

**М. Л. Гелетка**

**ЦИФРОВІ ТЕХНОЛОГІЇ В ПЕРЕКЛАДАЦЬКІЙ ДІЯЛЬНОСТІ**

2023

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
Національний аерокосмічний університет ім. М. Є. Жуковського  
«Харківський авіаційний інститут»

**М. Л. Гелетка**

**ЦИФРОВІ ТЕХНОЛОГІЇ В ПЕРЕКЛАДАЦЬКІЙ ДІЯЛЬНОСТІ**  
Навчальний посібник

Харків «ХАІ» 2023

УДК 81'33+81'322.4  
Н 50

Рецензенти: канд. пед. наук, доц. О. В. Карасьова,  
канд. філол. наук, доц. І. І. Морозова

**Гелетка, М. Л.**

Н50 Цифрові технології в перекладацькій діяльності [Електронний ресурс] : навч. посіб. / М. Л. Гелетка. – Харків : Нац. аерокосм. ун-т ім. М. Є. Жуковського «Харків. авіац. ін-т», 2023. – 67 с.

Описано принципи й алгоритми комп'ютеризованого перекладу з іноземної мови на українську і з української мови на іноземну. Розглянуто можливості глобальної мережі «Інтернет» і особливості спеціального програмного забезпечення для зручного і швидкого виконання перекладу письмових текстів.

Для студентів, аспірантів, викладачів, перекладачів, які опановують технології письмового перекладу з використанням інформаційних технологій.

Іл. 15. Табл. 3. Бібліогр.: 6 назв

**УДК 81'33+81'322.4**

© Гелетка М. Л., 2023  
© Національний аерокосмічний  
університет ім. М. Є. Жуковського  
«Харківський авіаційний інститут», 2023

## ЗМІСТ

Передмова.....	5
1. Інформаційні технології для технологічної підтримки роботи перекладача.....	6
2. Робота з електронними словниками.....	11
2.1. Словник «Мультилекс».....	12
2.2. Словник ABBYY Lingvo.....	13
2.3. Створення LUD-словника.....	15
2.4. Створення DSL-словника.....	16
2.5. Команди мови DSL.....	21
2.6. Словник Multitran.....	23
2.7. Словник Reverso Context.....	25
3. Системи автоматичного (машинного) перекладу (computer-aided/machine translation).....	28
3.1. Типологія систем машинного перекладу.....	28
3.2. Популярні онлайн-системи машинного перекладу.....	30
4. Системи автоматизованого перекладу (computer-assisted translation).....	32
5. SDL Trados Studio 2022.....	36
5.1. Особливості інтерфейсу SDL Trados Studio 2022.....	37
5.2. Ресурси, якими забезпечує перекладача система Trados.....	38
5.3. Здійснення перекладу за допомогою редактора SDL Trados Studio 2022.....	39
5.4. Основні функції редактора Trados.....	41
5.4.1. Застосування форматування.....	42

5.4.2. Переклад нетекстових файлів.....	43
5.4.3. Перелік швидких клавіш для роботи в SDL Trados 2022	44
5.5. Використання функції Winalign.....	44
5.6. Створення термінологічної бази Multiterm зі словника термінів за допомогою Excel і MultiTerm Convert .....	45
5.6.1. Створення аркуша Excel із термінологією.....	45
5.6.2. Перетворення аркуша Excel за допомогою MultiTerm Convert.....	46
5.6.3. Імпорт перетвореного аркуша Excel у MultiTerm.....	47
5.7. Визначення понять «проект» і «підпроект».....	49
5.7.1. Керування проектами.....	50
6. Аспекти локалізації прикладного програмного забезпечення.....	50
6.1. SDL Passolo Essential 2022 як засіб для локалізації.....	55
6.2. Інші засоби локалізації.....	57
7. Огляд найпоширеніших систем автоматизованого перекладу.....	57
8. Корпусна лінгвістика.....	60
Бібліографічний список.....	66

## ПЕРЕДМОВА

Курс сучасних цифрових технологій у перекладацькій діяльності є допоміжною дисципліною професійного циклу підготовки лінгвістів-перекладачів. Мета курсу – підготувати фахівців, які мають знання, вміння і навички в галузі комп'ютеризованого перекладу з англійської мови на українську і з української мови на англійську в обсязі, необхідному для таких дій:

- виконувати за допомогою відповідного програмного забезпечення (систем Translation Memory – SDL Trados Studio 2022, Wordfast, Startransit тощо та електронних словників і термінологічних баз) письмовий переклад текстів з англійської мови на українську і з української мови на англійську;
- правильно оформлювати результати роботи в електронному форматі;
- використовувати глобальну мережу «Інтернет» для підвищення якості перекладу.

Курс сучасних цифрових технологій у перекладацькій діяльності вивчають протягом першого семестру п'ятого курсу на основі вже здобутих студентами знань про комп'ютерні технології, а також набутих умінь і навичок у сфері письмового перекладу.

Мета посібника полягає в узагальненні практичного досвіду з використання електронних засобів, накопиченого автором та іншими перекладачами-практиками у процесі письмового перекладу технологічної документації. Посібник містить огляд сучасних електронних ресурсів, представлених сьогодні на ринку, які перекладач із максимальною користю для себе може застосовувати під час роботи. Посібник складається з восьми розділів. Кожний розділ містить основний текст, список адрес електронних ресурсів конкретної групи, за якою їх можна знайти в мережі «Internet», короткий опис кожного ресурсу та домашні завдання.

## 1. ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ДЛЯ ТЕХНОЛОГІЧНОЇ ПІДТРИМКИ РОБОТИ ПЕРЕКЛАДАЧА

**Інформаційні технології (ІТ)** являють собою комплекс методів і прийомів збирання, передавання, оброблення, збереження та донесення до користувача інформації, реалізованої сучасними засобами, а саме – комп'ютерами.

До основних завдань застосування інформаційних технологій у **лінгвістиці** належать:

- створення систем штучного інтелекту;
- створення систем автоматичного (машинного) та автоматизованого перекладу;
- створення систем автоматичного анотування та реферування текстів;
- створення систем генерування текстів;
- створення систем навчання мови;
- створення систем розуміння мовлення;
- створення систем генерації мовлення;
- створення автоматизованих інформаційно-пошукових систем;
- створення систем атрибуції та дешифрування анонімних і псевдоанонімних текстів;
- розроблення різних баз даних (словників, карток, реєстрів тощо) для гуманітарних наук;
- розроблення різного типу автоматичних словників;
- розроблення систем передавання інформації в мережі «Інтернет» тощо.

**Мета** посібника – ознайомити майбутніх фахівців з електронними словниками, системами автоматичного (машинного) та автоматизованого перекладу.

**Цільова аудиторія:** перекладачі, менеджери проектів бюро перекладів, викладачі та студенти старших курсів факультетів філології та перекладу, усі фахівці, чия професійна діяльність пов'язана з письмовим перекладом.

**Практичне значення:** вивчення можливостей цифрових технологій та набуття навичок їх використання, що дасть змогу розширити функціонал перекладача на міжнародному ринку перекладів, раціонально організувати роботу перекладача або групи перекладачів, підвищити ефективність роботи перекладача, підвищити якість перекладу.

### **Вимоги до рівня підготовки слухачів:**

- знати правила техніки безпеки під час роботи з ПК;
- мати базові навички роботи в середовищі Windows;
- мати базові знання про мережу «Інтернет» й електронну пошту та базові навички роботи з ними;
- уміти працювати у програмі Word на рівні впевненого користувача;
- уміти працювати у програмі Excel на базовому рівні;
- уміти працювати у програмі Power Point на базовому рівні.

У минулому основними інструментами письмового перекладача були друкарська машинка та колекція паперових словників. Тепер серед найнеобхідніших інструментів перекладача є персональний комп'ютер.

Перекладачам слід дотримуватися простого правила: на ПК мати мінімум два логічних сектори на жорсткому диску – для роботи і для особистого використання. Це допоможе мати оперативний доступ до робочих матеріалів і захистить від втрати важливої інформації в разі збою в роботі ПК. Важливо не зберігати робочу інформацію на жорсткому диску, де встановлена операційна система, оскільки можна втратити всю інформацію під час переустановлення системи.

Основні вимоги до комп'ютера: достатньо висока продуктивність, достатньо велика ємність оперативної пам'яті та ємність жорсткого диску, якісний монітор, бажано мати сканер і принтер, Wi-Fi-роутер, кардридер тощо [1, с. 12-13].

### **Мінімальна конфігурація комп'ютера, необхідна для роботи перекладача з використанням ІТ:**

- GigaByte GA-P35-S3 Socket775, PCI-E, LAN, SATA RAID – позначення материнської плати виробництва компанії GigaByte, що підтримує технології PCI-Express і SATA RAID і має вбудований мережевий адаптер (мережеву карту);
- Intel Core 2 Duo (E6550) 2330MHz 4MB 1333MHz S775 – позначення процесора, у цьому випадку двоядерного, виробництва компанії Intel з тактовою частотою 2,33 ГГц;
- DDR2 DIMM 1Gb PC5300 667MHz Samsung II – це модуль оперативної пам'яті на 1 Гбайт виробництва компанії Samsung, створений за технологією DDR2;
- GF-8600GT, 256Mb DDR3, 128bit, PCI-E, 2xDVI, HDTV – позначення відеокарти та підтримуваних нею технологій, 256 Мбайт – ємність вбудованої оперативної пам'яті відеокарти;



- WD 250Gb 7200rpm 16Mb – жорсткий диск (вінчестер) виробництва компанії Western Digital із ємністю 250 Гбайт і швидкістю 7200 оборотів на хвилину;
- NEC AD-7170S-01 18xDVD + R, 18xDVD-R, 8xDVD + RW, 6xDVD-RW, 8xDVD + R9, 12xDVD-RAM, 16x/48x/32x/48x, DVD +-RW, ivory SATA – пристрій для роботи із CD і DVD-дисками (з переліком підтримуваних форматів і швидкостей роботи).

Зазвичай спеціалісти з комп'ютерних технологій допомагають дібрати моделі з необхідною конфігурацією моделі, що підтримують відповідні стандарти.

Набуття навичок оброблення документів забезпечується одночасним використанням перекладачами таких програмних засобів:

- **браузерів** – програмного забезпечення для комп'ютера або іншого електронного пристрою, зазвичай під'єданого до інтернету, що дає можливість користувачеві взаємодіяти з текстом, зображеннями або іншою інформацією на гіпертекстовій вебсторінці (Internet Explorer, Mozilla Firefox, Google Chrome, Safari, Opera);
- **інформаційно-пошукових систем**, призначені для пошуку текстів (документів, їх частин, фактографічних записів) у сховищах (базах даних) за формальними характеристиками (Google, Bing, Rambler, Yahoo);
- **прикладних програм** – текстових редакторів і процесорів, електронних таблиць та інших засобів Microsoft Office (Word, Excel, Power Point, Outlook, OneNote, Access тощо);
- **видавничих або верстальних програм** – Adobe InDesign, Adobe FrameMaker, QuarkXPress, PageMaker тощо – досить знати про них;
- **систем автоматизованого проєктування** – Computer Aided Design: AutoCAD, Solidworks – досить знати про них, Adobe Reader, Fine Reader;

**\*Adobe Reader** – це інструмент для відкриття та використання файлів Adobe PDF, створених у програмі Adobe Acrobat. Хоча програма Reader не дає змоги створювати PDF-файли, за її допомогою можна переглядати та друкувати їх, а також керувати ними. Після відкриття PDF-файлу у програмі Reader відображаються різноманітні інструменти для швидкого пошуку інформації. За замовчуванням програма Adobe Reader X працює в захищеному режимі, забезпечуючи додатковий рівень захисту. У захищеному режимі зловмисні PDF-документи не можуть запускати випадкові виконувані файли або здійснювати запис у системні каталоги чи реєстр Windows.

**\*Fine Reader** – це програма для розпізнавання тексту дає змогу швидко і точно, без передрукування, змінювати зображення

документів і PDF-файли на електронні формати, придатні для редагування (Microsoft Word, Microsoft Excel, Microsoft Powerpoint, Rich Text Format, HTML, PDF/A, searchable PDF, CSV і текстові (plain text) файли). В останній, одинадцятій, версії ABBYY FineReader підвищилась точність розпізнавання, а оброблення документів відбувається до 99 відсотків швидше. Зрозумілий інтерфейс програми дає змогу одним натисканням миші розпізнати документи трьома мовами.

- **електронних словників** – комп'ютерних баз даних, що містять закодовані особливим чином словникові статті, які дають змогу здійснювати швидкий пошук потрібних слів, часто з урахуванням морфологічних форм і з можливістю пошуку словосполучень (прикладів вживання), а також із можливістю зміни напряму перекладу (ABBYY Lingvo, Multitran тощо);

- **систем машинного перекладу**, що здійснюють переклад текстів, ґрунтуючись на формальному «знанні» мови (синтаксисі мови – правил побудови пропозицій, правил словотвору) і використанні словників. Програма-перекладач спочатку аналізує текст однією мовою, а потім генерує цей текст іншою мовою (Prompt, «Сократ», Play, Pragma);

**\*Онлайн-перекладачі** – онлайн програми, що виконують переклад безпосередньо у вікні веббраузера, не вимагаючи встановлення програми-перекладача на комп'ютер користувача. При цьому зазвичай є обмеження на обсяг тексту, який можна перекласти за один раз.

- **систем автоматизованого перекладу (Computer Aided Translation)** – програми, призначені для спрощення процесу перекладу однотипних (або схожих) документів. Зрозуміло, що під час автоматизованого перекладу перекладач також бере участь, виконуючи ті самі функції. Завдання машини – запам'ятовувати перекладені фрагменти тексту та використавувати їх у наступних роботах. Така технологія називається translation memory (TM), про неї буде розказано трохи пізніше.

Слід зазначити, що знання та використання таких програм (Trados, Transit, Across Personal Edition, Wordfast тощо) стає обов'язковою вимогою для участі у великих розподілених проєктах.

Крім того, багато бюро та агентств перекладів, так само як і безпосередні замовники, все частіше стали ставити подібні вимоги до перекладачів під час роботи навіть над невеликими документами й текстами. Це пов'язано з тим, що компанії, тісно пов'язані з перекладацькою галуззю, прагнуть створити й постійно наповнювати власну базу TM для її використання в подальшому.

Використання CAT-програм часом дає змогу істотно скоротити час самого перекладу. Наприклад, якщо перекладач часто працює над технічними текстами (різними інструкціями тощо), однотипними або схожими за структурою документами (наприклад, пакетом митних декларацій), зазначені вище програми значно допоможуть у процесі перекладу. Однак якщо перекладач працює з художніми текстами, користь від подібних програм буде вельми сумнівною;

• **систем штучного інтелекту для лінгвістики і перекладознавства:**

– **NLTK** – це популярна бібліотека Python для таких завдань оброблення природної мови, як токенізація, створення коренів, додавання тегів, синтаксичний аналіз і аналіз настроїв;

– **Stanford CoreNLP** – це набір інструментів оброблення природної мови, розроблений Стенфордським університетом, який має такі можливості, як розпізнавання іменованих об'єктів, аналіз настроїв і вирішення кореференції;

– **Apache OpenNLP** – це набір інструментів оброблення природної мови з відкритим кодом, який надає такі можливості, як виявлення речень, токенізація, тегування частин мови та розпізнавання іменованих об'єктів;

– **spaCy** – це бібліотека Python для розширених таких завдань оброблення природної мови, як розбір залежностей, розпізнавання іменованих об'єктів і класифікація тексту;

– **GATE** – це платформа на основі Java для оброблення природної мови, яка надає ряд інструментів для таких завдань, як вилучення інформації, аналіз тексту та аналіз настроїв;

– **LDC** – це некомерційна організація, яка надає доступ до ресурсів лінгвістичних даних для науково-дослідних цілей;

– **WordNet** – це лексична база даних, яка організовує слова в такі семантичні зв'язки, як синоніми, антоніми та гіперніми;

– **ConceptNet** – це семантична мережа, яка представляє концепти та їхні зв'язки з іншими концептами у спосіб, який можна використовувати для таких завдань розуміння природної мови, як усунення неоднозначності слів і семантична подібність;

– **Google Cloud Natural Language API** – це хмарний сервіс, який надає такі можливості, як аналіз настроїв, розпізнавання об'єктів і класифікація вмісту;

– **Amazon Comprehend** – це хмарний сервіс оброблення природної мови, який надає такі можливості, як розпізнавання об'єктів, аналіз настроїв і моделювання тем.

**Домашнє завдання.** Find 10 similar texts of about 100 words. They have to be identical. For example, sales contracts (not the whole paper, but

one issue like FORCE MAJEURE, PAYMENT etc.), manuals, user guides (for NOKIA MOBILE PHONES, PHILIPS FOOD PROCESSORS, ASUS LAPTOPS), pharmaceutical product descriptions (painkillers PARACETAMOL, IBUPROFEN). Highlight and learn their terminology. Then translate them into Ukrainian using either on-line translator, finding their Ukrainian translations in the Internet or doing it yourself. Thus you will have 10 .docx files in English and 10 equivalent .docx files in Ukrainian.

## 2. РОБОТА З ЕЛЕКТРОННИМИ СЛОВНИКАМИ

**Комп'ютерна лексикографія** – напрямок лексикографії, який займається розробленням електронних словників.

Для створення словників використовуються спеціальні програми – бази даних, корпуси текстів тощо. Безліч лексикографічних програм можна поділити на таке:

- програми підтримки лексикографічних праць;
- автоматичні словники різних типів.

Основні завдання у процесі створення електронного словника:

- визначення форми заголовної одиниці;
- визначення обсягу інформації, яка міститиметься в описі лексичної одиниці;
- вибір способу організації словникової статті.

Структура словникової статті в різних електронних словниках може мати різний вигляд, обсяг статті може містити 1–99 зон, у яких фіксується різнотипна лінгвістична й екстралінгвістична інформація:

- заголовне слово;
- один або кілька еквівалентів (у перекладному словнику);
- інформація про тематичну приналежність слова;
- граматична інформація;
- дефініція;
- контексти;
- інформація про лексичну сполучуваність;
- семантичні ієрархічні зв'язки заголовного слова;
- інформація про стилістичні характеристики слова;
- джерело та ін. [5].

## 2.1. Словник «МультиЛекс»

Програми «МультиЛекс» давно відомі вітчизняним користувачам (рис. 2.1). За роки свого існування програми набули популярності передусім завдяки тому, що були створені на основі реальних паперових словників. Останньою версією «МультиЛекс» стала сьома. Зараз словник не дуже популярний, оскільки його місце посіли сучасніші програми, які буде розглянато далі.

