

УДК 81'246 + 81'33 + 044.42] : 001(4) (045)  
DOI: <https://doi.org/10.32589/2311-0821.1.2023.286184>

**А. В. Корольова**  
Київський національний лінгвістичний університет, Україна  
e-mail: [alla.korolyova@knlu.edu.ua](mailto:alla.korolyova@knlu.edu.ua)  
ORCID ID: <https://orcid.org/0000-0001-5541-5914>

## МУЛЬТИЛІНГВАЛЬНИЙ КОРПУС І ЙОГО ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ ДЛЯ ДОСЛІДЖЕННЯ ЄВРОПЕЇСТИКИ

### Abstract

The paper proposes the method of working with the functionality of the computer programs AntConc, WordSmith, WordList, MonoConc Pro, CATMA, and KORP, which can be used to study the multilingual corpus texts on the topic of European Studies. Various debatable views of foreign scientists – representatives of Corpus Linguistics – regarding the content of the concept of the multilingual corpus have been considered. Besides, there has been formulated a working definition of the multilingual comparative corpus that belongs to the combined type of thematically oriented corpora of texts in different languages, grouped into sub-corpora, with their translation into other languages (or with the possibility to use computer programs to translate texts). The paper also defines the role of statistics in Corpus Linguistics, which deals with the analyzed computer programs and allows to calculate the frequency of words or collocations use, construct diagrams of the frequency of word or collocation use in sub-corpora of texts, etc. thus there has been made a preliminary conclusion that the computer toolkit of the corpus managers AntConc, WordSmith, WordList, MonoConc Pro, CATMA and KORP makes it possible to construct both individual Key Word in Context (KWIC) and concordance lists of search elements on the subject of European Studies; distinguish between the functionality of one or another contextual meaning of search units in their most probable left-handed and right-handed valency in different languages; see the results of statistical processing of the information from corpus tags; save and print the results; support different formats of text data (txt, doc, rtf, html, etc.).

**Keywords:** multilingual corpus, combined type of text corpus, European Studies, computer programs, corpus managers.

### Анотація

У статті запропонована методика роботи із функціоналом комп'ютерних програм AntConc, WordSmith, WordList, MonoConc Pro, CATMA й KORP, які можна використовувати для дослідження текстів мультилінгвального корпусу за тематикою Європеїстики. Розглянуто різні дискусійні погляди зарубіжних учених – представників корпусної лінгвістики – щодо змісту поняття мультилінгвальний корпус. Сформульовано робоче визначення мультилінгвального зіставного корпусу, який належить до комбінованого типу тематично орієнтованих різномовних корпусів текстів, об'єднаних у підкорпуси, з їхнім перекладом іншими мовами (або з можливістю застосування комп'ютерних програм для виконання перекладу текстів). Визначено роль корпусної лінгвостатистики, якою оснащені проаналізовані комп'ютерні програми і яка дозволяє здійснювати обчислення частоти вживання слів чи колокацій, будувати діаграми частоти вживання слова / колокації в підкорпусах текстів тощо. Зроблено попередній висновок про те, що комп'ютерний інструментарій корпус-менеджерів AntConc, WordSmith, WordList, MonoConc Pro, CATMA й KORP надає змогу конструювати як окремі Key Word in Context (KWIC), так і конкордансні списки пошукових елементів за тематикою Європеїстики; розмежовувати відтинки того чи іншого контекстуального значення пошукових одиниць в їхній найбільш імовірній лівобічній та правобічній валентності в різних мовах; побачити результати статистичної обробки інформації з корпусних тегів; зберігати та роздруковувати результати; підтримувати різні формати текстових даних (txt, doc, rtf, html й ін.).

**Ключові слова:** мультилінгвальний корпус, комбінований тип корпусу текстів, Європеїстика, комп'ютерні програми, корпус-менеджери.

**Вступ.** Сучасні дослідження в галузі корпусної лінгвістики (О. Андрушенко, В. В. Жуковська, І. В. Мейзерська, В. А. Широков, А. М. McEnery, A. Kilgariff, S. P. Rauf, P. Resnik & N. A. Smith та ін.) визначають корпус текстів не суто як продукт технологічного прогресу чи зручний інструмент для пошуку необхідної інформації, а насамперед як нову ідеологію у вивченні мов, що орієнтує дослідника на текст чи ширше – дискурс, які становлять новий об'єкт теоретико-лінгвістичної рефлексії.

Робота з корпусом текстів допомогла відкрити можливості не лише для більш швидкого й ефективного розв'язання актуальних на сьогодні лінгвістичних завдань, а й для реалізації більш амбітних цілей, принципово нових, які раніше були надскладними для їхньої практичної реалізації через великий обсяг інформації, що потребувала селекції й подальшого оброблення. Одним із таких завдань є необхідність узагальнення наявних результатів дослідження мікроеволюції кожної мови протягом одного-двох століть, які пов'язані: 1) з малопомітними, та все ж таки змінами в правилах поєднання слів і комбінацій їхніх значень; 2) зі змінами частоти вживання лексичних і граматичних варіантів різних конструкцій; 3) із фіксацією абсолютно нових явищ мови та практичного зникнення тих, що тривалий час уважалися традиційними.

Оптимальне розв'язання сформульованої проблеми стає можливим у разі залучення ресурсів (значних за обсягом колекцій текстів) мультилінгвальних / багатомовних корпусів та наявного комп'ютерного інструментарію, що дозволить технологічно забезпечити виконання вище окреслених масштабних завдань, зокрема й діахронічних (Андрушенко, 2022; Жуковська, 2013). Для доведення цієї гіпотези розглянемо функціонал комп'ютерного програмного забезпечення і застосуємо його інструменти для опрацювання текстів мультилінгвального корпусу з такої тематичної сфери знання, як Європейістика.

Вибір Європейістики (англ. European studies), з одного боку, як галузі наукового знання, зосередженої на аналітиці процесів європейської інтеграції, а з іншого, як ресурсу наукової та медійної інформації, присвяченої висвітленню питань діяльності всіх європейських інституцій (Європейського парламенту, Європейської ради, Ради Європейського Союзу, Європейської комісії, Суду Європейського Союзу, Європейської Рахункової палати, Європейського центрального банку та ін.), зумовлений тим фактором, що обрана сфера (одна з небагатьох) сама по собі вже містить тематичну розмітку у вигляді так званих тематичних підкорпусів різномовних текстів.

