISO 4217, напр.: \$ 24 USD, € 24, £ 24.

8) писати адреси згідно усталеного зразка із точним заповненням відповідних полів, напр.:

- Ім'я
- Організація
- Вулиця
- Місто
- Країна
- Поштовий індекс

9) уникати уживання неусталених на міжнародному рівні форм звертання до осіб чоловічої та жіночої статі, які можуть зумовлювати комунікативні конфлікти, напр.:

Англійська мова: *s/he*:

Німецька мова: *Innen: Liebe StudentInnen!*

10) по можливості, замінювати вербальну інформацію невербальними усталеними символами, знаками або піктограмами, які дозволяють заощаджувати кошти на перекладі й зрозумілі носіям інших мов, напр.: піктограми операційної системи Windows дозволяють працювати у ній з різними мовами інтерфейсу без знання цих мов.

Засоби невербальної комунікації можуть варіюватись залежно від регіону, але в контексті міжкультурної комунікації важливим фактором для полікультурної публіки є можливість їх однозначної ідентифікації з референтом, визначення їх ілокуції та предикації.

Прикладом невербальних засобів такого типу можуть слугувати піктограми у міжнародних аеропортах та на міжнародних вокзалах, які вживаються на позначення аеропорту, ескалатора, туалету, ресторану, камери схову, поліцейного офісу тощо [Мищенко, с. 154].

Стандартизуються не тільки зображення символів, але й їхня форма та колір. Так, наприклад, червоний колір викликає збудження та занепокоєння. Тому червоні символи сигналізують небезпеку й використовуються для знаків, що вказують на небезпеку для життя чи здоров'я людини або пов'язані з гасінням вогню. У всіх заборонних знаках присутній червоний колір, який здебільшого комбінується з білим. Жовтий та помаранчевий кольори також викликають збудження, але їм властива особливість привертати

увагу. Тому комбінація чорної та помаранчевої фарб використовується у випадках, коли необхідно попередити про небезпеку. Крім того, ці фарби добре видно і вони спонукають до обережності. Синій колір впливає на людину заспокійливо. Тому синій колір у комбінації з білим використовується для передавання інформації, експлікації вимог чи рекомендацій. Характерна властивість зеленого кольору також заспокійлива. Тому комбінування зеленого й білого кольорів застосовують для позначення безпечних місць, а також для вказування на установи з надання першої допомоги хворим чи потерпілим, напр.:



Рис. 4.2. Приклад локалізації піктограм

Але створюючи багатомовні сайти, дизайнери веб-сторінок повинні враховувати значення кольорів у конкретній країні чи регіоні, оскільки нехтування національними традиціями не виключає можливості комунікативного конфлікту.

Проте, в процесі невербальної комунікації на міжнародному рівні слід обережно використовувати піктограми, жести та міміку. Це зумовлено тим, що для кожної культури характерні специфічні норми поведінки, жести та міміка у певних комунікативних ситуаціях як реакція на почуте, сказане, зроблене іншими учасниками комунікації. Їхня функція полягає у передаванні невербальної інформації і головне тут, не які жести використовує мовець, а як вони інтерпретуються іншими учасниками комунікації. Тому, крім конвенціоналізованих технічних жестів, які використовуються поліцейськими, пожежниками, персоналом аеропортів та залізниць, станцій метро, які також сьогодні унормовуються й стандартизуються на міжнародному рівні, незнання значення специфічних для певної культури сигналів в

процесі міжкультурної комунікації, як правило, призводить до комунікативного конфлікту.

Використана література та інформаційні ресурси:

1. Долинський Є.В., Скиба К.М. Інформаційні технології у професійній діяльності перекладача : навчальний посібник з інформаційних технологій для студентів першого (бакалаврського) рівня, а також другого (магістерського) рівня вищої освіти, галузі знань «Гуманітарні науки», спеціальності «Філологія», «Переклад». Хмельницький : ХНУ, 2016. 195 с.

2. Дольник І. М., Галій Л. Г. Проблеми локалізації українських веб-сайтів (на матеріалі веб-сайтів українських бюро перекладів). Фаховий та художній переклад: теорія, методологія, практика: збірник наукових праць / за заг. ред. С.І. Сидоренка. К.: Аграр Медіа Груп, 2021. С. 142-148. <https://er.nau.edu.ua/bitstream/NAU/49890/1/2021%20GSTI%20Conference%20Proceedings-142-148.pdf>

3. Міщенко А. У контексті міжкультурної комунікації. Наукові записки. Серія: Філологічні науки. Вип. 104 (1). С. 151-158. <https://core.ac.uk/download/pdf/53035645.pdf>

4. Шокина В.И. Перевод и локализация текстов в сфере индустрии красоты (на материале франко- и русскоязычных сайтов). Вестник ВолГУ. Сер. 9. Вып. № 15. 2017. С. 86-89. <https://glebov.com.ua/zasoby-likalizatsii-saytiv.html>

5. Göpferich S. Textproduktion im Zeitalter der Globalisierung : Entwicklung einer Didaktik des Wissenstransfer / Susanne Göpferich. 3. Aufl. Tübingen : Stauffenburg Verl., 2008. 497 S.