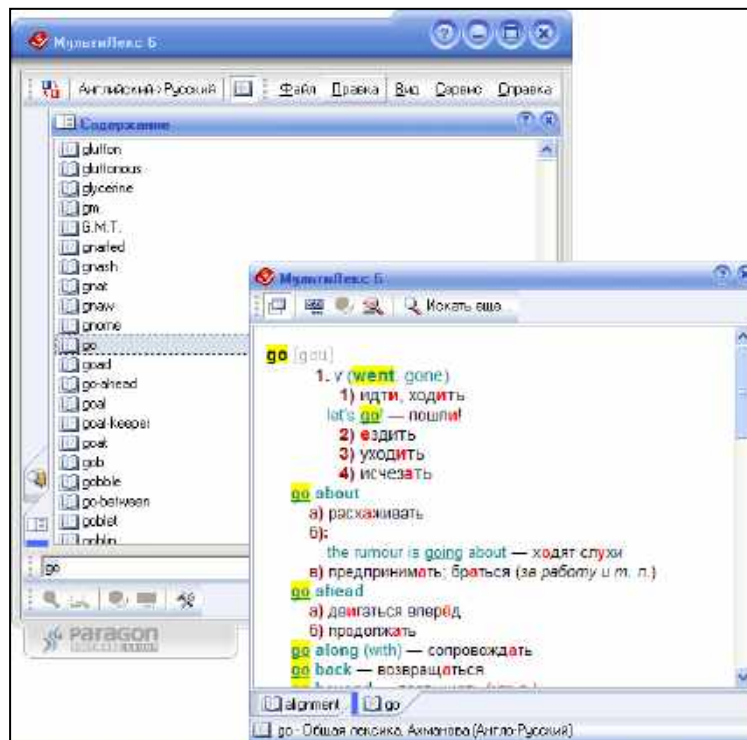


Рис. 2.1. Вікно словника «МультиЛекс»

Продукцію компанії «МультиЛекс» подано у трьох комплектаціях – стандартній, професійній та «боксовій». Як і раніше, «МультиЛекс» називається тільки «боксова» комплектація, а дві інші комплектації отримали нову назву «АльфаЛекс». Насправді, це та сама програма, але відрізняється лише обсягом словникових баз.

У стандартну jewel-версію входять словники В. К. Мюллера (170000 слів) і О. І. Смирницького (160000 слів). У професійній версії до них додаються ще 22 словники (два мільйони перекладів загалом). «Боксова» комплектація «МультиЛекс» містить 29 словників із загальним обсягом понад шість мільйонів перекладів. Так, у найпершу редакцію ввійшли такі словники, як «Новий великий англо-російський словник» за редакцією академіка Ю. Д. Апресяна і тлумачний словник Webster New World Dictionary & Thesaurus.

Для пошуку необхідно ввести слово, і програма одразу відшукає його в усіх активних у цей момент словниках. Причому це

повнотекстовий пошук, який здійснюється не тільки за індексом, але і за всіма перекладами і прикладами застосування. Можна здійснювати і перехресний пошук, натиснувши мишею два рази на будь-яке слово у словниковій картці, і все, що з ним пов'язано, також буде знайдено.

Слід ознайомитися з режимом «швидкого перекладу». Наводимо курсор до незнайомого слова, натискаємо «гарячу» клавішу, після чого перед нами з'являється картка зі словника з перекладом цього слова. Цікаво, що цей режим працює не тільки в текстовому процесорі або веббраузері.

Можна перекладати слова з графічного інтерфейсу будь-яких додатків для Windows.

Текст може бути озвучений, але для цього задіяний не людський голос, а Microsoft Agent.

Словник містить адаптивну самонавчальну технологію «МультиЛінкер». У всіх документах за її допомогою програма знаходить і виділяє слова, які вже були перекладені за допомогою словників «Мультилекс». Можна отримати переклад будь-якого із цих слів, прослухати їх звучання тощо. І все це можна зробити не закриваючи документ. Працює ця технологія тільки з Microsoft Office 2010.

Трохи ускладнює установлення програми обов'язкова процедура реєстрації, без якої більшість функцій програми буде заблоковано. Реєстрацію можна здійснити як через інтернет, так і через телефон.

Як доповнення до «боксової» версії словника додається «МультиЛекс Мобайл» – комплект словників для планшетів і смартфонів, що працюють на основі всіх популярних платформ – Pocket PC, Palm OS, Symbian і Windows Mobile Smartphone Edition.

Отже, попри занадто високу вартість, словник був однією з кращих пропозицій на ринку електронних словників, але сьогодні пошукові технології стрімко розвиваються, тому «МультиЛекс» є непоганим словником за обсягом опрацьованої лексики, проте не найнеобхіднішим інструментом перекладача.

## **2.2. Словник ABBY Lingvo**

ABBYY Lingvo поєднує простоту інтерфейсу і величезний обсяг наданої інформації. До складу англо-російської версії Lingvo 16 увійшло 69 словників, загальний обсяг яких – 3,2 млн словникових статей або 8 млн перекладів. Оскільки користувачеві під час роботи потрібно одразу кілька спеціалізованих словників, в одній із попередніх версій програми розробники передбачили групування словників за тематиками, у так звані «книжкові полиці». Типово вже пропонується

кілька таких тематичних груп, але користувач може створювати їх і самостійно.

Словникові статті в ABYBY Lingvo містять, мабуть, найбільшу кількість прикладів вживання (рис. 2.2). Але натисканням однієї кнопки можна приховати всі приклади, залишивши лише переклади. Користувач також може подивитися всі граматичні парадигми за вибраним словом.

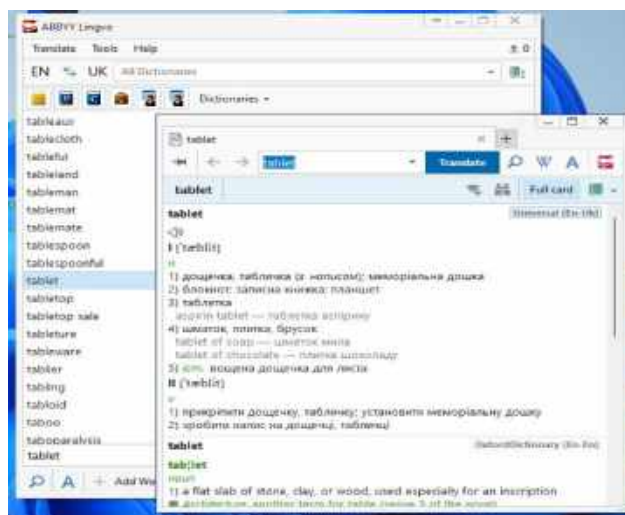


Рис. 2.2. Вікно словника ABYBY Lingvo x6

Ще у версії Lingvo 16 з'явилася нова функція – послівний переклад. Якщо виділити в тексті фразу та натиснути комбінацію «гарячих» клавіш, то у вікні програми відобразяться найуживаніші значення слів, що входять до фрази. Для кожного з перекладених слів можна буде відкрити й окрему словникову картку з переліком усіх значень, транскрипцією, прикладами вживання і вимови. Таким чином, можна отримати фактично підрядковий переклад тексту.

У попередній версії ABYBY Lingvo з'явився навчальний модуль Lingvo Tutor, який допомагає вчити іноземні слова. Навчання відбувається в ігровій формі: програма пропонує іншомовне слово, а користувач повинен вручну написати одне з його значень. У LingvoTutor x6 до послуг користувача вісім тем для навчання: «Погода», «Робота», «У готелі», «У ресторані», «Туризм», «Одяг», «Покупки» та «Екскурсії».

Як і в «МультиЛекс», у Lingvo є пошуковий механізм, зокрема перехресний. Особливістю словника є те, що в ньому озвучено не так багато слів, проте озвучені слова живими дикторами.

Ще одна особливість словника – DSL-компілятор, що входить до складу Lingvo, за допомогою цього компілятора користувачі можуть створювати власні словники, які потім можна під'єднувати до програми. Найкращі користувацькі словники є на сайті <https://www.lingvolive.com>, звідки їх можна безкоштовно завантажити в разі потреби.

Нарешті до складу дистрибутива ABBYY Lingvo x6 уведено версії словника для мобільних платформ – Palm OS, Windows Mobile і Symbian series 60. Таке оновлення дає змогу користуватися словником незалежно від місця перебування користувача.

ABBYY Lingvo x6 містить 220 загальнолексичних, тлумачних і тематичних словників для російської, англійської, німецької, французької, іспанської, італійської, португальської, китайської, турецької, української, латинської, угорської, грецької, данської, казахської, нідерландської, норвезької, польської, татарської, фінської мов.

Lingvo x6 підтримує локальні словники двох видів: прості LUD (слово-переклад) і просунуті DSL з усім оформленням системних словників ABBYY Lingvo, зображеннями, звуками та відео. LUD-словники можна створювати, доповнювати та редагувати під час роботи з програмою. Для створення DSL-словників використовується програма DSL Compiler x6 і команди мови DSL. Доповнювати і редагувати такий словник, одночасно працюючи в Lingvo не можливо, а от видалити його з програми, оновити та під'єднати можна знову.

### 2.3. Створення LUD-словника

У головному вікні Lingvo x6 треба перейти в меню «Сервіс», вибрати опцію «Створити/Редагувати картку» та заповнити всі необхідні поля так, як зображено на рисунку 2.3.

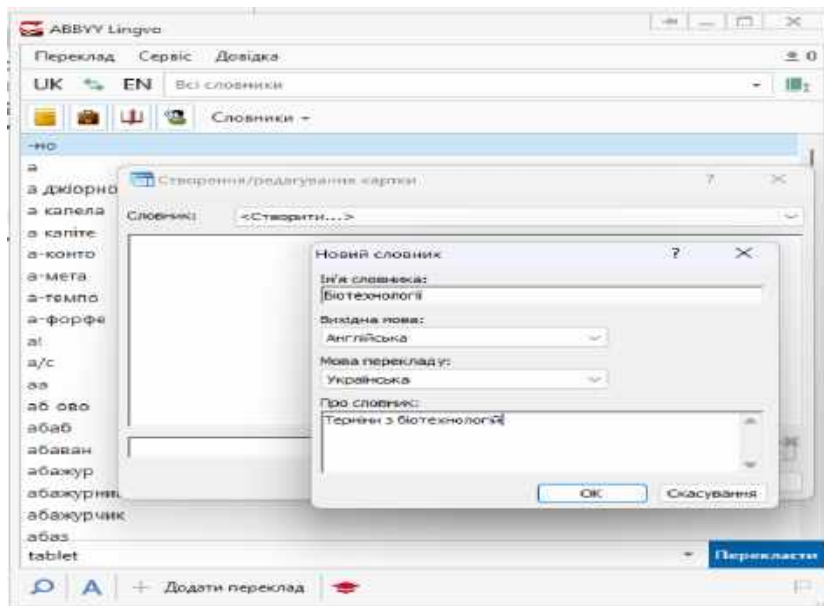


Рис. 2.3. Редагування картки словника ABBYY Lingvo x6



Тоді програма бачить створений словник і щойно додане до нього слово (рис. 2.4).

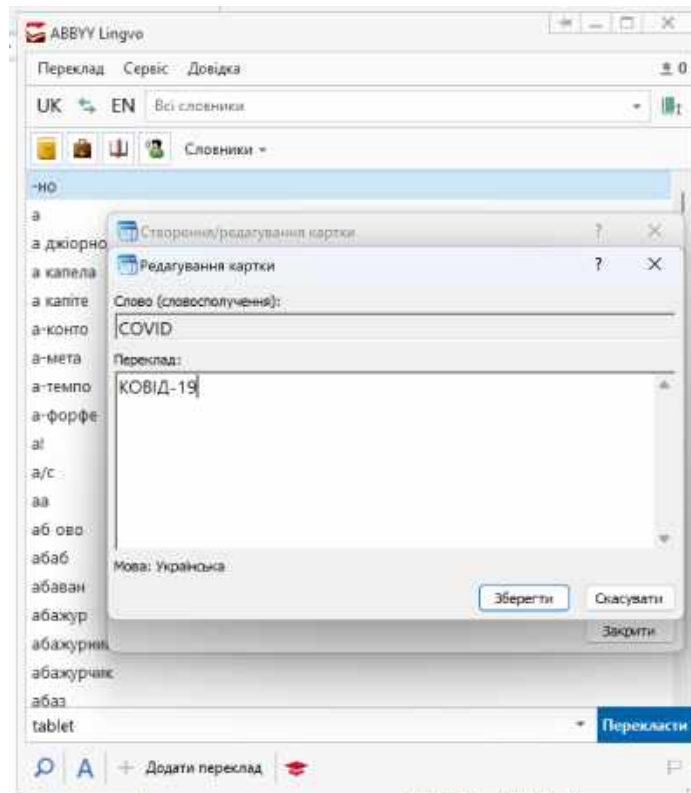


Рис. 2.4. Створення картки словника ABBYY Lingvo x6

## 2.4. Створення DSL-словника

Перший словник називатиметься «Мій перший словник» для перекладу з англійської мови на українську, а складатися словник буде з однієї картки з перекладом англійського словосполучення *to call it a day* на українську мову.

Порядок створення DSL-словника:

1. Файл словника DSL можна створити в будь-якому текстовому редакторі, наприклад у програмі «Блокнот».
2. Відкрити новий файл. На початку файлу розмістити заголовну частину:

```
# NAME "My first dictionary"  
# INDEX_LANGUAGE "English"  
# CONTENTS_LANGUAGE "Ukrainian"
```

3. Кожен рядок починається зі службової частини, за якою в лапках вказана інформація користувача, відокремлена від службової частини одним пропуском. У наведеному прикладі в заголовній частині вказано

ім'я словника – **Мій перший словник**, вихідна мова – **English**, і мова перекладу – **Ukrainian**.

4. Після створення заголовної частини можна починати створення картки словника. У кожній картці має бути заголовок і текст. Заголовок містить перекладне слово або словосполучення, що має розміщуватися на окремому рядку і починатися з першої позиції. Текст картки може містити переклад, коментарі, примітки й будь-яку іншу інформацію, яку автор словника визнає корисною. Текст картки має розміщуватися в рядку, наступному за заголовком, при цьому текст картки може містити будь-яку кількість рядків. Кожен рядок тексту картки має починатися з пропуску або символу табуляції (клавіша «ТАВ»). Цей простий спосіб дає змогу відокремити заголовок від тексту картки.

5. Можна скопіювати у файл словника текст картки, наведений нижче:

**call it a day**

**завершити робочий день**

6. Зберегти файл словника в кодуванні Unicode (у списку «Кодування» діалогового вікна «Зберегти як» вибрати пункт Unicode) з ім'ям **FirstDictionary**. Закрити «Блокнот».

7. Змінити розширення текстового файлу словника на **.dsl**.

8. Запустити програму **DSL Compiler** (меню Пуск> Програми> ABBYY Lingvo x5> DSL Compiler x5).

9. У діалоговому вікні DSL Compiler у групі «Властивості вихідного файлу» вибрати файл **FirstDictionary.dsl**, що містить створений словник, відзначити для нього кодування **Unicode**.

10. Переконатися в тому, що опція «Автоматично додавати розмітку», необхідна для пошуку за вмістом карток словника, відзначена. У цьому випадку програма DSL Compiler автоматично додасть необхідну розмітку у словник.

11. Натиснути кнопку «Компілювати». Якщо компіляцію словника здійснено успішно, з'явиться діалогове вікно «Про словник» з повідомленням про те, що словник створено. У результаті компіляції буде створено два файли: файл словника **FirstDictionary.dsl** і файл **FirstDictionary.dde**. У файлі First буде зазначено виявлені під час компіляції помилки.

12. Якщо у процесі компіляції з'явилося повідомлення про те, що не знайдено файл анотації, можна створити словник і без анотації. У повідомленні треба натиснути кнопку «Так» – процес компіляції буде продовжено.

14. Якщо не вдалося скопіювати словник, то, можливо, заголовна частина словника або його картка містять помилки.

Переглянути список помилок можна у файлі FirstDictionary.dde. Файл із розширенням .dde створюється під час компіляції в тому випадку, якщо відзначено опцію «Створювати файл помилок» (\*. dde) в діалоговому вікні DSL Compiler. Ім'я цього файлу збігається з ім'ям файлу словника. Слід виправити помилки і повторити компіляцію.

15. Під'єднати словник до програми АBBYU Lingvo x6 (меню Сервіс> Додати словник із файла...).

16. Перший користувацький словник створено. Словник складається з однієї картки, але є повноцінним словником АBBYU Lingvo. Після ознайомлення з мовою DSL і з роботою програми DSL Compiler, можна створити великі словники, які стануть у пригоді під час роботи та навчання.

17. У своєму словнику можна за допомогою команд DSL мови виділяти ділянки тексту кольором, жирним шрифтом, курсивом або підкресленням; виділяти в картці різні зони; вставляти в картку посилання на інші картки та на вебсайти; створювати значок для свого словника; вставляти зображення або звук у картку й багато іншого.

18. Можна набрати текст словника в будь-якому текстовому редакторі, зокрема Microsoft Word. При цьому необхідно дотримуватися вимог щодо оформлення заголовної частини словника, заголовка та тексту картки. Потім створений текст потрібно зберегти у форматі .txt з кодуванням Unicode. Наприклад, у Microsoft Word 2000 така послідовність: зі списку «Тип файлу» діалогового вікна «Зберегти як» потрібно вибрати пункт «Кодований текст». У Microsoft Word 2002 така послідовність: «Тип файлу» діалогового вікна «Зберегти як» потрібно вибрати пункт «Звичайний текст» (\*. txt), натиснути кнопку «Зберегти», у діалоговому вікні, у списку «Інше кодування» слід вибрати пункт Unicode. Якщо ж текстовий редактор не дає змоги вибрати кодування Unicode, просто треба зберегти текст у форматі .txt, а потім відкрити файл у програмі «Блокнот». У меню «Файл» вибрати команду «Зберегти як» і у списку «Кодування діалогового вікна» вибрати пункт Unicode.

21. Щоб додати нові картки або відредагувати скомпільований і під'єднаний до АBBYU Lingvo x6 словник, слід від'єднати словник від АBBYU Lingvo x6, відкрити файл словника .dsl. Внести зміни, скомпільувати словник і під'єднати його до АBBYU Lingvo x6.

22. Нові картки можна створити в окремому файлі, а потім під'єднати цей файл до основного файлу словника за допомогою директиви # INCLUDE.

23. Анотація містить довідкову інформацію про словник (відомості про автора, тему, дату створення тощо). Ця інформація відображається в АBBYU Lingvo x6 у разі подвійного натискання миші на значок

словника. Для створення анотації можна використовувати «Блокнот». Файл анотації має збігатися з ім'ям файлу словника. Треба зберегти файл у кодуванні Unicode, а потім змінити розширення файлу на **.ann**. Під час компіляції цей файл буде використано для створення анотації.

24. Можна додати у файл анотації опис словника трьома мовами – англійською, російською, українською. Щоб текст анотації було відображено мовою інтерфейсу, потрібно додати такі теги перед описом словника відповідною мовою:

```
# LANGUAGE "English" – для англійської мови;  
# LANGUAGE "Russian" – для російської мови;  
# LANGUAGE "Ukrainian" – для української мови.
```

Елемент DSL-словника має такий вигляд:

```
#NAME «My first dictionary»  
#INDEX_LANGUAGE «English»  
#CONTENTS_LANGUAGE «Ukrainian»  
blogorrhea  
[s]Blogorrhea.wav[/s]  
[b]1.[/b] [p]сущ.[/p]  
[m1]1) [p]блог.[/p] блого[']рея['] [com]([i]мережева графоманія, іноді ЖЖ-графоманія, рідко – блогографоманія[/i])/[com]/[m]  
[m2][*][ex][lang name=«English»]Blogorrhea is writing when you have nothing to say.[/lang] – Блогорея – написання об'ємних постів ні про що. (літературний переклад)/[ex]/[*]/[m]  
[m2][com]Джонатан Янг, «Короткий посібник для блогерів» [i](Jonathan Yang, author of The Rough Guide to Blogging)/[i]/[com]/[m]  
[m3][s]guide.jpg[/s]/[m]  
[m1][b]English-English[/b]/[m]  
[m1]1) pathologically incoherent, repetitious blogging[/m]  
[m1]2) incessant or compulsive holding forth on-line[/m]  
[m1]3) wearisome volubility aimed at an unsuspecting virtual readership[/m]  
[m1][*][url] meadhunter.blogspot.com/[url]/[*]/[m]
```

**Додаткова інформація.** Словник мовою DSL являє собою послідовність карток. Кожна картка складається із заголовка (заголовного слова або словосполучення) і тексту картки. У процесі використанні словника в системі ABBYY Lingvo заголовки показано у «Списку заголовних слів», а текст картки (разом із заголовком) подано у вікні картки як переклад або тлумачення.