**Аналіз останніх досліджень і публікацій.** Попередній огляд наукових праць із цього питання дає підстави припустити, що визначення дистинктивних критеріїв укладання оптимальної типології мультилінгвальних корпусів текстів перебуває в стадії розроблення.

Одним із типів мультилінгвальних корпусів є зіставні корпуси (англ. comparable corpora), створенням та вивченням яких займаються здебільшого закордонні вчені (А. McEnery, Z. Xiao, A. Kilgariff), а в Україні така практика лише починається. Основне питання, яке обговорюється в зарубіжних дослідженнях, пов'язане з визначенням принципів побудови такого типу корпусів, а також сфер їхнього призначення та використання.

У корпусно орієнтованих студіях представлені полярні погляди щодо інтерпретації поняття “зіставний корпус”, який розглядають або 1) як мономовний, або 2) як двомовний / мультилінгвальний, або 3) як паралельний.

1. До зіставних мультилінгвальних корпусів текстів (L. Bowker, J. Pearson) відносять “декілька аналогічних за будовою (структурою) підкорпусів текстів окремої тематики, які не містять текстів перекладу” (Bowker, Pearson, 2002, р. 93). Інші автори (Barzilay, Lee, 2003; Elhadad, Sutaria, 2007) цілком згодні зі своїми колегами в тому, що “зіставний корпус повинен містити текстові документи, обмежені рамками фахової мови з конкретної предметної сфери й призначені для використання членами певної соціальної групи” (Bowker, Pearson, 2002, р. 12).

Натомість із-поміж представників цього напрямку не все так однозначно. Деякі фахівці (F. Zanettin, O. Culo, S. H. Schirra, S. Neumann, M. Vela та ін.) висловлюють думку про те, що один із підкорпусів має включати письмові тексти певною мовою, а інший – тексти, перекладені цією мовою, оскільки це необхідно для вивчення перекладацьких трансформацій. Тим часом інша група дослідників (B. Cartoni, L. Deléger) переконують у тому, що подібний тип корпусу повинен містити підкорпуси текстів однією мовою (англійською, французькою, польською тощо), але які презентують різні функціональні типи мовлення й різну тематику (Cartoni, Deléger, 2011).

2. До двомовних або мультилінгвальних зіставних корпусів (H. Afli, L. Barrault, H. Schwenk) відносять тексти близької тематики двома й більше мовами, метою яких є винятково зіставлення мов, а не власне переклад. Виходячи з такого розуміння, “зіставний корпус визначають як колекцію текстів аналогічної тематики різними мовами, що не містить їх переклади” (Afli, Barrault & Schwenk, 2012, р. 448).

Проте тут також панує і протилежна думка (A. M. McEnergy та ін.), згідно з якою обстоюється можливість включення паралельних текстів до зіставного мультилінгвального корпусу. Зокрема С. Роф не заперечує, щоб у корпус такого типу були додані різномовні тексти, окремі фрагменти яких мали б переклади іншими мовами (Rauf, 2012).

3. Третій підхід (F. Scarpa) спостерігається в дослідженнях, які значно розширюють уявлення про межі мультилінгвальних зіставних корпусів і відповідно пропонують власні принципи для їхньої побудови. Представники цього підходу переконані, що такі корпуси повинні включати як паралельні (оригінали мовою А та їхні переклади мовою В), так і неперекладені різномовні тексти близької тематики (Scarpa, 2010, р. 124). Знову ж таки з приводу такої позиції дискусія до кінця не вичерпана. Наприклад, М. Гідер категорично не згоден включати до такого типу корпусів тексти без перекладу певною мовою. Він вважає, що поняття зіставного корпусу поширюється лише на вирівняні (перекладені) корпуси текстів (Guidère, 2010, р. 95).

І хоча дискусія щодо завершення укладання типології зіставних корпусів текстів ще триває, проте попередньо мультилінгвальні корпуси можна об’єднати в три типи:

1) одномовні корпуси, в яких представлені тексти оригіналу для спеціальних цілей та їхні тлумачення цією ж мовою (наприклад, підкорпус текстів англійською мовою, підкорпус текстів французькою мовою тощо);

2) двомовні чи мультилінгвальні корпуси, в яких представлені різномовні тексти оригіналів близької тематики для зіставного аналізу;

3) мультилінгвальні корпуси, в яких представлені тексти оригіналу та їхній переклад, що отримали назву паралельних корпусів.

Виходячи з представленої типології, можна припустити, що до різновидів мультилінгвальних корпусів належать:

– так звані одномовні корпуси з підкорпусами текстів для спеціальних цілей конкретними мовами;

– зіставні мультилінгвальні корпуси, що містять текстові масиви близької тематики двома і більше мовами, що належать до одного функціонального стилю, дискурсу тощо;

– паралельні (вирівняні) корпуси, що включають оригінальні тексти та їхні переклади різними мовами і створюються для вивчення різних аспектів перекладу.

У корпусних студіях пропонують виділити ще один тип зіставних мультилінгвальних корпусів – так звані *комбіновані мультилінгвальні корпуси текстів*. На нашу думку, комбінований мультилінгвальний корпус повинен містити колекцію підкорпусів різномовних текстів тієї самої тематичної сфери та їхні переклади. У цьому випадку такою тематичною сферою є Європеїстика.

Зважаючи на сказане, можемо сформулювати робоче визначення мультилінгвального зіставного корпусу текстів як комбінованого типу тематично орієнтованого корпусу, що складається з оригінальних різномовних текстів близької тематики, об'єднаних у підкорпуси, з їхнім перекладом іншою або іншими мовами (або ж із можливістю застосування комп'ютерного програмного забезпечення для перекладу текстів).

**Мета статті** – теоретично обґрунтувати і практично продемонструвати ефективність використання в мультилінгвальному корпусі комп'ютерного програмного забезпечення для дослідження Європеїстики.

**Методологія проведення дослідження з викладом та обговоренням основних наукових результатів.** Перш ніж розглядати потенціал програмного забезпечення для його використання в мультилінгвальному корпусі, слід детальніше розкрити характеристики цього типу корпусу.