6. Reiß K. Grundlegung einer allgemeinen Translationstheorie Katharina Reiß; Hans J. Vermeer, Niemeyer, 1984. 245 S.

Програми для локалізації додатків:

Multilizer. Сайт: <http://www2.multilizer.com/>

POEditor. Сайт: <https://poeditor.com/>

Програми для перекладу сайтів: CatsCradle.

Сайт: <http://www.stormdance.net/software/catscradle/overview.htm>

СЕМІНАРСЬКЕ ЗАНЯТТЯ 4

Теоретичні питання для обговорення:

1. Поясніть різницю між термінами "локалізація" та "адаптація"!

2. Які етапи передбачає локалізація веб-сайтів?

3. Яких висококваліфікованих спеціалістів потребує залучення процес локалізації веб-сайтів?
4. Що означає "локалізація програмного забезпечення"?
5. Які методи локалізації веб-сайтів Ви знаєте?
6. Які два рівні локалізації веб-ресурсів виокремлює Сюзанне Гьопферіх?
7. Проілюструйте на прикладах адаптацію одиниць чужої метричної системи до національної, подання дат, часу та чисел в тексті у формі, конвенціоналізований в межах національної культури!
8. Які спеціалізовані інструменти використовують для перекладу аудіовізуальних матеріалів?
9. Як називається процес створення фахового тексту незалежно від національної традиції?
10. Що варто брати до уваги перекладачеві при інтернаціоналізації різного типу текстів для полікультурної публіки (веб-сторінок)?
11. Яка роль засобів невербальної комунікації у процесі локалізації?
12. Чи важлива стандартизація не тільки зображення символів (піктограм), але й їхня форма та колір?

Практичні завдання:

Аналіз якості та типові помилки при локалізації веб-сайтів України для німецької цільової аудиторії (або ж навпаки: німецьких сайтів з україномовним контентом).

Інструкція до виконання: 1) здійсніть пошук веб-сайтів України, які мають німецькомовний контент; 2) оберіть за власним уподобанням один ресурс для аналізу; 3) підготуйте презентацію для демонстрації результатів дослідження.

Лабораторна робота 4:

Виконання перекладацького завдання на платформі MateCat. Студентам пропонується колаборативна робота над спільним проектом – технічний переклад німецької інструкції блендера Blender. У ході роботи над проектом студентам буде надана можливість використовувати створену базу пам'яті перекладів (Лабораторна робота 3), яка дозволить зберегти одноманітність при перекладі термінології та однорідність цільового тексту.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

Амеліна С. М., Тарасенко Р. О. Підготовка перекладачів у зарубіжних університетах до застосування інформаційних технологій: монографія. Київ : «Центр учбової літератури», 2017. 256 с.

Амеліна С.М. Особливості інформаційно-технологічної підготовки майбутніх перекладачів у вищих навчальних закладах Німеччини / С.М.Амеліна, Р.О.Тарасенко // Вища школа. 2014. №8. С. 94-100.

Баевский В.С. Лингвистические, математические, семиотические и компьютерные модели в истории и теории литературы. М.: Языки славянской культуры, 2001. 336 с.

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

Глуховцева К.Д. Лінгвістичний атлас лексики народного побуту українських східнослов'янських говірок. Луганська : Альма-матер, 2003. 182 с.

Гринчишин Д., Капелюшний А., Сербенська О., Терлак З. Словник-довідник з культури української мови. Львів: Фенікс, 1996. 368 с.

Долинський Є.В., Скиба К.М. Інформаційні технології у професійній діяльності перекладача : навчальний посібник з інформаційних технологій для студентів першого (бакалаврського) рівня, а також другого (магістерського) рівня вищої освіти, галузі знань «Гуманітарні науки», спеціальності «Філологія», «Переклад». Хмельницький : ХНУ, 2016. 195 с.

Дольник І. М., Галій Л. Г. Проблеми локалізації українських веб-сайтів (на матеріалі веб-сайтів українських бюро перекладів). Фаховий та художній переклад: теорія, методологія, практика: збірник наукових праць / за заг. ред. С.І. Сидоренка. К.: Аграр Медіа Груп, 2021. С. 142-148. <https://er.nau.edu.ua/handle/NAU/49890>

Жуковська В.В. Вступ до корпусної лінгвістики: навчальний посібник. Житомир: Вид-во ЖДУ ім. І. Франка, 2013. 142 с. http://eprints.zu.edu.ua/18909/1/korpusna_lingv.pdf

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.

Некряч Т.Є. Деталь у цілісній картині світу Е. Гемінгвея: 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_IIf_2014_1_8

Селегей В. Електронні словари і комп'ютерна лексикографія. [Електронний документ]. Режим доступу: www.lingvoda.ru

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.

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.

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

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.

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.

Фокін С.Б. Комп'ютерні інформаційні технології в перекладацькій та перекладознавчій діяльності. К.: Київ. нац. ун-т ім. Т.Шевченка, 2019. 254 с.

Шокина В.И. Перевод и локализация текстов в сфере индустрии красоты (на материале франко- и русскоязычных сайтов). Вестник ВолГУ. Сер. 9. Вып. № 15. 2017. С. 86-89. <https://glebov.com.ua/zasoby-likalizatsii-saytiv.html>

Навчальне видання

Інноваційні технології перекладу

Укладач: Антоніна КОРОЛЬ

© Чернівецький національний
університет, 2023