На початку текстового файлу зазначається заголовок, що містить:

**ім'я словника (# NAME "ім'я словника"),**  
**вихідну мову (# INDEX\_LANGUAGE "ім'я мови"),**  
**мову перекладу (# CONTENTS\_LANGUAGE "ім'я мови").**

Якщо словник складається з декількох файлів, то для їх під'єднання використовуйте директиву **# INCLUDE**. Через пробіл або табуляцію в лапках слід зазначити повний шлях до файлу. Оскільки символ «\» використовується у DSL для вказівки, що наступний за ним символ є текст, а не ім'я тега, у шляху до файлу слід вживати подвійні символи «\\», наприклад **#INCLUDE "C:\\Dictionaries\\UniverseE.dsl"**

Крім того, можна вказувати шляхи щодо каталогу, у якому міститься основний файл DSL (тобто той, у якому є директива INCLUDE). У цьому випадку потрібно або вказати ім'я файлу без шляху, або поставити на початку шляху крапку «.», наприклад

**# INCLUDE "UniverseE.dsl"**  
**# INCLUDE. "\\ includes \\ UniverseE.dsl"**

**Примітка.** Якщо текст у словнику було набрано в ANSI-кодуванні, то необхідно вказати кодову сторінку вихідного тексту. Для цього, після визначення вихідної мови та мови перекладу, додайте визначення кодової сторінки: **# SOURCE\_CODE\_PAGE "ім'я кодової сторінки"** (табл. 2.1).

Таблица 2.1

Можливі імена кодових сторінок

Ім'я кодової сторінки	Номер кодової сторінки у Windows
Latin	1252
Cyrillic	1251
EasternEuropean	1250

Під час збереження тексту словника в ANSI-кодуванні важливо врахувати, що вихідна мова і мова перекладу мають належати до однієї кодової сторінки. Якщо це не так, текст має бути збережено в кодуванні Unicode.

Якщо текст словника зберігається в кодуванні Unicode, то фактичне визначення кодової сторінки не потрібно.

## 2.5. Команди мови DSL

Усі команди мови DSL мають вигляд «[a]», де a – конкретна команда (один або кілька символів). Відповідна їй команда вимикання – «[/ a]»:

[b], [/ b] – напівжирний текст;

[i], [/ i] – курсивний текст;

[u], [/ u] – підкреслений текст;

[c], [/ c] – кольоровий текст;

[\*], [/ \*] – виділення тексту як вторинного відображення;

[mN] – встановлення зсуву абзаца щодо лівого краю картки, де N – довільна цифра від 0 до 9, абзац після цієї команди до експліцитної команди скасування зсуву («[/ m]») у поточній картці буде зміщено від лівого краю на число пробілів, кратне числу N у команді;

[trn], [/ trn] – виділення поля перекладу;

[ex], [/ ex] – виділення поля прикладу;

[com], [/ com] – виділення поля коментарів;

[! trs], [! trs] – виключення з індексування всередині поля;

[s], [/ s] – виділення поля мультимедіа (використовується для вставлення зображень або звукових файлів);

[url], [/ url] – посилання на вебсторінку;

[p], [/ p] – виділення приміток (у разі натискання на примітку з'являється роз'яснювальний текст);

[lang] [/ lang] – мова слова або словосполучення; вказується для слів усередині картки, мова яких відрізняється від мови перекладу (у цьому випадку для слів можливий пошук, переклад із картки, пошук тощо); параметри: ім'я мови або ідентифікатор мови (наприклад, [lang id = 1]); ім'я мови вказується в лапках, наприклад [lang name = "Russian"]; список DSL-мов наведено в розділі «Підтримувані мови»;

[ref] [/ ref] – посилання на картку в тому самому словнику, до якого належить картка, що містить це посилання (еквівалентна висновком заголовка картки, на яку створюється посилання, у подвійні кутові дужки << >>);

[sub] [/ sub] – нижній індекс;

[sup] [/ sup] – верхній індекс.

Під час відображення тексту картки система ABBYY Lingvo форматує кожен абзац відповідно до ширини вікна картки. При цьому межі абзаца визначаються за такими правилами:

- кожен рядок – один абзац;

- символ подкартки «@» – початок абзаца.

Це означає, що кожний перший рядок абзаца починається з першої позиції вікна картки, а весь абзац розділяється на рядки по межі слова залежно від ширини вікна.

Рекомендується встановлювати відступ абзаца щодо лівого краю картки. Для цього використовується команда «[mN]», де N – довільна цифра 1–9. Абзац після цієї команди до експліцитної команди скасування відступу («[/ m]») у поточній картці буде зрушено від лівого краю на число пробілів, кратне числу N у команді.

Щоб текст перекладу було зміщено на один відступ щодо заголовка картки, слід форматовувати картку таким чином:

**little**  
**[m1] 1) маленький [/ m]**  
**[m1] [\*] @ little one**  
**малюк**  
**@**  
**[/ \*] [/ m]**  
**[m1] 2) небагато, мало [/ m]**  
**[m1] 3) дрібниця, дрібниця [/ m]**  
**[m1] [\*] @ little by little**  
**помалу, поступово**  
**@**  
**[/ \*] [/ m]**

Щоб розмістити в тексті картки посилання на сторінку в інтернеті, у dsl-тексті картки потрібно вказати адресу сторінки всередині тегів [url], [/ url], наприклад [url] <http://www.lingvo.ru> [/ url].

Щоб вставити зображення в текст картки, у dsl-тексті картки треба зазначити ім'я файлу із зображенням усередині поля мультимедіа [s], наприклад [s] mypicture.bmp [/ s].

Дозволено такі формати зображень: BMP, Bitmaps (\*. bmp), PCX (\*. pcx), DCX (\*. dcx), JPEG (\*. jpg), TIFF (\*. tif)

Щоб вставити звук у картку, у dsl-тексті картки треба вказати ім'я звукового файлу всередині поля мультимедіа [s], наприклад [s] mysound.wav [/ s].

**Зауваження.** Прослуховування звукових файлів можливе лише за наявності звукової плати. Дозволено формат звукових файлів Wave Sound (\*. wav). Щоб вставити в картку відео, у dsl-тексті картки треба зазначити ім'я файлу із зображенням усередині поля мультимедіа [s], наприклад [s] movie.avi [/ s]. Цей файл має міститися в папці словника або в папці ABBYY Lingvo [4].

## 2.6. Словник Multitran

Multitran – інтернет-система, до якої входять такі електронні словники: англо-український і українсько-англійський; німецько-український і українсько-німецький; іспансько-український і українсько-іспанський; французько-український і українсько-французький; нідерландсько-український і українсько-нідерландський; італійсько-український і українсько-італійський; латисько-український і українсько-латиський; естонсько-український і українсько-естонський; японсько-український і українсько-японський тощо (рис. 2.5).

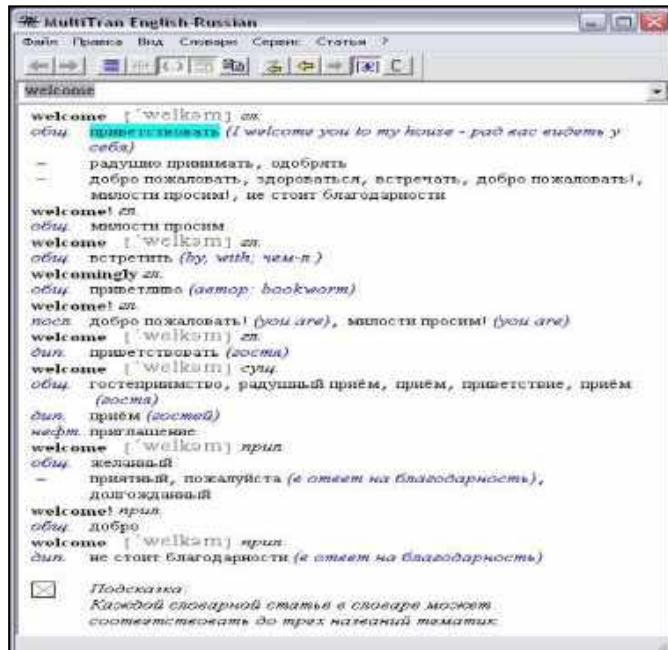


Рис. 2.5. Вікно словника Multitran

Щодня на сайт зазначеного електронного словника <https://www.multitran.com/m.exe?l1=1&l2=33&s=aircraft&l2=33> заходять понад 90 тис. користувачів, а сервер обробляє понад 1,5 млн пошукових запитів.

Активно працюють форуми Multitran, де будь-який користувач може поставити запитання та поспілкуватися з колегами-перекладачами. Періодично проводять зустрічі учасників форуму (row-wow) в офлайн-форматі.

Словникова база Multitran створена шляхом сканування, розпізнавання і перероблення великої кількості паперових словників та об'єднання отриманих перекладів слів у єдину базу даних.

У Multitran передбачено можливість поліпшення якості словникової бази, хоча щоб вивірити всю базу в кілька мільйонів слів потрібні десятки років. Кожен зареєстрований користувач словника може залишати повідомлення про помилки, знайдені у словнику. Ці



повідомлення записуються у відповідні словникові статті і відображаються в усіх користувачів, поки помилку не буде виправлено. З моменту впровадження цієї функції надійшло приблизно 8000 повідомлень про помилки, багато було потім виправлено. Таким чином, користувачі не тільки додають на сайт свої терміни (уже додано більше 180 000), а й допомагають виправляти помилки. Нижче наведено вигляд онлайн-версії словника Multitran (рис. 2.6).



Рис. 2.6. Вікно онлайн-словника Multitran

### Характеристики словника Multitran:

- понад 5 млн термінів, виражених різними частинами мови;
- терміни кожної мовної пари словника доступні у двох напрямках перекладу;
- понад 800 предметних областей (тематик);
- можливість активного самостійного поповнення словника користувачами;
- понад 1000 перекладачів поповнюють онлайн-версію Multitran безпосередньо на сайті;
- алфавітний, морфологічний, фразовий пошук (за словосполученнями);
- можливість підрядкового перекладу тексту;

- автоматичний пошук стійких словосполучень у тексті запиту;
- одночасний пошук у словниках, на форумі, а також у базі англійських та українських речень і базі посилань;
- у робочі дні (приблизно з 10:00 до 18:00) спостерігаються затримки і випадкові (мимовільні) відмови в обслуговуванні запитів.

### Види доступу до словника Multitran:

- через інтернет в онлайн-режимі – сайт Multitran;
- в автономному режимі на персональних комп'ютерах з Microsoft Windows – локальна або мережева установка;
- в автономному режимі на кишенькових комп'ютерах (Pocket PC), а також на смартфонах з операційною системою Symbian і Android.

## 2.7. Словник Reverso Context

Онлайн-словник Reverso Context (рис. 2.7) містить велику базу даних реальних прикладів та потужну лінгвістичну пошукову систему. У словнику подано вимову слів, розмовник, картки та багато іншого. Необхідно ввести лексичну одиницю або сказати, що треба перекласти, і з'являться результати запиту: найбільш значущі переклади, за якими наслідують приклади використання, вилучені із джерел, перекладених професіоналами (сайтів, офіційних документів, діалогів фільмів, газетних статей, описів продукції). Після цього можна зберігати результати пошуку та надані приклади в розмовнику, вивчати нові слова і вирази за допомогою карток. Якщо зареєструватися, то останні

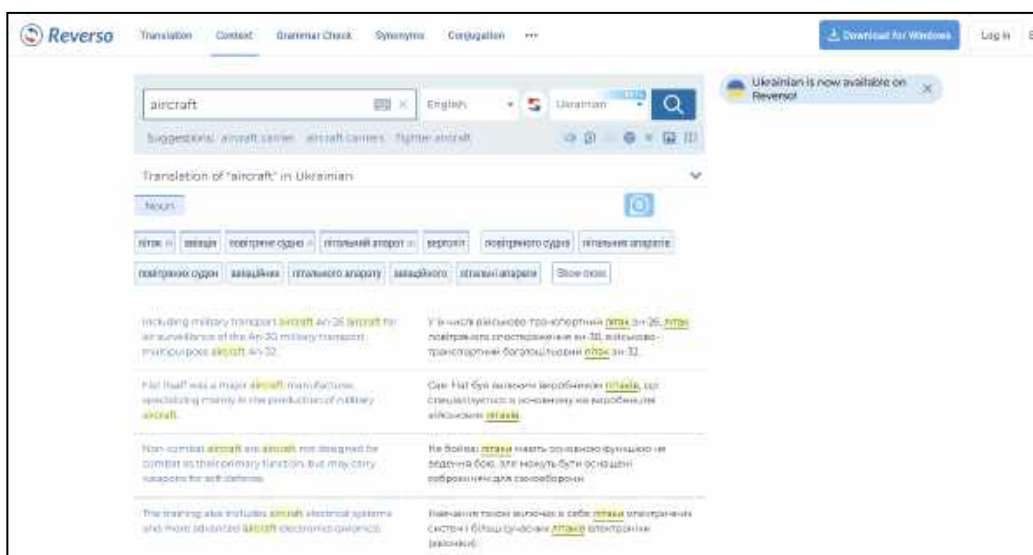


Рис. 2.7. Вигляд онлайн-версії Reverso Context

пошуки також будуть доступні на вебсайті Reverso Context на всіх пристроях користувача.

Reverso Context – великий багатомовний словник, який має багато контенту та різних функцій: містить спеціалізовану термінологію для бізнесу, фінансів, медицини чи техніки, більше ідіом, зворотніх перекладів. З простотою та потужністю пошукової системи та великим словниковим запасом Reverso Context надає релевантні переклади та приклади в контексті майже для будь-якого слова чи вираження. Завдяки зручному інтерфейсу Reverso Context підходить як для новачків, так і для професіоналів. Це потужний, цікавий, навчальний та безкоштовний інструмент.

### **Характеристики словника Reverso Context:**

- забезпечує миттєвий переклад 11 мовами: іспанською, французькою, італійською, англійською, португальською, німецькою, польською, голландською, арабською, російською та івритом;
- приклади використання в контексті (початковий текст та перекладений текст із виділеним пошуковим терміном та його перекладами), вилучені з бази даних, що складається із сотень мільйонів проіндексованих слів;
- словникові списки на основі особистого вибору прикладів та перекладів користувача, доступних в автономному режимі;
- природна вимова повних прикладів речень;
- один клік, щоб побачити зворотні переклади, подробиці частоти, визначення або сполучення, коли це можна застосувати;
- історія пошуку, доступна в автономному режимі та доступна на вебсайті Reverso Context, якщо користувач зареєструвався безкоштовно;
- пропозиції слів та виразів користувачу під час введення тексту;
- flashcards, щоб допомогти запам'ятати слова або фрази, які користувач шукав раніше.

Окрім зазначених базових електронних словників, здобувачам, що навчаються на прикладних лінгвістів, слід знати про такі онлайн-словники:

1. <http://www.thefreedictionary.com/>
2. <http://www.merriam-webster.com/dictionary/>
3. <http://dictionary.reference.com/browse/>
4. <http://www.oxforddictionaries.com/definition/english/>
5. <http://en.wikipedia.org/wiki/>

**Домашнє завдання.** Create a one-word dsl-dictionary applying a picture if possible. To perform the task use one of the following words.

### 1. **Cheugy**

This word (pronounced *choo-gee*) has swooped in to replace old fan-favorite «basic». It refers to the painfully mainstream or, along the same vein, someone hanging onto things that were cool years ago but would now be deemed basic or «cheugy».

### 2. **Rent free**

Can't stop thinking about someone? Sounds like they're living rent free in your mind. This is most often used when you can't get something out of your brain – whether it's a song, video, experience or person. They're stuck in there and they're not even paying rent for the space they're occupying.

### 3. **Vibe check**

Is someone acting shady or negative? Sounds like they didn't pass the vibe check. This describes when you check in on someone's vibe and assess what it's giving. Good vibes? You've passed the vibe check.

### 4. **Main character**

This one is fairly self-explanatory. When you call someone the main character, they're basically the core person in a situation. For example, in a video where someone seems to really stand out or is especially well dressed, funny, etc. you can identify them as the «main character» or as having «main character energy».

### 5. **Caught in 4k**

Caught someone red-handed and have the receipts to prove it? You've caught them in 4k with solid, digital evidence. Another popular phrase on TikTok, it can often be accompanied by the camera emoji to really bring home the fact that you've been exposed.

### 6. **Understood the assignment**

You can use this for someone that is really killing it in some (or any) aspect of their lives. They're giving it their all and doing an amazing job. Whether it's related to something they're doing, wearing, etc. they really understood the assignment.

### 7. **Bussin'**

You can use this when something is really, really good – often relating to food. If you're eating a really great meal you could say, «This food is bussin'». Want to add some extra emphasis? Just double up: «Wow this pizza is really bussin' bussin'».

### 8. **The blueprint**

This describes when someone is so perfect, or did so well at something, that they are deemed the original (comparable to «OG»). They are the

blueprint, an icon, a legend. They did it first and the rest simply followed in their exemplary footsteps.

### 9. Ate that

This essentially refers to someone doing a great job. For example if someone is performing a very impressive dance, when they're done, you might turn to your friend and say «they ate that». It can also be abbreviated to «they ate» or even «left no crumbs» (since, as we established, they ate it all).

## 3. СИСТЕМИ АВТОМАТИЧНОГО (МАШИННОГО) ПЕРЕКЛАДУ (COMPUTER-AIDED/MACHINE TRANSLATION)

**Машинний переклад (МП, Machine Translation)** — переклад текстів з однієї природної мови на іншу за допомогою спеціальної комп'ютерної програми.

Програми-перекладачі просто незамінні, коли виникає необхідність: швидко зрозуміти загальний зміст тексту; швидко проаналізувати багатомовну інформацію з інтернету; оптимізувати переклад великого обсягу текстів з однієї тематики (процес відбувається у два етапи: машинний переклад і постредагування людиною); зробити інформацію на сайті зрозумілою іноземним користувачам, вмонтувавши функцію перекладу на сайт; швидко листуватися з іноземними партнерами, а також вільно спілкуватися в соціальних мережах і на форумах; витягти інформацію з текстів великого обсягу з огляду на лінгвістичні характеристики слів. Наприклад, з'ясувати, скільки разів у тексті трапляються власні імена, які це імена і які дії пов'язані із цими іменами.