Оскільки запропоноване вище робоче визначення мультилінгвального корпусу загалом відповідає критеріям як зіставного, так і паралельного корпусів, то вочевидь слід зупинитися на характеристиках комбінованого мультилінгвального корпусу текстів, ресурси якого представляють колекцію текстів із тематикою Європеїстика.

Мультилінгвальний тематично орієнтований корпус комбінованого типу має кілька унікальних властивостей, що відрізняють його від інших типів зіставних корпусів. До таких властивостей варто віднести: а) багатомовність; б) паралельність; в) тематична об'єднаність; г) письмовий вид сучасних текстів (медійних та наукових).

Головні характеристики корпусу, до яких належать багатомовність і паралельність, передбачають наявність у його ресурсній базі оригінальних текстів різними мовами з їхніми відповідними перекладами іншими мовами (або з можливістю їхнього машинного перекладу). Наступні характеристики – це тематична медійно-дискурсивна та наукова спрямованість, пов'язана зі сферою Європеїстики, а також письмові види текстів. У корпус такого типу можуть бути включені як повноформатні тексти, так і їхні фрагменти.

Відповідно до завдань і специфіки будь-якого корпусу, у цьому випадку і комбінованого мультилінгвального корпусу, для їхньої реалізації обирається як наявне програмне забезпечення, так і розробляються цільові корпусні програми та їхній індивідуальний інструментарій.

Наразі найвідомішими і найпопулярнішими є комп'ютерні програми AntConc, WordSmith, WordList, MonoConc Pro, CATMA й KORP, які термінологічно називають ще корпусними менеджерами. Деякі з цих програм можуть виконувати функції

корпусних інструментів, наприклад, WordList. Крім цих програм, використовують інструменти й інших програмних продуктів, розроблених для виконання різних етапів корпусного аналізу.

Початковий етап методики роботи з корпусом передбачає аналіз корпусних менеджерів (програм) та принципів побудови конкордансів, що надають різні можливості для отримання необхідної інформації з корпусу.

Корпусний менеджер – це спеціальний пошуковий інструмент, який використовує програмні засоби для пошуку даних у корпусі, отримання статистичної інформації та надання результатів користувачеві у вигляді горизонтальних рядків із пошуковим елементом посередині.

Для опрацювання корпусів першого порядку використовують зазвичай корпус-менеджер (або комп'ютерну програму) AntConc (<http://www.antlab.sci.waseda.ac.jp/software.html>), яка є найбільш доступною комп'ютерною програмою та найбільш затребуваною користувачами. Вона підтримується на будь-якому комп'ютері, оснащеному операційною системою Microsoft Windows.

AntConc містить низку інструментів, до яких можна отримати доступ, натиснувши клавішу табуляції в меню інструментів або використовуючи функціональні клавіші F1-F7. Це – KWIC; Plot (відображає наявність пошукових елементів у тексті у вигляді штрих-коду і надає можливість візуально оцінити, як часто і в якій частині тексту трапляється потрібний елемент); File View; Clusters; N-Grams; Collocate; Keyword; Word; Wordcloud. Методологію роботи з програмою AntConc для дослідження ключових слів із семантикою резилентності розроблено у праці Я. В. Капранова (Karpunov, 2022, pp. 22-32).

Для дослідження Європеїстики можна використовувати найбільш ефективний інструмент цієї програми – Concordance (Johns, 1991), що представляє результати пошуку потрібного елемента з його лівобічними та правобічними валентними зв'язками шляхом оброблення великих масивів текстів. Після відповідної команди Concordance представляє задану користувачем кількість фрагментів текстів, що містять пошукові одиниці. На основі отриманих даних можна зробити висновок про контекстне вживання заданої одиниці у конкретному дискурсивному просторі.

Наприклад, до колекції текстів мультилінгвального корпусу був завантажений текст англійською мовою “Consolidated Version of the Treaty on the Functioning of the European Union” для пошуку в ньому інформації, пов'язаної із завданнями і функціями Європейського Парламенту (European Parliament) як законодавчого органу ЄС. Для пошуку в цьому тексті одиниці European Parliament був задіяний інструмент Concordance програми AntConc, що дав змогу побачити її контекстне вживання – лівобічний контекст, з якого був зрозумілий зміст цілого тексту, де розкривалися завдання Європейського Парламенту та його повноваження як органу ЄС (Рис. 1).

Генерування рядків конкордансу, або його лінгвістична наочність, досягається за допомогою баз даних (Data-driven approach) і представляє фрагменти текстів корпусу з пошуковою одиницею (O'Keeffe et al., 2007).

Прийом вертикального прочитання рядків конкордансу надає змогу скласти тематичний профіль пошукової одиниці в заданому тексті чи всьому підкорпусі або корпусі.



AntConc  
File Edit Settings Help

Target Corpus  
Name: temp  
Files: 1  
Tokens: 105663

KWIC Plot File View Cluster N-Gram Collocate Word Keyword Wordcloud  
Total Hits: 389 Page Size: 100 hits 1 to 100 of 389 hits