### 3.1. Типологія систем машинного перекладу

**Rule-based Machine Translation (RBMT, Машинний переклад, який ґрунтується на правилах мови).** Ця технологія враховує словникову інформацію та аналізує граматичні правила конкретних мов. Такі системи побудовано на основі лінгвістичного опису двох природних мов (двомовних словників та інших баз даних, що містять морфологічну, граматичну та семантичну інформацію), формальних граматик і власне алгоритмів перекладу. Якість перекладу залежить від обсягів лінгвістичних баз даних (словників) і глибини опису природних мов, тобто необхідного урахування максимальної кількості особливостей граматичної структури як вихідної мови, так і мови перекладу. Є два типи систем rule-based. **Системи типу Transfer** передбачають морфологічний, синтаксичний і семантичний

аналіз тексту вихідною мовою; перетворення у структуру цільової мови (TRANSFER); синтез тексту цільовою мовою. **Системи типу *Interlingua*** аналізують вхідний текст у термінах метамови і синтезують метаструктуру тексту мовою перекладу.

**Переваги RBM-систем:** синтаксична і морфологічна точність, стабільність і передбачуваність результату, можливість налаштування на предметну область. **Недоліки RBMT-систем:** складність і тривалість розроблення, необхідність підтримувати й актуалізувати лінгвістичні бази даних. Компанії-виробники систем машинного перекладу на основі технології rule-based – PROMT, Systran, LINGUATEC.

**Statistical Machine Translation (SMT, Статистичний машинний переклад).** Ця технологія передбачає пошук найбільш імовірного перекладу речення з використанням даних, отриманих із двомовної сукупності текстів. Такі системи перекладу працюють на основі порівняння великих обсягів корпусів паралельних текстів.

**Корпус паралельних текстів** – це тексти, що містять речення однією мовою і відповідні їм речення іншою мовою. Системи SMT здатні до «самонавчання»: чим більше паралельних корпусів текстів вони містять і чим точніше ці тексти відповідають один одному, тим краще результат статистичного машинного перекладу.

**Переваги SMT-систем:** адекватність перекладу, легкість у побудові за достатньої кількості паралельних корпусів, застосовуваність технології до будь-яких мовних пар.

**Недоліки SMT:** обмеженість паралельних корпусів у природі, нездатність у повному обсязі правильно інтерпретувати морфологію і синтаксис, спотворення інформації (дублювання, пропускання, підмінювання інформації). Компанії-виробники систем статистичного машинного перекладу – PROMT, Google, SDL Language Weaver, Microsoft, Asia Online, IBM.

**Hybrid Machine Translation (HMT, Гібридний машинний переклад).** У зв'язку з тим, що технології SMT і RBMT мають свої недоліки і складності, а також досягли певної межі у своєму розвитку, розробники рішень із МП розраховують на технологічний прогрес у створенні гібридної технології перекладу. Ця технологія передбачає суміщення методів RBMT і SMT. Такий підхід дає змогу поєднати переваги обох технологій (граматичну точність під час перекладу від RBMT і адекватність перекладу від SMT). Компанії-виробники систем гібридного машинного перекладу – DeepL, PROMT, Systran.

### 3.2. Популярні онлайн-системи машинного перекладу

**DeepL Translate** (<https://www.deepl.com/ru/translator>). Безкоштовний онлайн-перекладач, який можна використовувати для перекладу слів, речень або окремих виразів 24 мовами світу. Тут застосовуються уроки машинного навчання для швидкого, більш точного перекладу, ніж в інших подібних перекладачах. Завдяки DeepL Translate можна копіювати, поширювати та зберігати переклад у форматі текстового файлу. Програма самостійно пропонує автоматичне закінчення речення. Перекладачем можна користуватися безплатно, з обмеженням 5000 символів на переклад або ж із платною версією без обмежень. Зручно використовувати й мобільний додаток. Перекладач автоматично розпізнає вихідну мову, а користувачу дає змогу залишити оцінку й додати власний, більш влучний переклад слова. Також можна послухати звучання тексту перекладу – доступна і така функція.

**Yahoo Babel Fish** ([www.babelfish.yahoo.com](http://www.babelfish.yahoo.com)). Одну з перших систем автоматичного перекладу в інтернеті було створено в 1997 р. компанією AltaVista. Потім змінилося керівництво на компанію Yahoo, яка інтегрувала технологію Babel Fish у власні продукти Search та Toolbar та оснастила сервіс підтримкою нових мов. Онлайн-перекладач дає змогу перекладати на 38 мовних напрямів фрагменти тексту довжиною до 150 слів, а також сторінки вебсайтів. Якість перекладу з англійської на російську, яку надає Babel Fish, зазвичай така сама або перевищує якість Google Translate, але поступається Translate.ru. Під час перекладу втрачається поділ на абзаци.

**Google Translate** ([translate.google.com/translate\\_t](http://translate.google.com/translate_t)). Сервіс автоматичного перекладу текстів і вебсторінок цієї компанії, схожий із онлайн-перекладачем компанії PROMT, але за функціональністю, швидкістю роботи й зручністю інтерфейсу кращий. Має зміну напрямів перекладу, причому підтримуються не тільки європейські мови, але й японська, китайська, корейська та арабська, а також можливість працювати з вебсторінками та пошуковими запитами. Система Google Translate має додатковий інструментарій. За допомогою нього користувач сервісу може протягом декількох секунд додати на свій сайт модуль автоматичного транслювання сторінок з однієї мови на іншу, а також налаштувати у браузері миттєвий переклад слів з англійської. Якість перекладу з англійської на російську, виконана Google Translate, найчастіше гірша, ніж у Translate.ru, але такої кількості напрямів перекладу не має жодна інша система.

**Windows Live Translator** ([windowslivetranslator.com](http://windowslivetranslator.com)). Це розробка компанії Microsoft. Може перекладати введений користувачем текст (до 500 слів) і вебсторінки, надаючи оригінал документа й результат його оброблення у двовіконному інтерфейсі, що спрощує аналіз перекладеної інформації. На відміну від більшості інших автоматичних перекладачів, Windows Live Translator не дає змоги перекладати з англійської на російську, хоча й має можливість перекладу з російської на англійську. Так само, як і Google Translate, цей перекладач можна розмістити на власному сайті і за допомогою панелі інструментів Windows Live Toolbar інтегрувати у браузер Internet Explorer. Безплатний онлайн-переклад можна зробити також за такими адресами:

- <http://www.freetranslation.com>;
- <http://www.tranexp.com:2000/Translate/result.shtml>;
- <http://translation.langenberg.com>/та ін.

### **Цікаво знати**

#### ***Два «і» зіпсували репутацію урядові Італії***

Прем'єр-міністр Італії та медіамагнат Сільвіо Берлусконі якимось заявив, що досягти максимального технічного прогресу країни допоможуть три «і»: Internet, inglesa, impresa (інтернет, англійська мова та підприємництво). За іронією долі, саме через Internet і англійську мову члени уряду Італії опинилися в незручному становищі. На офіційному сайті уряду з'явилися біографії міністрів, перекладені англійською мовою явно за допомогою системи машинного перекладу. Біографія Лучіо Станца, міністра інновацій і колишнього віцепрезидента IBM, містила, наприклад, таке: *«Народився в Лусери (Фоджуа) 20 жовтня 1941 року. Парний і воно має двох дочок. У 1965 році хтось одержав учений ступінь по економії біля Університету Повних ротів Мілана»*. Дісталось й Ренато Руджеро, міністру закордонних справ (який, як і Станца, вільно говорить англійською мовою): *«У 1977 воно було Мегафоном президента Європейської комісії, (Роя) Ріючі Дженкінса, беручи участь до роботах, що принесли до запровадження в дію європейської грошової системи»*. Лише біографія самого прем'єр-міністра Берлусконі була складена на прийнятній англійській. Адміністрація сайту запевняла, що «експериментальні сторінки» з перекладеними біографіями містилися лише в тимчасовій кеш-пам'яті й більшості інтернет-користувачів були недоступні. Проте переклади були надіслані одним із читачів у газету «La Repubblica», видання що й



опублікувало повідомлення про те, як машинний переклад «заплямував» біографії членам уряду [3].

**Домашнє завдання. Translate the following text into Ukrainian using Google Translate and then edit it.**

Computer-assisted translation and machine translation

Some advanced computer-assisted translation solutions include controlled machine translation (MT). Integration of MT into computer-assisted translation has been implemented in various ways by various parties, especially by the language industry. Although this type of technology is neither widely known nor available to individual translators, carefully customized user dictionaries based on correct terminology significantly improve the accuracy of MT, and as a result, they improve the efficiency of translation process.

Translation memory software Translation memory (TM) programs store previously translated source texts and their equivalent target texts in a database and retrieve related segments during the translation of new texts.

Such programs split the source text into manageable units known as «segments». A source-text sentence or sentence-like unit (headings, titles or elements in a list) may be considered a segment, or texts may be segmented into larger units such as paragraphs or small ones, such as clauses. As the translator works through a document, the software displays each source segment in turn and provides a previous translation for re-use, if the program finds a matching source segment in its database. If it does not, the program allows the translator to enter a translation for the new segment. After the translation for a segment is completed, the program stores the new translation and moves onto the next segment. In the dominant paradigm, the translation memory, in principle, is a simple database of fields containing the source language segment, the translation of the segment, and other information such as segment creation date, last access, translator name, and so on.

#### **4. СИСТЕМИ АВТОМАТИЗОВАНОГО ПЕРЕКЛАДУ (COMPUTER-ASSISTED TRANSLATION)**

**Автоматизований переклад (АП, англ. Computer-assisted Translation)** – переклад текстів на комп'ютері з використанням комп'ютерних технологій. Від машинного перекладу (МП) відрізняється тим, що весь процес перекладу здійснюється людиною, комп'ютер

лише допомагає отримати готовий текст або за менший час, або з кращою якістю.

Ідея АП з'явилася з моменту появи комп'ютерів: перекладачі завжди ставали проти стандартної концепції МП минулих років, на яку було спрямовано більшість досліджень у галузі комп'ютерної лінгвістики, але підтримували використання комп'ютерів для допомоги перекладачам. У 1960-ті роки Європейське об'єднання вугілля і сталі (попередник сучасного Євросоюзу) стало створювати термінологічні бази даних під загальною назвою Eurodicautom. У Радянському Союзі для створення баз такого типу було створено «ВІНІТІ».

У сучасній формі ідею АП було розвинуто у статті Мартіна Кея від 1980 року, який висунув таку тезу: «By taking over what is mechanical and routine, it (computer) frees human beings for what is essentially human» (Комп'ютер бере на себе рутинні операції і звільняє людину для операцій, що вимагають людського мислення) [цит. за 6].

У наш час найбільш поширеними способами використання комп'ютерів під час письмового перекладу є робота зі словниками та глосаріями, пам'яттю перекладів (англ. Translation Memory, ТМ), що містить приклади раніше перекладених текстів, а також використання так званих корпусів, великих колекцій текстів на одному або декількох мовах, що дає стислий опис того, як слова і вирази реально використовуються в мові в цілому або в конкретній предметній області [6].

Для **локалізації** програмного забезпечення часто застосовуються спеціалізовані засоби, наприклад **Passolo**, які дають змогу перекладати меню та повідомлення у програмних ресурсах і безпосередньо в компільованих програмах, а також тестувати коректність локалізації.

Під **локалізацією** (мовною локалізацією) мають на увазі переклад та культурну адаптацію продукту до особливостей певної країни, регіону або групи населення. Найчастіше процес локалізації передбачає переклад і культурну адаптацію програмного забезпечення, відеоігор або вебсайтів. Процес локалізації може включати адаптацію графічного компонента, символів валют, формату дат, адрес і номерів телефону тощо.

Локалізація — це більше ніж переклад. Її кінцева мета — це не лише переклад тексту іншою мовою, а повністю адаптована мовна версія, яка має максимально природний вигляд для іноземного користувача.

У процесі локалізації відбувається модифікація елементів інтерфейсу, мультимедійних файлів, схем, числових форматів. Це особливо важливо під час створення мовних версій сайтів, запуску

мобільних застосунків, програмного забезпечення та інших ІТ-проектів на зарубіжному ринку.

**Домашнє завдання.** Translate the following Word menu into Ukrainian (рис. 4.1).

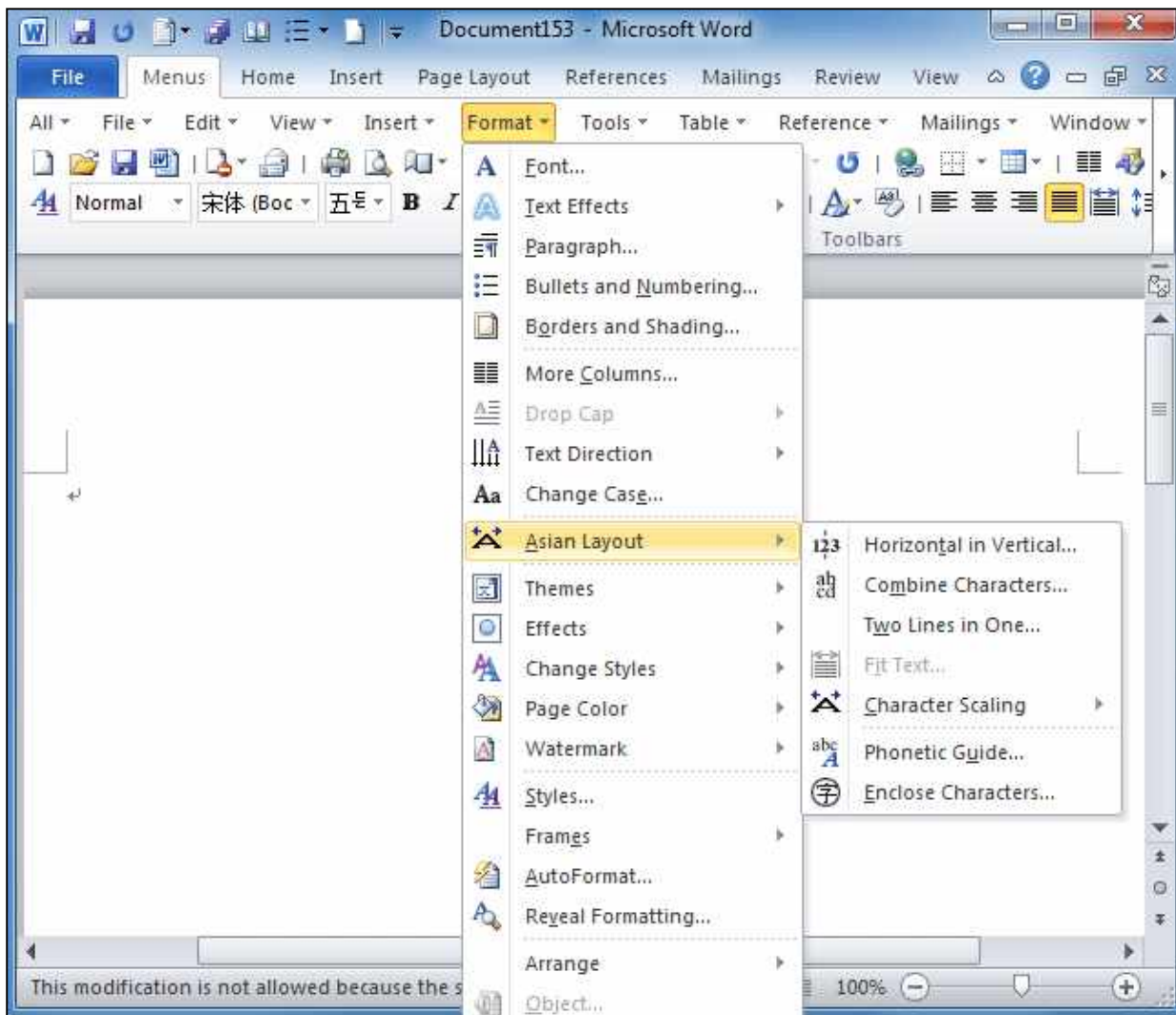


Рис. 4.1. Меню текстового редактора Word

Для **перекладу аудіовізуальних матеріалів** (передусім фільмів) також використовуються спеціалізовані засоби, наприклад **Swift**, у яких поєднано деякі аспекти пам'яті перекладів, але додатково забезпечено можливість появи субтитрів за часом, їх форматування на екрані, відповідність відеостандартам тощо.

У **синхронному перекладі** використання засобів автоматизованого перекладу за потребою обмежене. Одним із прикладів є використання словників, що завантажуються на КПК. Іншим прикладом може бути напівавтоматичний витяг списків термінів під час підготовки до синхронного перекладу у вузькій предметній області [3].

У вузьких предметних областях за великої кількості вихідних текстів і сталої термінології перекладачі можуть використовувати й машинний переклад, який може надати гарну якість перекладу термінології та стійких виразів. Перекладач у цьому випадку здійснює постредагування отриманого тексту. Більшість текстів Єврокомісії (передусім юридичні тести і поточна кореспонденція) перекладається з використанням МП.

**Пам'ять перекладів** (ПП, англ. translation memory, ТМ) – це база даних, де зберігаються виконані переклади. Технологія ТМ працює за принципом накопичення: у процесі перекладу в ТМ зберігається початковий сегмент (речення) і його переклад. Під час оброблення нового тексту, що надійшов для перекладу, система порівнює кожне його речення із збереженими в базі сегментами. Якщо ідентичний або подібний сегмент вихідного тексту знайдено, то переклад цього сегмента відображається разом із перекладом і зазначенням збігу у відсотках. Слова і фрази, які відрізняються від збереженого тексту, виділяються підсвічуванням. Отже, перекладачеві залишається перекласти тільки нові сегменти й відредагувати ті, що частково збігаються. Кожна зміна або новий переклад зберігаються в ТМ. А в результаті немає необхідності двічі перекладати те саме речення.

Зазвичай запис пам'яті перекладу складається з двох сегментів: вихідною (source segment) і кінцевою (target segment) мовами. Якщо ідентичний (або схожий) сегмент вихідної мови трапляється в тексті, сегмент кінцевою мовою буде знайдено в пам'яті перекладу та запропоновано перекладачеві як основа для нового перекладу. Автоматично знайдений текст може бути задіяно як є, відредаговано або повністю відкинуто. Більшість програм використовують алгоритм нечіткої відповідності (fuzzy matching). Цей алгоритм істотно поліпшує функціональні можливості програми, оскільки так можна знаходити речення, які лише нагадують шукані фрази, але так само придатні для подальшого редагування.

Крім прискорення процесу перекладу повторюваних фрагментів і змін, внесених у вже перекладені тексти (наприклад, нових версій програмних продуктів або змін у законодавстві), системи ПП також забезпечують однаковість перекладу термінології в однакових фрагментах, що особливо важливо для технічного перекладу. Проте якщо перекладач регулярно підставляє у свій переклад терміни з баз перекладів, що начебто є відповідними в запропонованому сегменті, то без контролю їх використання в новому контексті якість перекладеного тексту може погіршитися. У кожній конкретній системі ПП дані зберігаються у своєму власному форматі (текстовий формат у Wordfast, база даних Access в Deja Vu), але є міжнародний стандарт TMX (англ.

Translation Memory eXchange format), заснований на XML (мова програмування), і з яким можуть працювати практично всі системи ПП. Завдяки цьому результати роботи перекладачів можна обмінювати між різними CAT-програмами, тобто перекладач, що працює з OmegaT може використовувати ПП, створену в Trados і навпаки. Більшість систем ПП підтримують створення і використання словників користувача, створення нових баз даних на основі паралельних текстів (англ. alignment), а також напівавтоматичний витяг термінології з оригінальних і паралельних текстів.

Відповідно до опитувань 2022 року до найбільш популярних CAT-програм належать: SDL Trados Studio 2022, SmartCat, Deja Vu, OmegaT, SDLX Trados, Star Transit, WordFast, Across Personal Edition тощо.

У цьому посібнику буде детально розглянуто процес перекладу в програмі SDL TRADOS STUDIO 2022, що має найбільший попит.

## 5. SDL TRADOS STUDIO 2022

**Trados** – система автоматизованого перекладу, спочатку (з 1992 року) розроблена німецькою компанією Trados GmbH. Є одним зі світових лідерів у класі систем Translation Memory (ТМ, Пам'ять перекладів).

Система Trados складається з модулів, призначених для перекладу текстів різного формату: документів Microsoft Word, презентацій Power Point, текстів у форматі HTML та інших метаданих, документів FrameMaker, InterLeaf тощо, а також для ведення термінологічних баз даних (модуль MultiTerm). Остання версія Trados на сьогодні – SDL Trados Studio 2022.

Концепція Translation Memory передбачає виявлення в перекладному тексті фрагментів, переклади яких уже є в базі даних перекладів, і завдяки цьому скорочується обсяг роботи перекладача. Фрагменти, що залишилися не перекладеними, передаються далі перекладачу для ручного опрацювання або системі машинного перекладу (Machine Translation, МТ). Перекладач на цьому етапі може виділити знову перекладені фрагменти і додати нові пари паралельних текстів на двох мовах у базу даних. Така схема найкраще працює для однотипних текстів, де повторюваність словосполучень досить регулярна (наприклад, інструкції для користувачів, технічні описи тощо).

**Домашнє завдання. Translate SDL Trados Studio software mini-glossary into Ukrainian.**

1. **SDL Trados Studio.** An application for translating files, creating and managing translation memories, for automated project creation and automated translation.

2. **SDL MultiTerm.** A terminology management tool (separate application) that interacts with SDL Trados translation environments.

3. **SDL Passolo Essential.** A software localization tool that enables the translation of user interfaces.

\***SDL Trados 2007 Suite.** This is an older tool set based on other interface paradigms. Since the launch of SDL Trados Studio 2014 it is no longer included in the SDL Trados Studio product, but remains in widespread use. SDL Trados 2007 comprised the following distributives, which are no more separate and are all confined within **SDL Trados 2022** interface:

- SDL Trados Server Manager** is the client used by system administrators to manage serverbased translation memories and user access rights to translation memories.

- Synergy** is the SDL Trados 2007 Suite project management tool where projects are created and distributed and project team members open and translate project packages.

- Translator’s Workbench** is where translation memory administration is performed.

- TagEditor** is where tagged files are translated. The special editing environment ensures that information in tags is not lost during translation.

- SDL Passolo Essential.**

## 5.1. Особливості інтерфейсу SDL Trados 2022

Версії продукту оновлюються щороку. Останньою версією, яка вийшла на ринок, є SDL Trados 2022. У цьому посібнику буде розглянуто популярну версію від 2022 року, яка є надійною та зручною у використанні. За бажанням можна придбати ліцензійну нову версію (рис. 5.1).

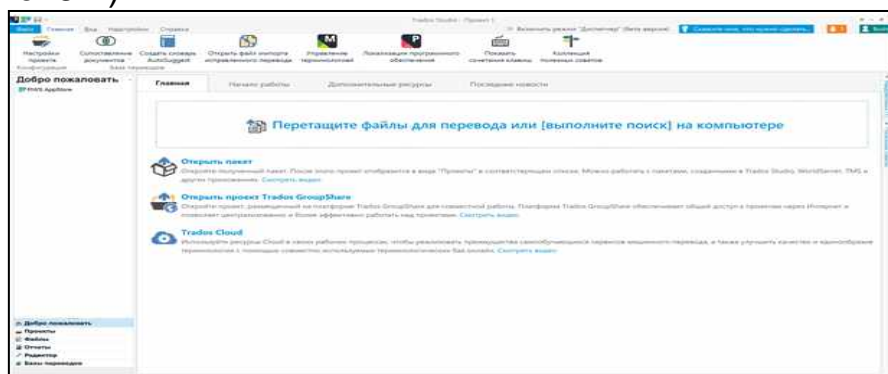


Рис. 5.1. Интерфейс SDL Trados 2022

Для відображення певного режиму роботи у SDL Trados Studio треба натиснути мишею на зображення з назвою режиму роботи або на значок цього режиму. Кнопки запуску режимів роботи розташовані в нижній частині панелі навігації (табл. 5.1).

Таблиця 5.1

### Режими роботи SDL Trados 2022

Назва режиму	Призначення
Home	У цьому режимі роботи можна виконувати такі команди: відкрити документ для перекладу відкрити підпроект; визначити основні налаштування за замовчуванням
Projects	Тут здійснюється перегляд проєктів та робота з ними. Можна вибрати проєкт, переглянути детальну інформацію щодо проєкту й файлу та відстежити статус файлу
Files	Тут здійснюється робота з файлами проєкту. У цьому режимі роботи можна виконати такі дії: відкрити файли для перекладу; відкрити файли для перевіряння; виконати пакетне оброблення файлів; крім того, можна побачити кількість слів і перебіг виконання перекладу цих файлів
Reports	Тут можна переглянути звіти з проєктів. У звітах наведено дані докладного аналізу, які безпосередньо передаються у процес планування проєкту та бюджету
Editor	Тут здійснюється переклад та редагування перекладу
Translation Memories	У цьому режимі роботи здійснюється створення баз перекладів та керування ними

## 5.2. Ресурси, якими забезпечує перекладача система Trados

1. Бази перекладів (Translation Memories). Файловий формат бази перекладів у SDL Trados Studio – .sdltm. Дозволено використання бази перекладів з SDL Trados Studio 2009 і 2011. Також можна під'єднати серверні бази систем SDL TM Server 2009 SP3, SDL TM Server 2011 і SDL WorldServer 2011 SP1 (версія 10.1).

2. Сервер автоматичного перекладу (Automated Translation Server). Переклад виконується автоматично, без участі людини. Додаток для автоматичного перекладу встановлюється на сервері,

доступ до якого здійснюється аналогічно доступу до серверної бази перекладів. Для під'єднання доступні такі сервери автоматичного перекладу: сервер автоматичного перекладу SDL за умовчанням, SDL BeGlobal, Google Translate.

3. Термінологічні бази. У SDL TRADOS STUDIO 2022 використовують файлові термінологічні бази даних SDL MultiTerm 2009 і 2011. Файловий формат – .sdltb. До того ж можна використовувати термінологічні бази, створені в більш ранніх версіях SDL MultiTerm: SDL MultiTerm 2007 (версія 7.5) і попередні випуски 7.x SDL MultiTerm iX (версія 6.x). Під час запуску термінологічних баз даних попередніх версій автоматично виконується їх оновлення до формату SDL MultiTerm 2011 (.sdltb). У MultiTerm Server 2009 SP3 і MultiTerm Server 2011 можна під'єднатися до серверних термінологічних баз даних.

4. Словники AutoSuggest (Автопідставлення). Редагування із використанням автопідставлення слів (AutoSuggest) – дуже важлива нова функція, що дає змогу прискорити ручний переклад.

Під час виправлення з використанням автопідставлення слів відстежується текст, який набирає перекладач – і після введення декількох перших символів слова з бази перекладів пропонується список слів і фраз мовою перекладу, які починаються із цих символів. Для автопідставлення слів можуть використовуватися *словники AutoSuggest* (можна створити новий словник AutoSuggest або вибрати наявний словник, створений яким-небудь користувачем). У цьому словнику наведено слова і словосполучення з баз перекладів .sdltm або файлів .tmx.

**Примітка.** У SDL Trados Studio 2022 для створення словника AutoSuggest досить бази перекладів із 10 тис. одиниць.

### **5.3. Здійснення перекладу за допомогою редактора SDL Trados 2022**

1. У рядку меню треба вибрати **Tools> Options**. Відкриється діалогове вікно **Options**.

2. У дереві навігації треба вибрати **Language Pairs> All Language Pairs> Translation Memory and Automated Translation**.

3. На сторінці **Translation Memory and Automated Translation** можна вибрати файлові і серверні бази перекладів. Слід вибрати заздалегідь створений тестовий файл бази перекладів **.sdltm** (пробну базу перекладів можна створити з двох текстових файлів: із текстом мовою оригіналу і з текстом перекладу завдяки функції **WinAlign**, яку буде описано далі).

Якщо під час налаштування не використовуються опції для мовної пари за замовчуванням, **English (United States) – German (Germany)**,



з'явиться діалогове вікно **Add Supported Language Pairs**. Треба встановити галочку поруч із мовною парою та натиснути кнопку «OK».

4. У дереві навігації вибрати **Language Pairs > All Language Pairs > Termbases**.

5. На сторінці **Termbases** можна вибрати термінологічні бази **SDL MultiTerm**. Треба вибрати файл заздалегідь створеної або запозиченої в іншого користувача тестової термінологічної бази з розширенням **.sdltb**.

6. У дереві навігації вибрати **Language Pairs > English (United States) > Ukrainian (Ukraine) > AutoSuggest Dictionaries**.

**Примітка.** Словники **AutoSuggest Dictionaries** можна налаштувати тільки для однієї мовної пари. На сторінці **AutoSuggest Dictionaries** можна вибрати словники **AutoSuggest**.

Наприклад, потрібно перекласти **Nokia 2610 User Guide.doc** з англійської мови (США) на українську (Україна). Для відкриття файлу й застосування параметрів настроювання, визначених раніше в діалоговому вікні **Options**, слід виконати такі дії:

1. Натиснути **Open Document** на панелі інструментів **Standard** в будь-якому режимі роботи. Відкриється діалогове вікно **Open Document** (рис. 5.2).

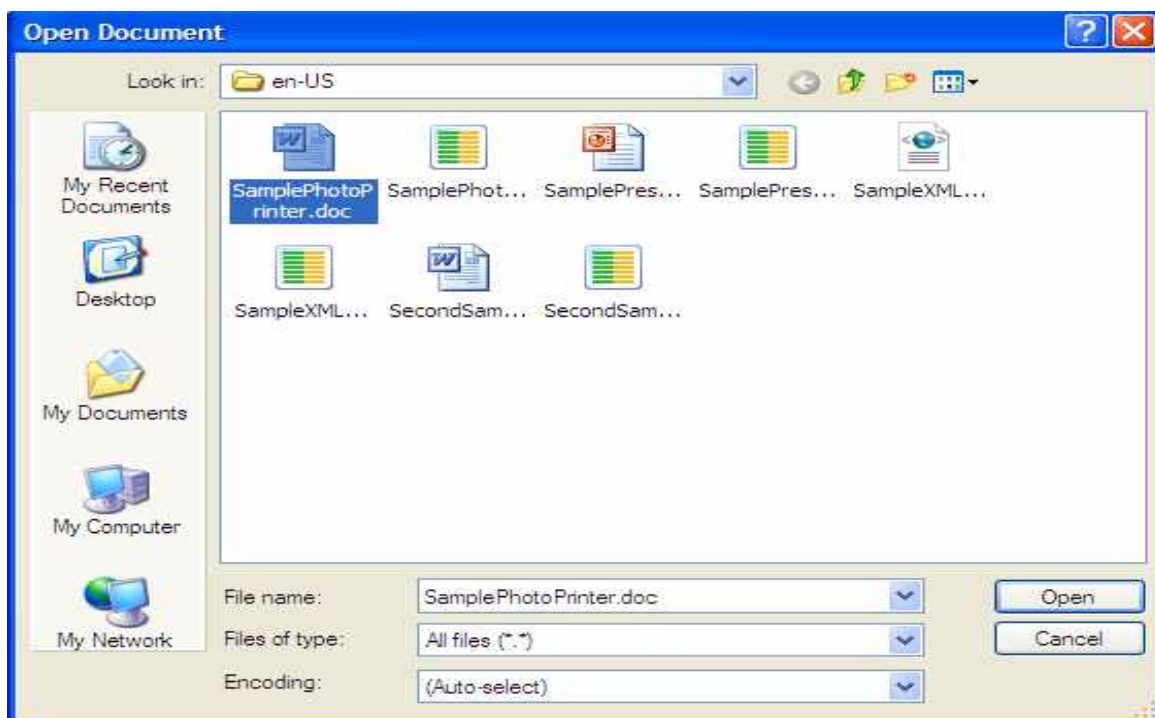


Рис. 5.2. Діалогове вікно **Open Document**

2. Вибрати необхідний файл (**Nokia 2610 User Guide.doc**) і натиснути **Open**. Відкриється діалогове вікно **Open Document**. У полі

**Translation Memory and Automated Translation** автоматично відобразатиметься база перекладів, обрана в попередньому розділі.

3. Натиснути **OK**: контент документа мовою оригіналу, що можна перекладати, визначається, розділяється на сегменти і додається у двомовний файл **.sdlxliff**. Файл відкриється для перекладу у вікні редагування режиму роботи **Editor**. Автоматично буде відкрито зазначені раніше бази перекладів за замовчуванням, термінологічні бази даних і словники **AutoSuggest**. Також буде автоматично застосовано всі настройки, визначені в діалоговому вікні **Options**.

4. На панелі інструментів натиснути **Save Document**: відкриється діалогове вікно **Save as**. Автоматично для збереження документа буде запропоновано ім'я **Nokia 2610 User Guide.doc\_en-US\_ua-UA.sdlxliff**. Натиснути **Save**, документ буде збережено із цим ім'ям. Автоматично буде створено проєкт із тим самим ім'ям, що й документ.

#### 5.4. Основні функції редактора Trados

У SDL Trados Studio перевірка та переклад документів здійснюються в режимі роботи **Editor**. Цей режим роботи має таке:

- панель навігації, на якій можна переглянути відкриті в цей час документи і вибрати необхідний;
- меню і панелі інструментів з інструментами редагування;
- вікно редагування, у якому здійснюється переклад і перевірка.

У вікні **Translation Results** відображаються результати пошуку в базі перекладів і створюються одиниці чорнового автоматизованого перекладу. Якщо автоматично підставлений результат не підходить, можна вибрати один із запропонованих тут результатів і застосувати до поточного сегмента.

У вікні **Term Recognition** відображено результати пошуку в термінологічній базі.

У вікні редагування режиму роботи **Editor** виконується переклад документів. У цьому вікні паралельно відображаються двомовний документ **.sdlxliff** із текстом, який потрібно перекласти, і готовий переклад. Оригінальний текст відображається ліворуч, а версія мовою перекладу – праворуч. Зміст документа розділити на сегменти (переважно за реченнями). Цільові сегменти можна змінювати.

У першому стовпчику відображені номери сегментів. Відображення номерів сегментів можна вимкнути. З вихідними та цільовими сегментами можна працювати як із двома окремими списками.

Між стовпцями вихідних і цільових сегментів розташовано стовпець статусу сегмента. Цей стовпець дуже важливий, тому що в

ньому показано поточний статус перекладу сегмента і джерело перекладу.

Наприклад, можна побачити, що переклад затверджено, а в базі перекладів знайдено повний збіг для цього сегмента.

Праворуч від цільових сегментів розміщено стовпець структури документа. У ньому вказано код, за яким можна визначити розміщення тексту сегмента в початковому документі. Для перегляду опису розміщення тексту сегмента в початковому документі наведіть курсор на цей код і клацніть по коду в цьому стовпці.

Відкрийте знову ваш документ, буде автоматично виконано пошук збігів для першого сегмента. Результати пошуку в базі перекладів з'являться у вікні **Translation Results**, найкращий результат буде підставлено в цільовий сегмент документа. Якщо в базі перекладів було знайдено повний контекстний збіг, то переклад підтверджується автоматично. Тип збігу відображається в стовпці статусу сегмента.

Контекстним збігом називається 100%-й збіг у відповідному контексті, коли, наприклад, попередній сегмент у документі повністю збігається з аналогічним сегментом у базі перекладів. У цьому випадку повний контекстний збіг означає, що це речення вже було перекладено на початку документа. Передбачається, що контекст у базі перекладів збігається з контекстом прикладу, де ця пара речень також є першою. виправлення повного контекстного збігу зазвичай не потрібне.

Під час переміщення в наступний сегмент також виконується пошук у термінологічній базі. Якщо в термінологічній базі даних буде знайдено необхідний термін (наприклад, «home screen»), то у вихідному сегменті термін виділиться червоними дужками. У вікні **Term Recognition** буде відображено переклад знайденого терміна «головний екран». Цей переклад уже підставлено до цільового сегмента, тому що його було знайдено в базі перекладів.

**Примітка.** Для підставлення перекладу терміна треба натиснути **Ctrl + Shift + L**, вибрати **Insert term translation** у вікні **Term Recognition** або **Translation > Show Translated Terms** у рядку меню.

#### 5.4.1. Застосування форматування

Наприклад, частина тексту в якомусь сегменті виділена жирним шрифтом (**battery replacement**), а частина – курсивом (*lithium*). Застосувати форматування до тексту перекладу можна способами, наведеними нижче.

Спосіб 1. Необхідне форматування можна вибрати у випадному списку **QuickPlace**. Цей список називається QuickPlace, оскільки за

його допомогою можна швидко помістити елементи з початкового сегмента до цільового, наприклад **теги, цифри, дати** та інші типи підстановних елементів. Для отримання додаткової інформації див. **Вставлення підстановних елементів (Placeables)**.

Для застосування форматування зі списку **QuickPlace** в цільовому сегменті слід вибрати текст **заміна батареї** та натиснути **Ctrl +, (кома)**. Під цільовим сегментом відображається список **QuickPlace**, що містить опції форматування. Текст з форматуванням, застосованим у початковому сегменті, виділяється жирним шрифтом. Для вибору першого зразка з текстом зі списку, у якому використовується жирний шрифт, натисніть **Enter** або **Tab**.

Треба повторити описані вище дії для виділення курсивом тексту *lithium*. На цей раз у випадному списку **QuickPlace** потрібно вибрати зразок із текстом, виділеним курсивом.

Спосіб 2. Крім того, форматування можна вибрати на панелі інструментів **QuickInsert**. При цьому до перекладу можна навіть застосувати форматування, яке відсутнє в початковому сегменті.

У цільовому сегменті слід вибрати текст **«battery replacement»** і натиснути на панелі інструментів **QuickInsert** або натиснути **Ctrl + B**.

У цільовому сегменті вибрати текст **«lithium»** і натиснути на панелі інструментів **QuickInsert** або **Ctrl + I**.

#### 5.4.2. Переклад нетекстових файлів

Досить часто перекладачам доводиться мати справу з документами інших форматів, які не призначено для роботи з текстом, але в яких може міститися текст. Наприклад, передбачається, що існує лише дві програми, здатні самостійно впоратися з форматом проєктів, створених в AutoCAD (файли з розширенням **.dxf** або **.dwg**) – це **across** і **Star Transit**.

Проблемою може стати необхідність перекладу креслень у форматі файлів AutoCAD (з розширенням **dxf** або **.dwg**). Складність полягає в тому, що спочатку цей формат не було призначено для роботи з текстом, тому можливості його набору та редагування надзвичайно обмежено. Із цієї причини перекладачам довелося розробити методики перекладу файлів, які дають змогу задіяти засоби автоматизації не безпосередньо, а за допомогою інших програм, наприклад **Microsoft Word**. Можна виділити дві найбільш прості: переклад за допомогою Microsoft Word і переклад із використанням буфера обміну.

Обидві методики прості й передбачувані. Перша полягає в тому, що текст з креслення AutoCAD копіюється в документ **Microsoft Word**.