	File	Left Context	Hit	Right Context
1	CELEX_12012E...	financial conditions, which enable them to fulfil their missions. The	European Parliament	and the Council, acting by means of regulations
2	CELEX_12012E...	right of access to documents shall be determined by the	European Parliament	and the Council, by means of regulations,
3	CELEX_12012E...	→ this paragraph only when exercising their administrative tasks. The	European Parliament	and the Council shall ensure publication of the
4	CELEX_12012E...	personal data concerning them. Article 16 (ex Article 286 TEC) 2. The	European Parliament	and the Council, acting in accordance with the
5	CELEX_12012E...	ny discrimination on grounds of nationality shall be prohibited. The	European Parliament	and the Council, acting in accordance with the
6	CELEX_12012E...	or sexual orientation. 2. By way of derogation from paragraph 1, the	European Parliament	and the Council, acting in accordance with the
7	CELEX_12012E...	and the Treaties have not provided the necessary powers, the	European Parliament	and the Council, acting in accordance with the
8	CELEX_12012E...	essary to facilitate such protection. Article 24 (ex Article 21 TEC) The	European Parliament	and the Council, acting by means of regulations
9	CELEX_12012E...	TEC) Within the scope of application of the Treaties, the	European Parliament	and the Council, acting in accordance with the
10	CELEX_12012E...	in agricultural products only to the extent determined by the	European Parliament	and the Council within the framework of Article 43(2)
11	CELEX_12012E...	pendence of the agricultural matters mentioned in this Title. 2. The	European Parliament	and the Council, acting in accordance with the
12	CELEX_12012E...	employment in the public service. Article 46 (ex Article 40 TEC) The	European Parliament	and the Council shall, acting in accordance with
13	CELEX_12012E...	the exchange of young workers. Article 48 (ex Article 42 TEC) The	European Parliament	and the Council shall, acting in accordance with
14	CELEX_12012E...	attain freedom of establishment as regards a particular activity, the	European Parliament	and the Council, acting in accordance with the
15	CELEX_12012E...	26.10.2012 EN Official Journal of the European Union C 326/69 The	European Parliament	and the Council, acting in accordance with the
16	CELEX_12012E...	grounds of public policy, public security or public health. The	European Parliament	and the Council shall, acting in accordance with
17	CELEX_12012E...	take up and pursue activities as self-employed persons, the	European Parliament	and the Council shall, acting in accordance with
18	CELEX_12012E...	of the person for whom the services are intended. The	European Parliament	and the Council, acting in accordance with the
19	CELEX_12012E...	without prejudice to the other Chapters of the Treaties, the	European Parliament	and the Council, acting in accordance with the
20	CELEX_12012E...	ards preventing and combating terrorism and related activities, the	European Parliament	and the Council, acting by means of regulations
21	CELEX_12012E...	system for external borders. For the purposes of paragraph 1, the	European Parliament	and the Council, acting in 2. accordance with the
22	CELEX_12012E...	and other relevant treaties. For the purposes of paragraph 1, the	European Parliament	and the Council, acting in 2. accordance with the
23	CELEX_12012E...	trafficking in human beings. For the purposes of paragraph 1, the	European Parliament	and the Council, acting in 2. accordance with the
24	CELEX_12012E...	326/78 EN Official Journal of the European Union 26.10.2012 4. The	European Parliament	and the Council, acting in accordance with the
25	CELEX_12012E...	of the Member States. For the purposes of paragraph 1, the	European Parliament	and the Council, acting in legislative procedure, shall
26	CELEX_12012E...	the areas referred to in paragraph 2 and in Article 83. The	European Parliament	and the Council, acting in accordance with the

Search Query  Words  Case  Regex Results Set All hits Context Size 10 token(s)

European Parliament Start  Adv Search

Sort Options Sort to right Sort 1 1R Sort 2 2R Sort 3 3R Order by freq

Progress 100%

Рис. 1. Concordance програми AntConc для візуалізації пошукової одиниці  
European Parliament у тексті  
“Consolidated Version of the Treaty on the Functioning of the European Union”  
мультилінгвального корпусу

Окрім заданого тексту, конкорданс може опрацювати й інші тексти підкорпусу European Parliament мультилінгвального корпусу іншими мовами. У цьому випадку користувач корпусу та корпусної програми фактично обробляє за невеликий проміжок часу великий масив інформації певної тематики одночасно, фіксуючи важливі для виконання власного дослідницького завдання особливості контекстного вживання пошукових елементів. Користувач за допомогою корпусу формує та вдосконалює навички корпусного аналізу в процесі оволодіння функціоналом корпусних інструментів.

Розглянемо ще одну, теж ефективну, комп'ютерну програму WordSmith Tools (<http://www.lexically.net/wordsmith/index.html>), яка створює конкорданс пошукового елемента після завантаження у вікно Concord того ж файлу тексту англійською мовою “Consolidated Version of the Treaty on the Functioning of the European Union”, конвертованого з формату pdf у формат txt. Після здійснення цієї операції на запит іншого пошукового елемента European Central Bank отримано її конкорданс.

N	Concordance	Set	Tag	ent.	Sent.	Para.	lead.	head.	ect.	Sect.	File	Date	%
1	«The Council shall act after consulting the European Central Bank. »2. The Council, on	441	43	365	73	n/a	n/a	0	0	0	0_12012E_TXT_EN_TXT.txt	:023.06.28 00:00	18,3%
2	as a third decision-making body of the European Central Bank. 2. ¶If and as long	462	1	382	1	n/a	n/a	0	0	0	0_12012E_TXT_EN_TXT.txt	:023.06.28 00:00	19,2%
3	deliberation to the Governing Council of the European Central Bank. 2. ¶The President	979	1	809	1	n/a	n/a	0	0	0	0_12012E_TXT_EN_TXT.txt	:023.06.28 00:00	37,4%
4	«The Council »shall act after consulting the European Central Bank. »3. For the	443	31	366	31	n/a	n/a	0	0	0	0_12012E_TXT_EN_TXT.txt	:023.06.28 00:00	18,3%
5	from the Commission and after consulting the European Central Bank. »4. The Council,	420	54	347	54	n/a	n/a	0	0	0	0_12012E_TXT_EN_TXT.txt	:023.06.28 00:00	17,3%
6	European Union C 326/233 »Article 9 »The European Central Bank »9.1. The ECB which,	0	0	n/a	n/a	0	0	0	0	0	0_12012E_TXT_EN_TXT.txt	:023.06.28 00:00	57,1%
7	»consulting the European Parliament and the European Central Bank, adopt the	0	0	n/a	n/a	0	0	0	0	0	0_12012E_TXT_EN_TXT.txt	:023.06.28 00:00	16,7%
8	act either on a »recommendation from the European Central Bank and after consulting	0	0	n/a	n/a	0	0	0	0	0	0_12012E_TXT_EN_TXT.txt	:023.06.28 00:00	17,2%
9	the Court of Justice, recommendations from the European Central Bank and requests from	0	0	n/a	n/a	0	0	0	0	0	0_12012E_TXT_EN_TXT.txt	:023.06.28 00:00	48,5%
10	the members of the »Executive Board of the European Central Bank and the Governors	0	0	n/a	n/a	0	0	0	0	0	0_12012E_TXT_EN_TXT.txt	:023.06.28 00:00	37,2%
11	the Court of Justice, recommendations from the European Central Bank and requests from	0	0	n/a	n/a	0	0	0	0	0	0_12012E_TXT_EN_TXT.txt	:023.06.28 00:00	47,7%
12	brought by the Court of »Auditors, by the European Central Bank and by the	0	0	n/a	n/a	0	0	0	0	0	0_12012E_TXT_EN_TXT.txt	:023.06.28 00:00	35,5%
13	»either on a recommendation from the European Central Bank and after consulting	0	0	n/a	n/a	0	0	0	0	0	0_12012E_TXT_EN_TXT.txt	:023.06.28 00:00	61,7%

N	Concordance	Set	Tag	ent.	Sent.	Para.	lead.	head.	ect.	Sect.	File	Date	%
1	in Article 129(4), the European Central Bank shall be entitled to impose fines or periodic »penalty payments on undertakings for failure to comply with obligations under its regulations and »decisions. 26.10.2012 ¶EN Official Journal of the European Union C 326/105 »Article 133 »Without prejudice to the powers of the European Central Bank, the European Parliament and the »Council, acting in accordance with the ordinary legislative procedure, shall lay down the measures »necessary for the use of the euro as the single currency. »Such measures shall be adopted after »consultation of the European Central Bank. ¶CHAPTER 3 »	0	0	n/a	n/a	0	0	0	0	0	0_12012E_TXT_EN_TXT.txt	:023.06.28 00:00	17,6%
2	of the Treaties, collect any information and carry »out any checks required for the performance of the tasks entrusted to it. ¶Article 338 »(ex Article 285 TEC) »1. »Without prejudice to Article 5 of the Protocol on the Statute of the European System of Central »Banks and of the European Central Bank, the European Parliament and the Council, acting in »accordance with the ordinary legislative procedure, shall adopt measures for the production of »statistics where necessary for the performance of the activities of the Union. ¶»26.10.2012 EN Official Journal of the European Union C 326/193 »»2.	0	0	n/a	n/a	0	0	0	0	0	0_12012E_TXT_EN_TXT.txt	:023.06.28 00:00	45,3%
3	as long as there are Member States with a derogation,	462	1	382	1	n/a	n/a	0	0	0	0_12012E_TXT_EN_TXT.txt	:023.06.28 00:00	19,2%

Рис. 2. Конкорданс програми WordSmith Tools для візуалізації пошукового елемента “Consolidated Version of the Treaty on the Functioning of the European Union” мультилінгвального корпусу

У створеному програмою конкордансі маркуємо необхідний для подальшого дослідження пошуковий елемент і за допомогою клавіші F8 отримуємо результати роботи інструмента “Розширений контекст”, який надає доступ до кількох фрагментів вихідного тексту. Розширений контекст сприяє збереженню включеності пошукового елемента в певній ситуації, містить відомості про час створення та використання тексту, джерело публікації тощо. Зазначена функція розширеного контексту допомагає реконструювати ситуацію, в якій зафіксовано вживання пошукового елемента European Central Bank.

Інструмент “Розширений контекст” для пошукового елемента наявний у наборі комп’ютерних сервісів програми MonoConc Pro (<http://www.athel.com/mono.html>), яка призначена для роботи в комп’ютерних мережах системи Windows у різних версіях (W95 і вище) та достатньо проста у використанні. Програма оснащена низкою функцій, серед яких: контекстуальний пошук, пошук за регулярним виразом, частотний пошук за тегами, за сполучністю слів та іншими опціями. Однак її обмеження пов’язані з тим, що вона є комерційною програмою.

Для того щоб побачити ширший контекст пошукового елемента (тобто фрагмент попереднього та наступного тексту), користувач обирає будь-який рядок у вікні отриманих первинних конкордансних результатів. Розширений контекст пошукової одиниці відображається у верхньому вікні конкордансу.

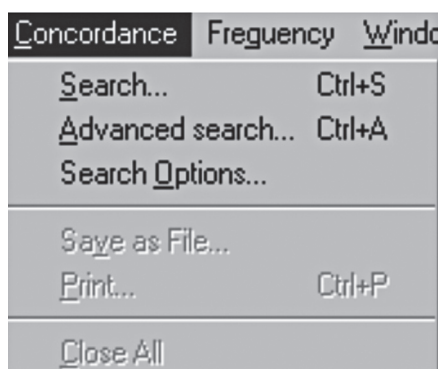


Рис. 3. Меню Конкордансу програми MonoConc Pro

Оскільки мультилінгвальний корпус – це тематично орієнтоване зібрання текстів з Європеїстики, то його тематичні блоки є підкорпусами текстів, які можна створити за допомогою, наприклад, тега < European >. Класифікацію текстів у підкорпуси з використанням тегів можна здійснити за допомогою інструментарію комп’ютерної програми CATMA (Computer Aided Textual Markup and Analysis) (<http://www.catma.de/download>). Версія CATMA 3.2 JAVA підтримується для Mac та Windows PC та доступна на сайті для завантаження. Заснована на відомій програмі “Usebase”, вона дозволяє обробити тексти у двох її складниках: Tagger та Analyzer (Рис. 4).

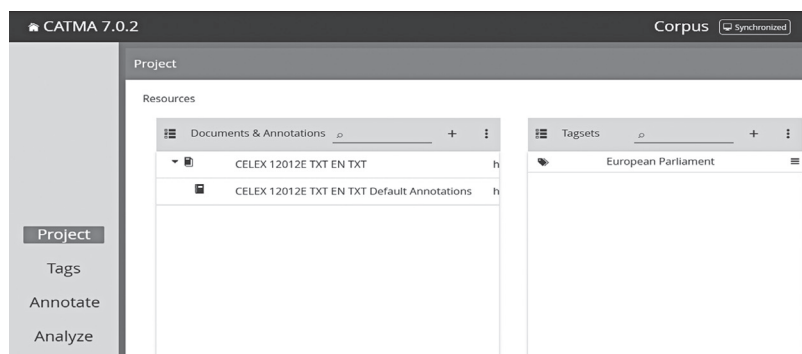


Рис. 4. Діалогове вікно комп’ютерної програми CATMA для створеного тега European Parliament

Ще одна досить відома програма KORP (<https://spraakbanken.gu.se/korp/>) може бути використана для дослідження Європеїстики як комп’ютерний пошуковий інструмент конкордансного типу, що включає колекції текстів різних мов, різної тематики, різних жанрів тощо. Наприклад, на запит пошукового елемента European у підкорпусі текстів іспанською мовою програма KORP побудувала конкорданс із можливістю представлення розширеного контексту шляхом натискання в меню функції Show context. Вікно конкордансу, окрім лівобічного і правобічного контекстів, містить праворуч метадані тексту (Рис. 5). Розширений контекст надав можливість ознайомитися зі змістом усього тексту, у якому йдеться про діяльність Європейського економічного співтовариства (ЄЕС), що став фундатором Європейського Союзу. Цікавими виявилися відомості про функціонування економічних інституцій ЄЕС,



про умови для вступу країн до ЄЕС, основні з яких такі: ратифікація Європейської конвенції з прав людини, реформування економіки й соціальної сфери, дотримання принципів плюралістичної демократії тощо.