Таке копіювання можна здійснювати як вручну (операція **Копіювати > Вставити** або сполучення «гарячих» клавіш **Ctrl + C> Ctrl + V**), так і з використанням спеціальних програм, що «витягують» текст із креслень, наприклад за допомогою спеціального плагіну **AutoCAD**. Використання цього плагіну також пов'язане з низкою складнощів, наприклад з тим, що текст вставляється у Word не в порядку за змістом, а в порядку експортування.

Після того, як текст був тим чи іншим способом «витягнений» із креслення, можна його перекласти за вже знайомою технологією. Далі текст редагується й вставляється у креслення знову. Отже, під час перекладу зберігається можливість використання попередніх баз ТМ і термінологічних баз, однак досить велика ймовірність виникнення помилок, пов'язаних із неухважністю під час експорту та імпорту тексту.

### 5.4.3. Перелік швидких клавіш для роботи в SDL Trados 2022

З переліком швидких клавіш SDL Trados 2022 можна ознайомитися у режимі **HOME** або інших, де на верхній панелі міститься відповідна позначка клавіши. Якщо навести на неї курсор миші, то висвітиться назва **Shortcuts**. Основні комбінації швидких клавіш:

- **Ctrl + Insert** – копіювання тексту з початкового сегмента в сегмент перекладу;

- **F6** – переключення між початковим сегментом і сегментом перекладу;

- **Ctrl + F3** – пошук Concordance в початковому тексті;

- **Ctrl + Shift + F3** – пошук Concordance в тексті перекладу.

- \* **Concordance Search** – це не повнотекстовий пошук у повній пам'яті перекладів, а приблизний пошук, розроблений для якнайшвидшого виявлення значущих результатів у великих наборах даних. Хоча можна використовувати Concordance Search для слів і символів, цей алгоритм оптимізовано саме для пошуку словосполучень.

### 5.5. Використання функції Winalign

Функція Winalign виконує вирівнювання, тобто створення бази даних накопичених перекладів (ТМ) із старих документів, коли в користувача є і оригінал, і переклад (рис. 5.3). Під час вирівнювання треба задати напрям перекладу (з якої мову на яку здійснюється переклад), далі відбувається автоматичне сегментування тексту – поділ його на сегменти. Перекладач сам має перевірити коректність відповідності сегментів (чи збігається сегмент тексту мовою оригіналу із

сегментом тексту перекладу), якщо сегменти поєдналися некоректно або з'явилися пусті сегменти, перекладач власноруч їх об'єднує повторно, додає або видаляє [1, с.111 –133].

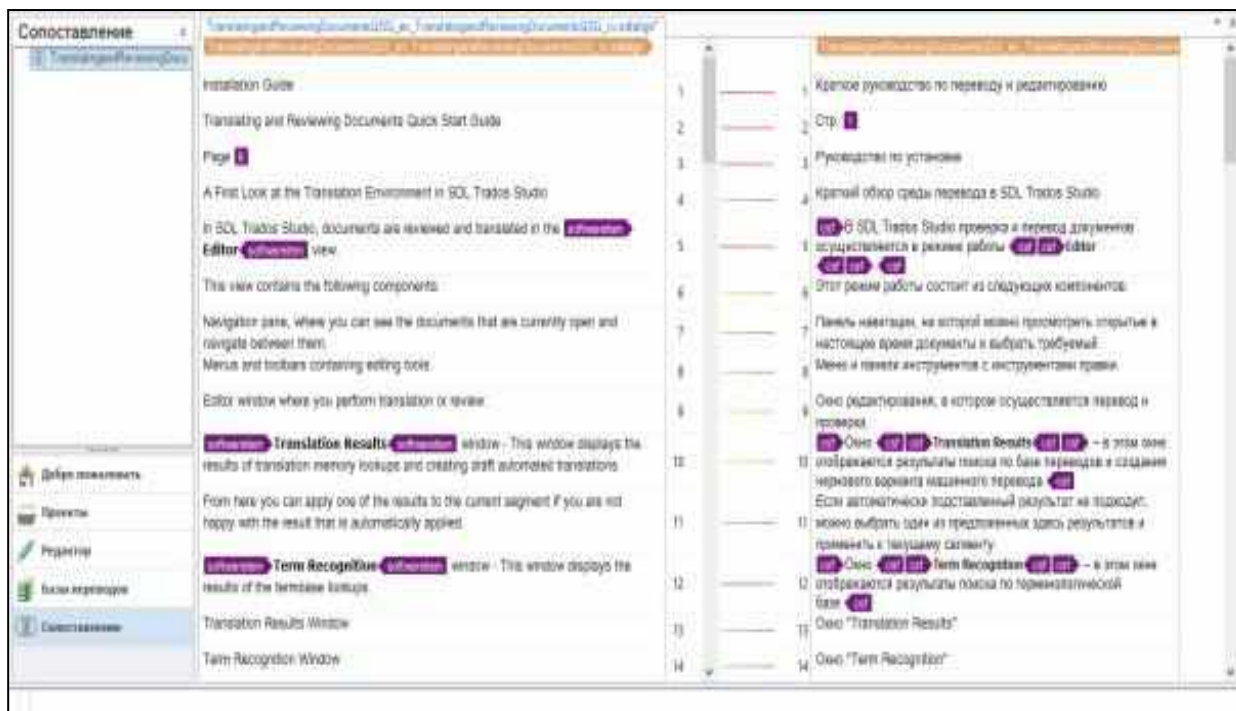


Рис. 5.3. Діалогове вікно WinAlign

## 5.6. Створення термінологічної бази Multiterm з словника термінів за допомогою Excel і MultiTerm Convert

Якщо наявний готовий словник перекладених термінів у форматі електронної таблиці Excel, то можна в простий спосіб конвертувати його в MultiTerm Convert.

### 5.6.1. Створення аркуша Excel із термінологією

Наприклад, є словник термінів (тема «Мобільні телефони»), що має вигляд таблиці Word (інші варіанти – аналогічно) і складається з трьох (або двох) стовпців:

- 1) терміни англійською мовою;
- 2) еквіваленти українською мовою;
- 3) визначення, коментарі, приклади тощо (цей стовпець користувач може створити на свій розсуд, можна обмежитися і двома).

Третій стовпець (і наступні) можна також використовувати для створення багатомовної термінологічної бази.

Натиснувши **Правлення> Виділити все**, треба виділити таблицю, копіювати та вставити її в Excel, у комірку A2. Отож таким чином стовпці (у комірках A1, B1, C1) матимуть такі назви: TermEng, TermRus, Def.

Потрібно зберегти аркуш Excel.

### 5.6.2. Перетворення аркуша Excel за допомогою MultiTerm Convert

Треба створити у вибраному місці папку **MultiTerm Bases**. Запустити **MultiTerm Convert** (рис. 5.4).

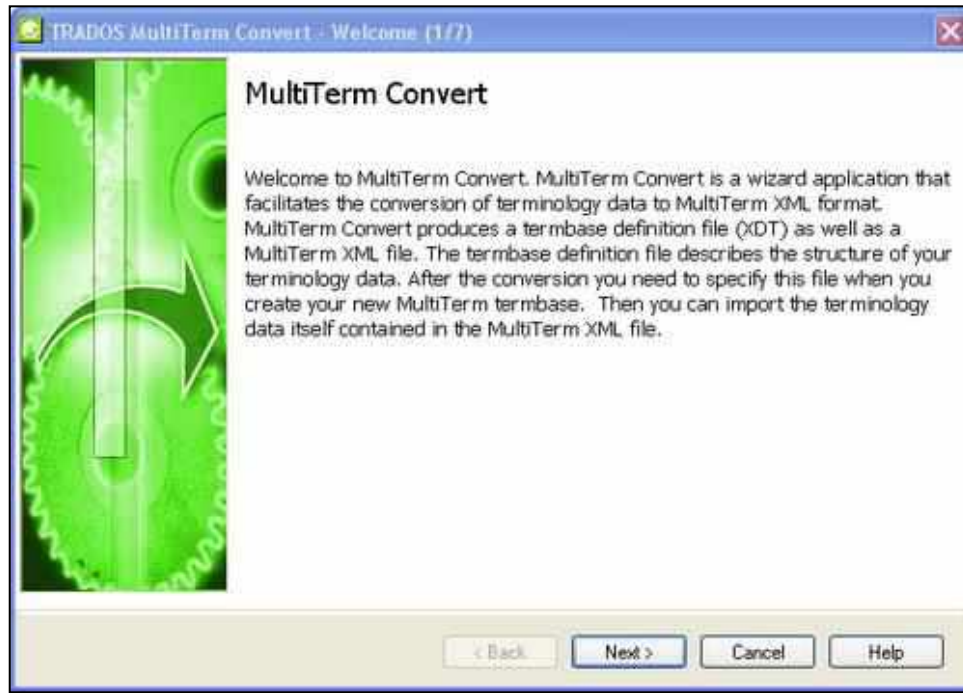


Рис. 5.4. Конвертор бази термінів у XML-формат

Потрібно натиснути **Next**. Залишити перемикач **New conversion session** і натиснути **Next**.

У вікні встановити перемикач **Microsoft Excel format**. Натиснути **Next**.

У вікні натиснути **Browse** праворуч від поля **Input file** і знайти аркуш **Excel**, який було створено (наприклад, NOKIA.xls). Потім потрібно вибрати цей аркуш, інші поля програма заповнить сама. Натиснути **Next**. Вибрати мову оригіналу: встановити перемикач **Index field** і вибрати відповідну мову. Повторюємо цю процедуру для мови перекладу. Для третього стовпця **Def** потрібно встановити перемикач **Descriptive field**, залишити **Text**. Натиснути **Next**.

За допомогою кнопки **Add** приєднати описове поле до структури ліворуч (виділити **Term** під **TermUa** і натиснути **Add**). З'явиться останній екран, коли ще можна повернутися і внести зміни. Потрібно натиснути **Next**, після чого почнеться процес перетворення. Після його закінчення натиснути **Next**, після чого з'явиться вікно з підтвердженням. Натиснути **Finish**.



### 5.6.3. Імпорт перетвореного аркуша Excel у MultiTerm

Потрібно запустити **MultiTerm**. У меню **Termbase** вибрати **Create termbase**. У вікні огляду папок вибрати місце зберігання термінологічної бази, папку **MultiTerm Bases**, і натиснути **OK**. Відкриється **Termbase Wizard**. Натиснути **Next**. У вікні встановити перемикач **Load an existing termbase definition file** і натиснути **Browse**. Потім потрібно перейти в папку, у якій було збережено перетворений аркуш **Excel** (папка **MultiTerm bases**), вибрати файл **NOKIA.xdt** та натиснути **Open**. Натиснути **Next**.

У вікні, що відкрилося, у полі **Name** треба ввести назву термінологічної бази (**NOKIA**), у двох інших (необов'язкових для заповнення) полях можна за бажанням написати будь-що. Потім слід натиснути **Next** чотири послідовні рази за запитами програми. Натиснути **Finish**. Порожню термінологічну базу створено. Тепер потрібно імпортувати терміни у створену термінологічну базу. Щоб можна було працювати одночасно з декількома термінологічними базами, потрібно всі такі бази об'єднати в один проєкт. У меню **Project** слід вибрати **Save Project As** (рис. 5.5).

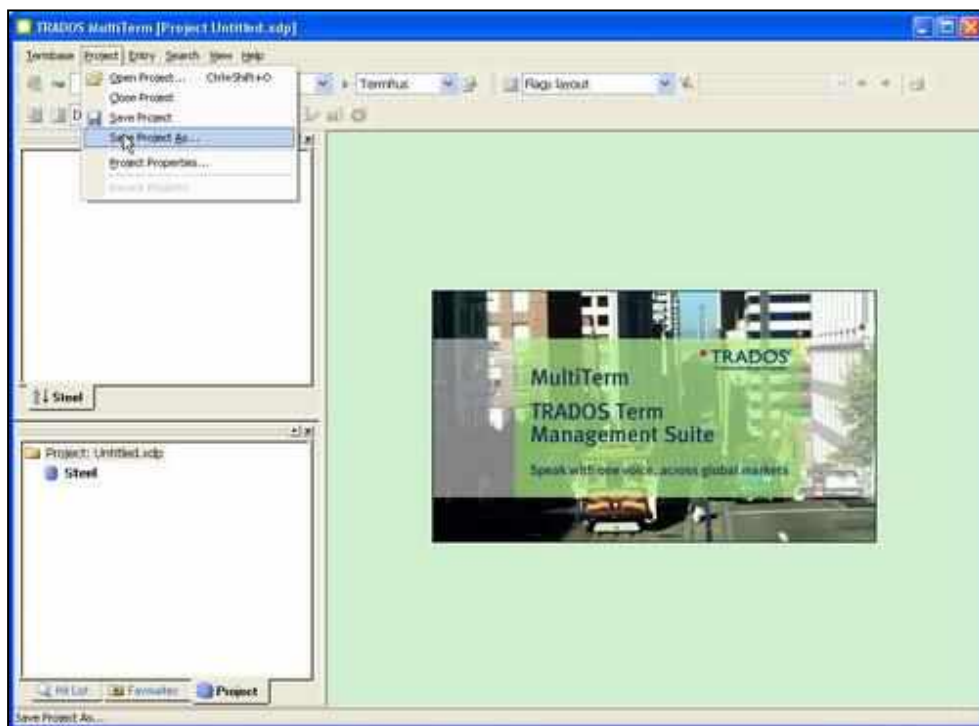


Рис. 5.5. Розміщення бази термінів у відповідному

Потрібно назвати проєкт (**MobilePhones**) і зберегти його в папку **MultiTerm bases**.

Далі натиснути правою кнопкою миші на назві термінологічній бази (**NOKIA**) і вибрати **Termbase Catalogue**. Відкрити закладку **Import**. Натиснути **Process**.



У вікні, що відкрилося, натиснути **Browse** і перейти в папку (зазвичай папка відкривається сама), де міститься файл, з якого програма внесе терміни у створену порожню базу, вибрати цей файл (**NOKIA.xml**) і натиснути **Open**. Натиснути **Next**. Зберегти файл винятків з ім'ям **NOKIAEx**. У цей файл вводяться дані про ті терміни, які не було введено **MultiTerm** за відомими йому причинами до термінологічної бази. Слід натиснути **Next** три рази за запитами програми. Натиснути **Finish**. Натиснути **OK**.

**Примітка.** Якщо необхідно імпортувати Excel-файл у наявну базу термінів або об'єднати дві бази разом, то можна скористатися підказками меню **Termbase>Package/DetachTermbase** (рис. 5.6).

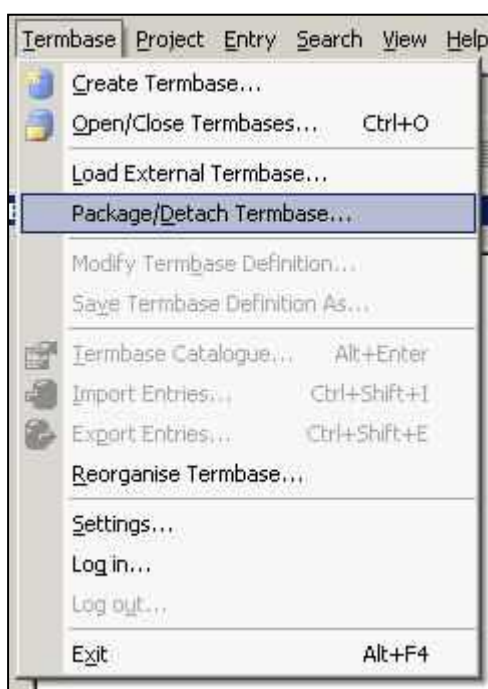


Рис. 5.6. Меню Termbase для роботи з термінологічними базами

**Примітка.** Переклад файлів із розширенням .pdf, .excel та ppt здійснюється так само як і текстових файлів .doc, .docx.

### **Домашнє завдання № 1. Translate the dialogue between two Trados 2011 users.**

Maria Elisa Manfrino, Italy:

I would like to buy SDL TRADOS STUDIO 2022 Freelance Plus but I was told that it does not include TagEditor and I presently use it a lot for translating excel and power point files.

Is that right?

Thank you.

Emma Goldsmith, Spain:

Studio 2011 doesn't include TagEditor (because it is not bundled with Trados 2007, which was the case of Studio 2009). However, there is no need for TagEditor to translate Ppt or Excel files. Just open them in Studio. Studio accepts a multitude of file formats and has a friendly Editor interface so it is streets ahead of TagEditor in every respect. If your clients need TTX files then you can convert the file to TTX format before opening it in Studio. In any case, if you are already using Trados 2007, you won't lose that licence so you could go on using Tag Editor if you really want to.

**Домашнє завдання № 2.** Prepare an excel-file with the dictionary that is valid for the ten texts you already have. Follow the procedure described above.

### 5.7. Визначення понять «проект» і «підпроект»

**Проект (Project).** Переклад усіх файлів і керування ними виконуються в межах проекту. Проект може складатися з одного або декількох файлів, які потрібно перекласти на одну або декілька мов. Також до його складу можуть входити довідкові матеріали, бази перекладів, словники AutoSuggest та інструкції для перекладачів. Після створення проекту менеджер може контролювати перебіг його виконання за допомогою інструментів керування проектами у SDL Trados Studio.

**Підпроект (Project Package).** Підпроект являє собою файлову структуру, що включає всі файли, які потрібно відправити члену проектної групи для роботи над проектом. Наприклад, файли, які необхідно перекласти або відредагувати. Після створення підпроекту його можна надіслати електронною поштою, викласти на сервер FTP або використувати інший бажаний спосіб відправки співробітнику, який буде працювати з файлами проекту. При цьому немає необхідності відправляти декілька файлів або повідомлень електронною поштою, а всю інформацію про проект можна відправити разом із файлами проекту в єдиній логічній структурі.

**Завершення підпроекти (Return Packages).** До складу завершеного підпроекту входять файли проекту, робота над якими завершена. Наприклад, перекладені або перевірені файли проекту. Також у ньому містяться файли, що потребують подальшого оброблення. Наприклад, якщо рецензент повертає перекладений файл із помилками, які має усунути перекладач. Створюваний завершений підпроект містить тільки файли проекту і не містить базу перекладів, термінологічні бази або словники AutoSuggest.

### 5.7.1. Керування проєктами

**PerfectMatch.** PerfectMatch належить до форми контекстної відповідності, за якої оновлені вихідні файли порівнюються з відповідним набором наявних двомовних документів, а не з базою перекладів. Сегменти, що збігаються, називаються одиницями PerfectMatch і перевіряються в контексті, тобто перевіряється відповідність сусідніх записів. Далі вони витягуються з наявних двомовних документів і переносяться в оновлені файли для перекладу. Оскільки під час зіставлення сегментів із застосуванням функції PerfectMatch виконується перевірка контексту, одиниці PerfectMatch зазвичай не потребують перекладу або редагування у процесі перекладу.

**Пакетне оброблення (Batch Task).** Це процес, що виконується щодо окремого файлу або групи файлів (наприклад, всього проєкту). Пакетне оброблення виконується системою. Наприклад, функція пакетного оброблення **Word Count** використовується для підрахування слів у вибраних файлах, результати чого відображаються в режимах роботи та звітах **SDL Trados Studio**. Функції пакетного оброблення можна включити в перелік завдань. Усі завдання виконуються (у зазначеному в переліку порядку) щодо всіх вибраних файлів.