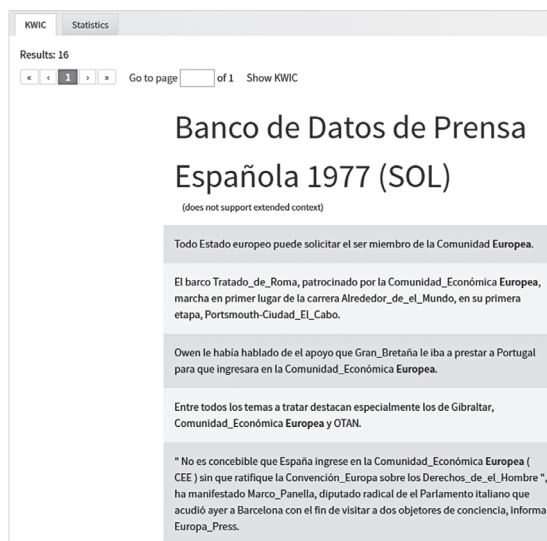
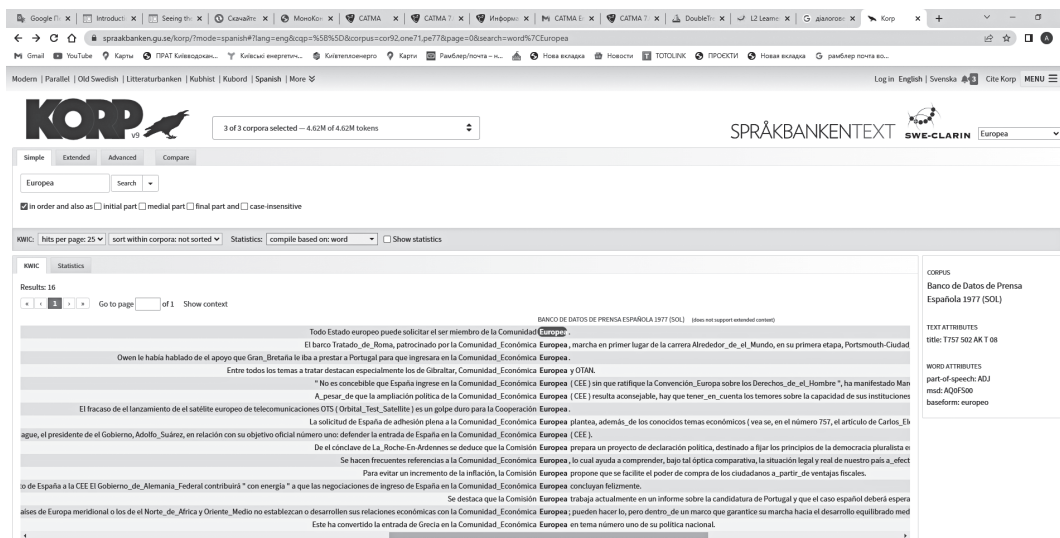


Рис. 5. Concordance програми KORP для візуалізації пошукового елемента European у підкорпусі текстів іспанською мовою з можливістю представлення розширеного контексту

Окрім стислого огляду комп'ютерних програм, які можна використовувати в процесі роботи з мультилінгвальним комбінованим корпусом, зокрема, для дослідження Європеїстики, вкажемо на роль і значення корпусної статистики, якою оснащені всі комп'ютерні програми для виконання різних дослідницьких завдань. Наприклад, програма KORP дозволяє отримати статистичні дані шляхом натискання в меню функції Show statistics (Рис. 6) та побудувати діаграми шляхом натискання в меню функції Graph (Рис. 7).

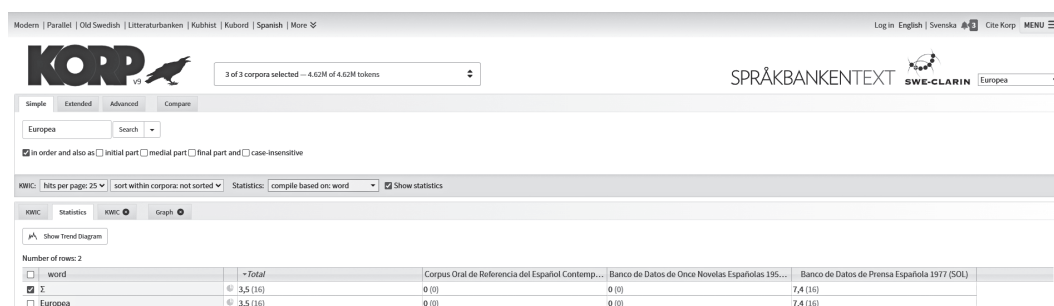


Рис. 6. Статистичні дані пошукового елемента European у підкорпусі текстів іспанською мовою програми KORP