**Перелік завдань (Task Sequence).** У перелік завдань входить кілька функцій пакетного оброблення. Після запуску переліку завдань усі вони виконуються щодо всіх вибраних файлів: у тому порядку, у якому зазначені в переліку. Є кілька наборів завдань пакетного оброблення. У версії Professional можна створювати нові переліки завдань.

## 6. АСПЕКТИ ЛОКАЛІЗАЦІЇ ПРИКЛАДНОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

Утвердження інформаційних технологій зумовило стрімкі темпи виробництва та поширення прикладних програмних продуктів світом, що також викликало потребу в їхній культурно-мовній локалізації (адаптації) для коректного функціонування та можливості розширення ринків збуту.

Термін «**локалізація**» (від лат. locus – місце) – це переклад вербального компонента програмного продукту відповідно до особливостей певної країни, регіону або групи людей, відповідно до їхніх соціально-культурних очікувань; перетворення готового

програмного продукту з метою використання його у країнах з іншою мовою; динамічний процес адаптації програмного забезпечення, вебсайтів або супутньої технічної документації до стандартів і мовних традицій країни, для якої призначається згадане ПЗ, вебсайт чи документація, які до того ж характеризуються глибоким 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 [5, с. 24]. Локалізація передбачає переклад документації та користувацького інтерфейсу, зміну методик розрахунків, а також деяке перероблення програми; відтворення програмної продукції засобами іншої мови як особливий жанр перекладу. Локалізацію програмних продуктів українською мовою називають українізацією, тобто впровадження української мови в програмному інтерфейсі.

Об'єктами, що підлягають локалізації на сучасному етапі розвитку інформаційних технологій, є програмне забезпечення, сайти, комп'ютерні ігри та мультимедійні засоби, де всі вказані об'єкти долучають також переклад супровідної документації, а останні два передбачають не тільки адаптацію графіки, а й голосове озвучування.

**Прикладне програмне забезпечення**, або програмні засоби (англ. software), – комплекс програмних засобів і супровідної документації, призначених для вирішення порівняно вузьких класів завдань у конкретних предметних галузях, розрахованих на певного споживача, а саме: науково-технічних, економічних, інженерних, конструкторських та інших спеціальних завдань у різних сферах людської діяльності. Процес локалізації передбачає такі етапи роботи перекладача: аналіз локалізованості (ступеня складності перекладу); укладення багатомовного словника проекту; локалізацію інтерфейсу й системних повідомлень; локалізацію довідкових файлів; складання локалізованої версії; локалізацію супутніх маркетингових матеріалів; лінгвістичне й функціональне тестування.

Умовно у процесі локалізації можна виокремити два аспекти: власне **лінгвістичний** та **технічний**, що постають комплексно.

Щодо лінгвістичного аспекту процес українізації як конкретний вид локалізації передбачає створення коректних та унормованих українізованих версій зарубіжних програмних продуктів із урахуванням таких базових особливостей, як-от: дотримання лексико-стилістичних, граматичних та синтаксичних норм, усталення комп'ютерної терміносистеми, уникання надмірної міжмовної інтерференції (з англійською мовою як мовою оригіналом та російською мовою як мовою посередником, оскільки найчастіше русифіковані версії програм

з'являються раніше через економічні причини), збереження прагматичної настанови та врахування культурної специфіки цільової мови (особливо для комп'ютерних ігор). Так, англійське слово *click* (*press one of the buttons on a mouse to select a function or item on the screen*) має декілька українських відповідників: 1) *цокати* (про підбори); 2) *клацати, клоцати* (про заціпку, металеві предмети); 3) *клікати* (*натискати і відпускати кнопку миші – калька з англійської мови*); 4) *натиснути*. Згідно зі статистикою локалізованих програм Microsoft найчастотнішим відповідником є *клацати*, потім *натиснути*. Отже, під час перекладу рекомендовано перш за все аналізувати контекст вживання, стилістичні норми й уникати зайвого калькування.

Цікавим випадком є команда *Burn* у програмі Nero, що записує компакт-диски. У її русифікованій версії перекладацьким відповідником є іменник *Прожиж*. З професіональної точки зору запис компакт-диску полягає у пропіканні його поверхні – способі малюванні точок на поверхні диску лазером. Але в українізованій версії замість відповідника *Пропікання* використовують термін *Записати*, таким чином уникаючи надмірної міжмовної інтерференції і надаючи зрозумілий термін замість професіоналізму.

З іншого боку, у зв'язку із впливом російської мови на українську, іноді калькований переклад залишається більш вживаним через свою усталеність. Наприклад, англійський термін *desktop* (*the working surface of a desk*) метафорично став позначати власне відображення вмісту екрана ПК під час його запускання (*the working area of a computer screen regarded as a representation of a notional desktop and containing icons representing items such as files*). Російською мовою його було метонімічно перекладено як *рабочий стол*. В українізованій версії маємо усталений відповідник *робочий стіл*. Зараз з'явився ще один відповідник *стільніця* (*верхня дошка, кришка стола*), що є перекладом прямого значення англ. *desktop* без метафоричного перенесення. Проте за статистикою локалізованих програм Microsoft новий термін не набув поширення.

Виникнення граматичних і синтаксичних труднощів зумовлене розбіжностями в будові англійської мови, що характеризується аналітичністю, та української мови, якій притаманна флективність. Наприклад, відомо, що стереотипними для прикладного англословного програмного забезпечення (продукції) є еліптичні конструкції, які нерідко вимагають змін граматичних форм, частин мови, змін порядку слів тощо та імперативних форм дієслів, які потребують заміни на неозначені форми дієслів.

Наприклад, переклад українською мовою деяких команд меню програми-перекладача SDL Trados 2022 із посиланням на Microsoft

Language Portal: «*Refresh View*» – «*Оновити вигляд*» (термін «*вигляд*» іноді змінюють на термін «*подання*»); «*Search Details*» – «*Ознайомитися докладніше*», російською «*Искать в подробных сведениях*»; «*Mark as complete*» – «*Позначити як завершений*» (замість «*завершені*», оскільки мова йде про один сеанс перекладу); «*Repeat Go to*» – «*Повторно перейти до*»; «*Batch tasks*» – «*Пакет задач*»; «*Failed to open File 1*» – «*Не вдалося відкрити Файл 1*» – як коректні переклади.

Для вирішення лінгвістичних труднощів локалізації корисними для перекладачів, окрім базових лексикографічних джерел, є спеціалізовані онлайн-ресурси та безкоштовні глосарії вже локалізованих програм. Одним із таких джерел є вищезгаданий сервіс Microsoft Language Portal, що являє собою багатомовний онлайн словник комп'ютерних термінів.

Технічний аспект локалізації полягає в перекладі програмних інтерфейсів і документації до програмних продуктів за умови тісної взаємодії перекладача із програмістами/виробниками програмної продукції. Наприклад, переклад елементів інтерфейсу має враховувати те, що в різних мовах один і той самий за змістом текст може відрізнятися довжиною (кількістю символів), наприклад: англ. «*Default*» – укр. «*За замовчуванням*»; англ. «*Add entry*» – укр. «*Додати елемент*»; англ. «*Hyperlink Display Text*» – укр. «*Текст гіперпосилання для відображення*» тощо.

Умовно все програмне забезпечення, яке необхідно локалізувати, можна поділити на дві групи: перша група – програмне забезпечення, яке розроблялося з урахуванням багатомовного використання; друга група – програмне забезпечення, яке створювалося для використання тільки однією мовою. Друга група програм більш складна для локалізації та потребує певних підготовчих операцій, крім того, вона може вимагати суттєвого перероблення програмного коду. Відповідно, з огляду на технічний аспект локалізацію виконують у двох варіантах: 1) локалізація безпосередньо ресурсних файлів (наприклад, виконавчі файли із розширенням \*.EXE, \*.DLL, \*.RESX, \*.RESOURCE, \*.XML, а також ресурси створювані Borland Delphi і C++ Builder, проекти Visual Basic); 2) локалізація елементів безпосередньо в коді програми; локалізація тексту в коді програми необхідна, якщо, наприклад, у коді програми задається текст, який потім буде показано користувачеві (це може бути назва клавіш, текст звіту тощо).

До початку власне перекладу вербальної інформації перекладачу слід виконати такі дії: визначити, які елементи користувацького інтерфейсу не винесені в ресурси і, отже, недоступні для локалізації; перевірити, як впливає зміна довжини рядків на функціональність програми; проаналізувати наслідки від використання в ресурсах символів з іншої мови [14].

Для вирішення завдань локалізації програмного забезпечення допоможуть спеціалізовані засоби автоматизованого перекладу, що працюють на основі накопичення пам'яті перекладу, наприклад SDL Passolo Essential 2011 SP6, Alchemy Catalyst, Lingobit Localizer та ін. За допомогою таких програм перекладачі мають змогу перекладати безпосередньо ресурсні файли, використовувати пам'ять перекладів, тестувати здійснений переклад на наявність помилок, створювати глосарії та автоматично перевіряти правопис.

Якщо в накопичених раніше перекладах, що зберігаються у форматі перекладених сегментів (фрагментів тексту мовою оригіналу та відповідних фрагментів цільовою мовою) трапляється фрагмент (речення/вираз), який збігається або достатньою мірою збігається з фрагментом, уже наявним у пам'яті перекладів, то програма шукає його в перекладацькій пам'яті й відображає на екрані, і перекладач може використовувати цей фрагмент у тому самому вигляді або змінити його. Додаткові функції таких програм дають змогу складати словники термінів (глосарії) і використовувати їх під час перекладу для підстановки термінів: програма автоматично знаходить наявні у словнику термінів слова або дає змогу переглянути їх переклад і перенести в текст [4; 5]. Для перекладу аудіовізуальних матеріалів (медіатекстів) використовуються інші спеціалізовані засоби, наприклад **Swift Elite**, у яких об'єднано деякі аспекти пам'яті перекладів, але додатково забезпечено можливість редагування субтитрів (появи за часом, їх форматування на екрані, відповідності відеостандартам тощо). Зазначені програми дають змогу перекладати меню та повідомлення безпосередньо у програмних ресурсах і відкомпільованих програмах, перевіряти коректність локалізації. Це значно полегшує роботу перекладачів і забезпечує високу якість адаптації програмного забезпечення.

Локалізація прикладного програмного забезпечення національною мовою є важливим фактором культурної ідентифікації українського етносу. Локалізація має економічне підґрунтя, адже дає змогу програмним продуктам іншомовного походження, зокрема прикладним, поширюватися ринками в інших країнах, зокрема Україні.

У процесі локалізації виокремлюють два аспекти: власне лінгвістичний та технічний, що постають комплексно. Лінгвістичний аспект полягає у здійсненні перекладу відповідних текстових/графічних/звукових елементів програм. У зв'язку із цим виникають труднощі створення коректних перекладів, що ґрунтуються на адекватній термінології з урахуванням лексико-граматичних, стилістичних та прагматичних вимог і відповідають соціокультурним очікуванням. Технічний аспект передбачає використання спеціальних

комп'ютерних засобів локалізації, зокрема систем автоматизованого перекладу, що полегшують роботу перекладача та зменшують рівень залежності перекладачів від програмістів у процесі їхньої сумісної діяльності в межах локалізації.

### 6.1. SDL Passolo Essential 2022 як засіб для локалізації

**Локалізація** – це адаптація програмного забезпечення таким чином, щоб його могли використовувати користувачі, які розмовляють іншими мовами, відповідно до особливостей, прийнятих у їхніх країнах. Зазвичай локалізується програмний інтерфейс і документація до програмного забезпечення, зокрема може знадобитися зміна одиниць виміру (наприклад, мір, ваги і довжини), а також перероблення програм у процесі їх створення. Процес перекладу – це лише частина локалізації. Особливістю ж процесу локалізації є те, що часто перекладач має тісно взаємодіяти з програмістами. Найбільш поширеними завданнями локалізації є такі: використання символів національних валют, зміна форматів дати і часу, переклад елементів інтерфейсу з урахуванням того, що в різних мовах той самий за змістом текст може відрізнятися довжиною (кількістю символів) тощо.

Умовно все програмне забезпечення, яке необхідно локалізувати, можна розділити на дві групи: 1) програмне забезпечення, яке розроблялося з урахуванням багатомовного використання; 2) програмне забезпечення, яке створювалося для використання тільки однією мовою. Друга група програм більш складна для локалізації та потребує певних підготовчих операцій, зокрема така група може вимагати суттєвого перероблення програмного коду. У вирішенні цього завдання вам може допомогти спеціалізоване програмне забезпечення, призначене для локалізації, наприклад **SDL Passolo Essential 2022 SP6, Alchemy Catalyst, Lingobit Localizer** та ін.

**SDL Passolo** – це спеціалізоване рішення для локалізації програмного забезпечення, підтримує безпосереднє оброблення чималої кількості спеціальних форматів файлів, враховуючи виконувані програми, файли ресурсів і файли на основі **xml**. Це рішення дає змогу перекладати текстові рядки на безліч мов, зокрема азійські та з написанням справа наліво (наприклад, іврит і арабська мова).

**SDL Passolo** – лідер серед засобів підтримки локалізації додатків **Microsoft.net**, які отримують все більше поширення в усьому світі та фактично стають платформою для розроблення мережевого програмного забезпечення, що використовується багатьма компаніями під час створення орієнтованих на клієнта вебдодатків.



Рішення SDL Passolo надзвичайно зручне у використанні і дає змогу оптимізувати процес локалізації програмного забезпечення. Від користувачів не потрібно ні тривалої і дорогої професійної підготовки, ні досвіду у програмуванні. Локалізація програмного забезпечення може виконуватися без доступу до вихідного початкового коду (мови програмування) і без залучення розробників. До того ж процес локалізації може бути початий ще до виходу остаточної версії програмного продукту. SDL Passolo забезпечує просту компіляцію та оброблення даних перекладу, а також обмін цими даними. Функція **Pseudo Translation** дає змогу перевірити додаток на можливість локалізації до початку власне перекладу.

SDL Passolo містить кілька редакторів **wysiwyg**, що дають змогу адаптувати інтерфейс користувача програмного забезпечення. Це редактори діалогових вікон, меню, бітових об'єктів, значків і курсорів. Під час адаптації інтерфейсу користувача виключається ймовірність випадкового видалення або зміни наявних елементів або структур.

Вбудована технологія пам'яті переказів у SDL Passolo дає змогу багаторазово використовувати раніше виконані переклади. Для автоматичного попереднього перекладу в нових проєктах може використовуватися текст навіть із тих програм, які ще не перекладалися за допомогою SDL Passolo. Технологія пошуку неточних збігів дає змогу знаходити як просто схожі фрази, так і тексти, що в точності збігаються. Такий підхід значно підвищує продуктивність перекладача і скорочує цикл перекладу.

Зазвичай локалізація програмного забезпечення передбачає участь безлічі фахівців, що працюють з абсолютно різними інструментами. Завдяки SDL Passolo стає можливим обмін даними між усіма основними системами пам'яті перекладів із підтримкою загальноприйнятих форматів обміну даними.

Функція контролю якості (QA) дає змогу виконувати перевірку правопису в текстах і забезпечує автоматичне розпізнавання усічених текстів, а також помилок у призначенні комбінацій клавіш та інших допоміжних засобів виклику команд. SDL Passolo виключає або автоматично розпізнає безліч потенційних помилок, які можуть виникнути у процесі локалізації програмного забезпечення.

Спеціальні версії SDL Passolo, призначені для особливо великих проєктів або роботи в розподіленому середовищі, дають змогу швидко організувати ефективний обмін даними із зовнішніми перекладачами. SDL Passolo включає в себе **VBA-сумісний механізм створення скриптів** і підтримує автоматизацію **OLE**. Функції SDL Passolo можна значно розширити шляхом безкоштовного завантаження вже готових макросів. Інтегроване середовище **IDE** дає змогу розробляти власні

рішення з локалізації відповідно до вимог кожного конкретного програмного продукту.

Тісна інтеграція **SDL Passolo** і **SDL Trados** дає змогу корпораціям і постачальникам лінгвістичних послуг спільно і більш ефективно використовувати бази перекладів під час перекладу користувацьких інтерфейсів, файлів довідки та документації. Комбінація **SDL Passolo** і **SDL Trados** довела свою стабільність і попит на ринку.

## 6.2. Інші засоби локалізації

**Alchemy Catalyst** – потужний інструмент для локалізації програмного забезпечення. Це візуальне середовище, яке включає всі аспекти керування проєктом локалізації. Програма розроблена ірландською компанією **Alchemy Software Development**, яка є лідером у цій галузі та спеціалізується на розробленні інструментів локалізації. У **Alchemy Catalyst** також використовують технологію **Translation Memory**, яка дає змогу найбільш швидко і ефективно виконувати переклад, а також використовувати отримані матеріали в подальших проєктах.

**Multilizer** – потужна програма для локалізації додатків, працює з безліччю різних форматів файлів, дає змогу перекладати на багато мов. Програма підтримує роботу з декількома мовами програмування, і розпізнавши їх, дає змогу вносити зміни у програмний код. Робота з додатком досить комфортна, і за наявності певних навичок його використання не потребує багато часу.

**Visual Localize** – професійний інструмент для локалізації додатків та інтерфейсів, є одним із лідерів серед програмного забезпечення для локалізації. Програма дає змогу швидко й комфортно виконувати локалізацію, забезпечуючи при цьому ефективність процесу, заощаджуючи час і скорочуючи затрати. Працює з будь-якими мовами, проте не підтримує роботу з програмами, написаними на мові Pascal.

## 7. ОГЛЯД НАЙПОШИРЕНІШИХ СИСТЕМ АВТОМАТИЗОВАНОГО ПЕРЕКЛАДУ

Перелік САТ-програм, вказаний нижче, охоплює тільки деякі з наявного та доступного програмного забезпечення. Перелік не є вичерпним і дає змогу лише оглянути найтипівіші властивості кожної з програм (табл. 7.1).