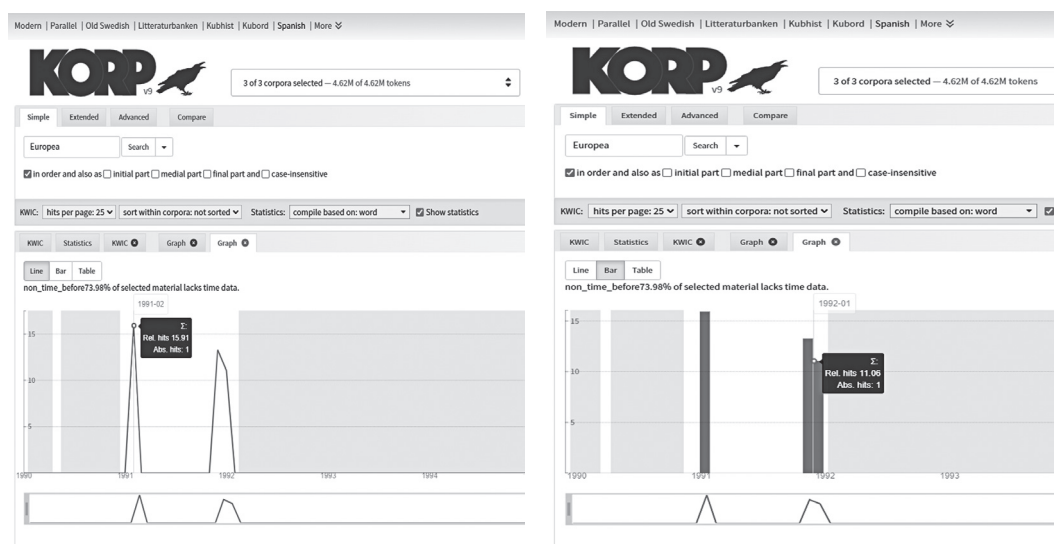


Рис. 7. Побудова діаграм у підкорпусі текстів іспанською мовою програми KORP

Сьогодні, коли текстові бази даних володіють конкордансом і корпус-менеджерами, до основних методів роботи з корпусами текстів можуть бути віднесені всі види корпусної лінгвостатистики: підрахунок абсолютної частотності слів, виділення найбільш змістових частотних слів за окремими підкорпусами (включаючи дискурсивно релевантні), побудова діаграм частотності слова за кожним підкорпусом, виділення кластерів (статистично значущих ланцюжків слів, розташованих контактено), корпусний пошук сполучуваності (колокацій, які можуть розташовуватися дистантно).

Зіставний аналіз кількісних і статистичних даних про частоту вживання слів або колокацій, про ступінь їхньої семантичної близькості допомагає встановлювати зв'язки між пошуковими одиницями різних мов як усередині тегів, так і між ними.

**Висновки та перспективи дослідження.** Підбиваючи підсумок критичного огляду комп'ютерних програм, які можна використовувати для роботи з мультилінгвальним зіставним корпусом текстів за тематикою Європеїстика, відзначимо, що функціонал комп'ютерних інструментів корпус-менеджерів AntConc, WordSmith, WordList,

MonoConcPro та CATMA дозволяє: будувати як окремі KWIC, так і повні конкордансні списки (KWIC дає змогу простежити, у яких контекстах (зокрема й розширених із наданням метаданих текстів: автора, джерело публікації, рік видання джерела тощо) та дискурсах вжито колокати, наприклад, European Central Bank, European Parliament, і яка їхня семантична близькість у різних мовах); розмежувати функціонал того чи іншого контекстуального значення пошукового елемента в його найбільш імовірній лівобічній та правобічній валентності; побачити результати статистичної обробки необхідної інформації з корпусних тегів; зберігати та роздруковувати результати; швидко обробляти запити та видавати результати; підтримувати різні формати текстових даних (txt, doc, rtf, html та ін.).

Перспективи подальших досліджень полягають в описі результатів практичної апробації запропонованої методики роботи з корпус-менеджерами AntConc, WordSmith, WordList, MonoConc Pro, CATMA та KORP для дослідження Європеїстики.

#### ЛІТЕРАТУРА

- Андрушенко, О. (2022). Комплексна методика дослідження фокусувальних адвербів у сучасних та історичних корпусах текстів англійської мови. *Innovative pathway for the development of modern philological sciences in Ukraine and EU countries* (pp. 26–54). Publishing House “Baltija Publishing”.
- Бобер, Н. М. (2020). *Когнітивно-семантична матриця фразових дієслів емоційних станів людини у Британському національному корпусі* [Дис. канд. філол. наук, Національний педагогічний університет імені М. П. Драгоманова].
- Жуковська, В. В. (2013). Ресурси корпусної лінгвістики у дослідженні історичної динаміки мови. В *Матеріали міжнародної наукової конференції “Слово і речення: синтактика, семантика, прагматика”* (с. 151–156). Київський університет імені Бориса Грінченка.
- Мейзерська, І. В. (2014). Корпусний підхід у сучасній лінгвістиці: перспективи і можливості застосування. *Науковий вісник кафедри Юнеско. Серія Філологія. Педагогіка. Психологія*, 28, 53–58.
- Широков, В. А., Бугаков, О. В., & Грязнухіна, Т. О. (2005). *Корпусна лінгвістика*. Київ: Довіра.
- Afli, H., Barrault, L., & Schwenk, H. (2012). Traduction automatique à partir de corpus comparables: extraction de phrases parallèles à partir de données comparables multimodales. *Actes de la conférence conjointe JEP-TALN-RECITAL*, 2, 447–454.
- Barzilay, R., & Lee, L. (2003). *Learning to Paraphrase: An Unsupervised Approach Using Multiple-Sequence Alignment*. Edmonton.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press.
- Bowker, L., & Pearson, J. (2002). *Working with Specialized Language: A Practical Guide to Using Corpora*. Routledge.
- Brawn, S. (2007). Designing and exploiting small multimedia corpora for autonomous learning and teaching. In E. Hidalgo, L. Querada, J. Santana (Eds.), *Corporain the Foreign Language Classroom: Selected Papers from the Sixth International Conference on Teaching and Language Corpora* (TaLC 6, p. 32–33.). Rodopi.
- Cartoni, B., & Deléger, L. (2011). Découverte de patrons paraphrastiques en corpus comparable: une approche basée sur les n-grammes. In *Actes de la 18e conférence sur le Traitement Automatique des Langues Naturelles. Articles courts* (p. 182–187). Montpellier.