Таблиця 7.1

## Перелік систем автоматизованого перекладу

Назва	Формати файлів, з якими можна працювати у програмі
Déjà Vu	Microsoft Office (Word, Excel, Powerpoint, also embedded objects, and Access), Help Contents (CNT), FrameMaker (MIF), PageMaker, QuarkXPress, QuickSilver/Interleaf ASCII, Java Properties (.properties), HTML, HTML Help, XML, RC, C/Java/C++, IBM TM/2, Trados Workbench, Trados BIF (old TagEditor), Trados TagEditor, JavaScript, VBScript, ODBC, TMX, EBU, InDesign (TXT, ITD, INX, IDML), GNU GetText (PO/POT), OpenOffice, OpenDocument SDLX (ITD), ResX, XLIFF (XLF, XLIF, XLIFF, MQXLIFF, unsegmented and segmented SDLXLIFF), Visio (VDX), PDF, Transit NXT PPF, WordFast Pro TXML
GlobalSight	Text ANSI/ASCII/Unicode for Windows, Text for Apple Macintosh, HTML, XML (ASP.NET, ASP, JSP, XSL), SGML, MS Word for Windows, MS Excel, MS PowerPoint, RTF, RC, Adobe FrameMaker, Adobe InDesign
Lokalize	Gettext PO, Qt ts, XLIFF, TMX
memoQ	.MIF, InDesign formats (.INDD, .INX, .IDML), .XML, .DITA, .XML, .MM, .PO, .HTML, .HMT, .SHT, .properties, .DOC, .RTF, .BAK, .DOT, .DOCX, .XLS, .XML, .XLSX, .XLSM, .XLS, .XLT, .PPT, .PPS, .POT, .PPTX, .PPSX, .POTX, .SLDX, .VDX, .HHC, .HHK, .ODT, .ODF, .TXT, .INF, .INI, .REG, .PDF, .SVG, .SDLPPX, .TTX, .SDLXLIFF, .TMX, .TXML, .RESX, .XLF, .XLIF, .XLIFF, XLIFF:doc
MemSource	.doc, .docx, .rtf, .ppt, .pptx, .xls, .xlsx, .htm, .html, .properties, .tmx, .ttx, .xhtm, .xhtml, .xml, .dita, .ditamap, .xliff, .sdlxliff, .txt, .csv, .resx, .idml, .mif
MetaTaxis	Microsoft Word, Excel and Powerpoint, all kinds of text formats, XML, HTML, XLIFF, RTF, TRADOS Studio (SDLXLIFF), TagEditor (TTX), POT/PO, Manual Maker, several further formats
OmegaT	Plain text, HTML, XHTML, StarOffice, OpenOffice.org, OpenDocument (ODF), MS Office Open XML, Help & Manual, HTML Help Compiler (HCC), LaTeX, DokuWiki, QuarkXPress CopyFlow Gold, DocBook, Android Resource, Java

Закінчення табл. 7.1

Назва	Формати файлів, з якими можна працювати у програмі
	Properties, Typo3 LocManager, Mozilla DTD, Windows RC, WiX, ResX, INI files, XLIFF, PO, SubRip Subtitles, SVG Images
Open Language Tools	XLIFF, HTML/XHTML, XML, DocBook SGML, ASCII, StarOffice/OpenOffice/ODF, PO, .properties, .java (ResourceBundle), .msg/.tmsg (catgets)
SDL Trados	Features four translation environments: dedicated TagEditor, MSWord Interface, SDLX, the integrated interface SDL TRADOS STUDIO 2022. Additional filters for translating with TagEditor available: Word, Excel, PowerPoint, OpenOffice, InDesign, QuarkXPress, PageMaker, Interleaf, Framemaker, HTML, SGML, XML, SVG, Includes SDL MultiTerm for terminology management and Project Management Dashboard for automating tasks and tracking.
Wordfast PRO	MS Word, Excel, PowerPoint (all versions), PDF, SGML, HTML, XML, InDesign, FrameMaker, tagged documents

Системи автоматизованого перекладу (CAT) – це програми, що у процесі перекладу зберігають у пам'яті початкові та відповідні перекладені сегменти тексту й дають змогу повторно використовувати збережені сегменти під час перекладу схожих фрагментів тексту. Переклад текстів і збереження початкових і відповідних перекладених сегментів у пам'яті перекладу становить основу для подальших подібних перекладів.

Отже, системи автоматизованого перекладу працюють за принципом лінгвістичної пам'яті, пропонуючи варіанти збереженого перекладу. Не слід плутати системи автоматизованого перекладу з системами машинного перекладу, що автоматично генерують переклад тексту за допомогою комп'ютерних програм. Особливо рекомендується застосування систем автоматизованого перекладу під час перекладу часто оновлюваних або схожих текстів, що містять повторювані фрагменти. До таких текстів належать: технічні довідники та каталоги продукції; інструкції з експлуатації, вебсайти, програмне забезпечення та довідкові файли; документація для різних варіантів продукту одного і того ж виробника або для схожих варіантів продуктів різних

виробників; документація, складена з використанням правил контрольованої мови.

Застосування систем автоматизованого перекладу сприяє використанню термінології та фраз-кліше, прийнятих у певній компанії, у всіх перекладних документах.

Щоб працювати у програмі, яку вимагає замовник, достатньо розібратися із принципами роботи однієї із перелічених програм (наприклад, SLD TRADOS STUDIO 2022). Інші програми діють за тими ж принципами та відрізняються лише деякими особливостями, наприклад інтерфейсами та додатковими функціями.

## 8. КОРПУСНА ЛІНГВІСТИКА

У сучасних умовах постійного зростання інформаційних потоків проблема автоматичного опрацювання природномовної інформації набуває все більшого значення. Комп'ютерні технології, впроваджені в лінгвістику, запропонували свої технічні можливості для опрацювання, збереження та відбирання мовного матеріалу. Це спричинило зародження та розвиток нової галузі мовознавчих досліджень – корпусної лінгвістики. Базуючись на використанні чималої кількості емпіричного матеріалу та залученні комп'ютерних технологій для отримання об'єктивних висновків про функціонування мовних одиниць, лінгвістичні корпуси швидко перетворилися на неоціненний лінгвістичний ресурс. Отримані в результаті корпусного аналізу дані сприяють формулюванню якісно нових висновків про мову та окреслюють такі напрями лінгвістичних досліджень, які до появи корпусів не привертали уваги дослідників. Спираючись передусім на реальний «живий» мовний матеріал, а не на мовну інтуїцію та інтроспекцію, корпусні дослідження дають змогу абстрагуватися від суб'єктивності дослідника, наблизитися до об'єктивного вивчення мови.

**Корпусна лінгвістика** – розділ комп'ютерної лінгвістики, що займається розробленням загальних принципів побудови та використання лінгвістичних корпусів (корпусів текстів) із застосуванням комп'ютерних технологій [2, с. 63]. Під лінгвістичним, або мовним, корпусом текстів розуміється великий, поданий у форматі для розпізнавання комп'ютером, уніфікований, структурований, розмічений, філологічно компетентний масив мовних даних, призначений для вирішення конкретних лінгвістичних завдань (рис. 8.1).

До поняття «корпус текстів» входить також система керування текстовими та лінгвістичними даними, яку останнім часом найчастіше називають корпусним менеджером.

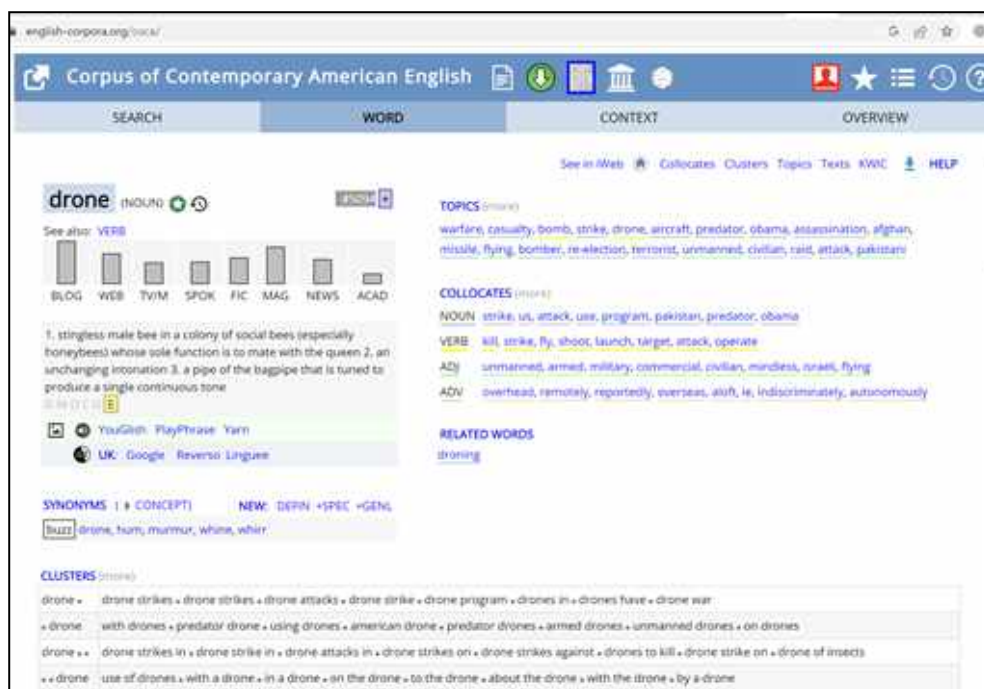


Рис. 8.1. Сучасний корпус американського варіанта англійської мови

Пошук у корпусі дає змогу за будь-яким словом побудувати **конкорданс** – список усіх вживань певного слова в контексті з посиланнями на джерело.

Використання корпусів стає дедалі популярнішим у навчанні та вивченні мови. Викладачі та студенти все частіше використовують корпус для здійснення швидкого різноманітного пошуку та аналізу найрізноманітнішої лінгвальної інформації, а конкорданс визнається зручним інструментом для ефективного опанування мови. Корпуси текстів є неоціненним ресурсом реальних прикладів вживання тих чи інших лексичних і граматичних одиниць, за умови, що користувачі мають доступ до корпусів та обізнані з технікою роботи з корпусом. Навчання за допомогою баз даних (data-driven learning) іноземній мові, коли корпус стає головним джерелом і способом отримання лінгвістичних знань й умінь, має все більше прихильників серед сучасних студентів і вчителів. За цього підходу базоване на корпусі програмне забезпечення використовують для підтримки інтерактивної навчальної діяльності, а залучення «живого» мовлення з корпусу сприяє пробудженню у студентів інтересу до процесу вивчення мови та

заохоченню їх до самостійного вивчення автентичного мовного матеріалу. Студент виконує роль дослідника мови, а викладач лише спрямовує здобувача та сприяє його навчанню.

Приклади популярних корпусів текстів:

1. Корпус сучасної американської англійської мови містить тексти, що являють собою усне мовлення, художню прозу, популярні журнали, газети та наукову літературу обсягом 425 млн слів.

2. Корпус DeReKo (das Deutsche Referenz Korpus) складається з белетристики, наукових і публіцистичних текстів німецької мови і містить понад 4,1 млрд слововживань.

3. Британський національний корпус містить понад 100 млн слів розмовної та письмової англійської мови. Підкорпус письмової мови становить 90 % всього корпусу і включає в себе газети, періодичні наукові видання й журнали для різних вікових груп, популярну наукову фантастику, опубліковані та неопубліковані листи, шкільні та університетські твори тощо. Підкорпус усного мовлення включає в себе множини контекстів: від мови формальних ділових або урядових зустрічей до радіошоу й телефонних розмов.

4. Чеський національний корпус – синхронічний морфологічно розмічений корпус, який представляє сучасну чеську мову. Масив текстів, зібраних у корпусі, розділений на синхронічну (SYNKORP) і діяхронічну частини (DIAKORP). **Синхронічна частина** містить такі частини: письмові тексти (понад 100 млн слововживань), розмовні (750 тис. слововживань) і діалектну мову. Обсяг **діяхронічної частини** – 1 750 тис. слововживань. Ця частина включає тексти із 14 століття. Сьогодні лінгвісти-дослідники зосереджуються на текстах 19 століття. Довгострокова мета DIAKORP полягає в тому, щоб створити корпус, що охоплює період із 1850 року дотепер, і поєднати дані із серією текстів у SYNKORP.

5. Національний корпус української мови – вибірка текстів сучасної української мови, обсягом близько 50 млн слововживань. Корпус містить прозові, поетичні та драматургічні тексти відповідних хронологічних меж і метатекстову розмітку.

Отже, великі національні корпуси дають змогу дослідникам здійснювати автоматичний пошук і систематизацію емпіричного матеріалу, швидко обробляти чималу кількість мовних даних за допомогою спеціальних комп'ютерних програм.

Корпусна лінгвістика являє собою одну з галузей, що стрімко розвивається. Отже, прогрес у сфері комп'ютерних технологій

передбачає прогрес у створенні і вдосконаленні програм автоматичного оброблення тексту.

### **Домашнє завдання**

1. Зареєструйтеся на сайті Oxford English Corpus за посиланням:  
<https://www.english-corpora.org/>
2. Ознайомтеся з основними характеристиками корпусу.

The following is a short tour of the COCA corpus, including new features in March 2020. You can click on any of the links below to carry out sample searches, and then return to this page (for more searches) by clicking on TOUR at the top of the page.

### **TEXTS**

The COCA corpus contains about 1 billion words in nearly 500,000 texts from 1990 to 2019 – which are nearly evenly divided between spoken, fiction, magazines, newspapers, academic journals, blogs, other web pages, and TV/Movie subtitles (120-130 million words in each genre). In addition, there are 20 million words each year from 1990-2019 (with the same genre balance each year), which means that you have very recent data – not just stale data from 25-30 years ago. This is the only corpus of English that is 1) large 2) recent and 3) has a wide range of genres.

### **SEARCHES**

As with the other BYU corpora, you can do many kinds of searches, such as single words (mysterious, skew), phrases (make up, on the other hand), «fuzzy phrases» (more \* than, \* bit), wildcards (\*icity, \*break\*), lemma/forms (SKIM, BUY \* NOUN), alternants and NOT (fast|slow rate), pretty -NOUN), part of speech (rough NOUN, VERB money), synonyms (=beautiful, =gorgeous NOUN, =clever =man), customized word lists (@clothes, @colors @clothes), and any combination of these (GET him|her to VERB, PUT on POSS @clothes). Click on the box next to SECTIONS in the search form to see the results from each genre and time period, e.g. rough NOUN or =gorgeous NOUN). You can also compare words, such as nouns that appear with *small* and *little*, adjectives with *men* and *women*, or nouns with *utter* and *sheer*.

Note that because the corpus has an architecture that is designed for speed (unlike other large 10-20 billion word corpora), even searches like the following only take a few seconds: got VERB-ed, I VERB PRON VERB, phrasal verbs, or NOUN NOUN.



## GENRES

Because the corpus has so much data from each of the eight genres listed above, you can do powerful searches to compare among these genres. This lets you know, for example, whether a particular word or phrase or grammatical construction is more informal or formal. For example, look for the frequency of *I guess*, *muffled*, *validity*, or forms of *need + to + VERB*. Via the chart display, you can also see the frequency of the word or phrase in subregisters as well, such as movie scripts, children's fiction, women's magazines, or medical journals.

You can also limit searches to particular genres, such as \*ize verbs in ACADEMIC, past tense verb + *up* in TV/MOVIES; nouns near *chair* in FIC; or adjectives in ACADEMIC-Medicine (click on SECTIONS in the search form to see the selected genres, if they're not already visible). Better yet, you can easily and quickly compare all of the results in one genre (or sub-genre) with another. For example past tense verb + *up* in TV/M vs ACAD; hard + NOUN in MAG vs ACAD; nouns in NEWS-Money vs NEWS; adjectives in ACAD-Medicine vs ACAD; synonyms of *strong* in TV/M vs ACAD; or nouns near *chair* in ACAD vs FIC, or ADJ + WOMAN in FIC vs ACAD.

## CHANGES OVER TIME

The COCA corpus is the only large corpus of English that contains data (20 million words of data, with the same genre balance) in each year from 1990-2019. This allows you to see the frequency of any word or phrase over time, such as *gift* (as a verb), *awesome*, or BE likely a|the. You can also compare *all* words in different periods, such as -ed verbs, the suffix *-friendly*, or words ending in *-ism*. You can also see the frequency of each matching form in each five year period, such as *-ism* words since the early 1990s. You can also look at syntactic constructions over time, such as end up V-ing, *get* passive (*got hired*; an increase in each five year period), *so not* ADJ (*I'm so not interested in her*), and «quotative *like*» (*he's like, I'm not going*). You can compare collocates over time so see changes in meaning: *green*, *web*, *engine*, and you can see the frequency of collocates in each five year period since the early 1990s (*nuclear*, *crisis*) to examine changes in American culture and society during this time (again, just click on the box by SECTIONS to see the frequency by section).

## BROWSE/WORDS

Finally – but perhaps most importantly – you can **browse through the top 60,000 words** in the corpus, and to see detailed information on any of these words. You can search by word form (e.g. *\*ism*, *un\*able*, *\*break\**, *fl?p*), rank order (e.g. words near

the 8,000th, 28,000th or 48,000th most frequent word), pronunciation, e.g. words rhyming with *light* (note the many variant spellings for this sound), or three syllable words accented on the last syllable, or any combination of these (e.g. two syllable words that rhyme with *stay* where the stress is on the second syllable, mid-frequency two syllable adjectives ending in \*y, or low frequency five syllable \*ism words). You can also search for a specific word via WORD in the search form.

When you see a word of interest, you can get a wealth of information on that word, including frequency information, definition, translation, link to audio, images, and videos, as well as synonyms, related word forms, topics (words that co-occur anywhere in the ~500,000 texts), collocates, clusters, and concordance / KWIC lines. Click on any of the links below for examples of this "word-based" information for a few sample words.

As you click on any of these links, note how you can easily and quickly move from one type of search to another (via the links at the top of those pages), and how you can easily move from one word to another. You can also see a history of all of these word-centered searches, and you can add words to your "favorites" list, for later review. Note of this functionality is available from any other large corpus of English.

3. Зайдіть на вкладку **BROWSE/WORDS** та проаналізуйте слова «aircraft», «applied linguistics», «human» з огляду на стилістику та лексикологію.

## БІБЛІОГРАФІЧНИЙ СПИСОК

1. Ільченко, М. Л. Сучасні інформаційні технології у перекладацькій діяльності: навч. посіб. для студ. гуманіт. ф-т. вищ. навч. закладів / М. Л. Ільченко. – Харків : Нац. аерокосм. ун-т «Харків. авіац. ін-т», 2015. – 52 с.
2. Петрасова, С. В. Сучасні інформаційні технології в лінгвістиці : навч. посіб. / С. В. Петрасова, Н. Ф. Хайрова. – Харків : ФОП Панов А. М., 2020. – 124 с.
3. Тарасенко, Р. О. Інформаційні технології в перекладацькій діяльності: навч. посіб. / Р. О. Тарасенко, С. М. Амеліна. – Київ : ЦП «КОМПРИНТ», 2017. – 253 с.
4. Жуковська, В. В. Вступ до корпусної лінгвістики: навч. посіб. / В. В. Жуковська. – Житомир : Вид-во ЖДУ ім. І. Франка, 2013. – 140 с.
5. Основи інформатики та прикладної лінгвістики : консп. лекцій для здобувачів вищ. освіти першого (бакалаврського) рівня зі спец. 035 «Філологія» / укладач Н. М. Волосова. – Кам'янське: ДДТУ, 2018. – 84 с.
6. Тарануха, В. Ю. Інтелектуальна обробка текстів: навч. посіб. / В. Ю. Тарануха. – Київ : КНУ ім. Т. Шевченка, 2014. – 80 с.

Навчальне видання

**Гелетка Маргарита Леонідівна**

## **ЦИФРОВІ ТЕХНОЛОГІЇ В ПЕРЕКЛАДАЦЬКІЙ ДІЯЛЬНОСТІ**

Редактор А. Г. Литвин

Зв. план, 2023

Підписано до видання 31.08.2023

Ум. друк. арк. 3,7. Обл.-вид. арк. 4,19. Електронний ресурс

---

Видавець і виготовлювач

Національний аерокосмічний університет ім. М. Є. Жуковського

«Харківський авіаційний інститут»

61070, Харків-70, вул. Чкалова, 17

[http:// www.khai.edu](http://www.khai.edu)

Видавничий центр «ХАІ»

61070, Харків-70, вул. Чкалова, 17

[izdat@khai.edu](mailto:izdat@khai.edu)

Свідоцтво про внесення суб'єкта видавничої справи  
до Державного реєстру видавців, виготовлювачів і розповсюджувачів  
видавничої продукції сер. ДК № 391 від 30.03.2001