- Culo, O., Schirra, S.H., Neumann, S., & Vela, M. (2008). Empirical Studies on Language Contrast Using the English-German Comparable and Parallel Corpus. In N. Calzolari (Ed.), *Workshop abstracts / Sixth International Conference on Language Resources and Evaluation* (p. 47–51). Palais des Congrès Mansour Eddahbi.
- Elhadad, N., & Sutaria, K. (2007). Mining a Lexicon of Technical Terms and Lay Equivalents. In *ACL BioNLP Workshop* (p. 49–56). Prague.
- Guidère, M. (2010). *Introduction à la traductologie. Penser la traduction: hier, aujourd'hui, demain*. De Boeck Université.
- Johns, T. (1991). Should You Be Persuaded – Two Samples of Data-driven Learning Materials. *Classroom Concordancing: ELR Journal*, 4, 1–16.
- Kapranov, Ya. (2022). AntConc corpus manager and its possibilities for keywords with resilience semantics search. In R. Vasko (Ed.), *Language. Culture. Discourse* (p. 22–32). PC TECHNOLOGY CENTER.
- Kilgariff, A. (2001). Comparing Corpora. *International Journal of Corpus Linguistics*, 6, 97–133.
- McEnery, A. M. (2003). Corpus Linguistics. In R. Mitkov (Ed.), *The Oxford Handbook of Computational Linguistics* (p. 448–463). Oxford University Press.
- McEnery, A., & Xiao, Z. (2007). Parallel and Comparable Corpora: What is Happening. In *Incorporating Corpora: Translation and the Linguist* (p. 18–31). Multilingual Matters.
- O’Keeffe, A., McCarthy, M., & Carter, R. (2007). *From Corpus to Classroom : Language Use and Language Teaching*. Cambridge University Press.
- Rauf, S. (2012). *Efficient Corpus Selection for Statistical Machine Translation: thèse de Doctorat*. Université du Maine.
- Reppen, R. (2010). *Using Corpora in the Language Classroom*. Cambridge University Press.
- Resnik, P., & Smith, N. A. (2003). The Web as a parallel corpus. *Computational Linguistics*, 29(3), 349–380.
- Scarpa, F. (2010). *La traduction spécialisée: une approche professionnelle à l’enseignement de la traduction*. University of Ottawa Press.
- Zanettin, F. (1998). Bilingual Corpora and the Training of Translators. *Meta*, 4(43), 616–630.

## REFERENCES

- Andrushenko, O. (2022). Kompleksna metodyka doslidzhennya fokusuval'nykh adverbiv u suchasnykh ta istorychnykh korpusakh tekstiv anhliys'koyi movy. V *Innovative pathway for the development of modern philological sciences in Ukraine and EU countries* (pp. 26–54). Publishing House “Baltija Publishing”.
- Bober, N. M. (2020). *Kohnityvno-semantychna matrytsya frazovykh diyesliv emotsiynykh staniv lyudyny u Brytans'komu natsional'nomu korpusi* [Dys. kand. filol. nauk, Natsional'nyy pedahohichnyy universytet imeni M. P. Drahomanova].
- Zhukovs'ka, V. V. (2013). Resursy korpusnoyi linhvistyky u doslidzhenni istorychnoyi dynamiky movy. V *Materialy mizhnarodnoyi naukovoyi konferentsiyi “Slovo i rechennya: syntaktyka, semantyka, prahmatyka”* (s. 151–156). Kyiv: Kyivskyy universytet imeni Borysa Hrinchenka.
- Meyzers'ka, I. V. (2014). Korpusnyy pidkhid u suchasniy linhvistytsi: perspektyvy i mozhlyvosti zastosuvannya. *Naukovyy visnyk kafedry Yunesco. Seriya Filolohiya. Pedahohika. Psykholohiya*, 28, 53–58.
- Shyrovkov, V. A., Buhakov, O. V., & Hryaznukhina, T. O. (2005). *Korpusna Linhvistyka*. Dovira.

- Afli, H., Barrault, L., & Schwenk, H. (2012). Traduction automatique à partir de corpus comparables: extraction de phrases parallèles à partir de données comparables multimodales. *Actes de la conférence conjointe JEP-TALN-RECITAL*, 2, 447–454.
- Barzilay, R., & Lee, L. (2003). *Learning to Paraphrase: An Unsupervised Approach Using Multiple-Sequence Alignment*. Edmonton.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics : Investigating Language Structure and Use*. Cambridge University Press.
- Bowker, L., & Pearson, J. (2002). *Working with Specialized Language: A Practical Guide to Using Corpora*. Routledge.
- Brawn, S. (2007). Designing and exploiting small multimedia corpora for autonomous learning and teaching. In E. Hidalgo, L. Quereda, J. Santana (Eds.), *Corporain the Foreign Language Classroom : Selected Papers from the Sixth International Conference on Teaching and Language Corpora* (TaLC 6, p. 32-33.). Rodopi.
- Cartoni, B., & Deléger, L. (2011). Découverte de patrons paraphrastiques en corpus comparable: une approche basée sur les n-grammes. In *Actes de la 18e conférence sur le Traitement Automatique des Langues Naturelles. Articles courts* (p. 182–187). Montpellier.
- Culo, O., Schirra, S.H., Neumann, S., & Vela, M. (2008). Empirical Studies on Language Contrast Using the English-German Comparable and Parallel Corpus. In N. Calzolari (Ed.), *Workshop abstracts / Sixth International Conference on Language Resources and Evaluation* (pp. 47–51). Palais des Congrès Mansour Eddahbi.
- Elhadad, N., & Sutaria, K. (2007). Mining a Lexicon of Technical Terms and Lay Equivalents. In *ACL BioNLP Workshop* (p. 49–56). Prague.
- Guidère, M. (2010). *Introduction à la traductologie. Penser la traduction: hier, aujourd'hui, demain*. De Boeck Université.
- Johns, T. (1991). Should You Be Persuaded – Two Samples of Data-driven Learning Materials. *Classroom Concordancing: ELR Journal*, 4, 1–16.
- Kapranov, Ya. (2022). AntConc corpus manager and its possibilities for keywords with resilience semantics search. In R. Vasko (Ed.), *Language. Culture. Discourse* (p. 22–32). PC TECHNOLOGY CENTER.
- Kilgariff, A. (2001). Comparing Corpora. *International Journal of Corpus Linguistics*, 6, 97–133.
- McEnery, A. M. (2003). Corpus Linguistics. In R. Mitkov (Ed.), *The Oxford Handbook of Computational Linguistics* (p. 448–463). Oxford University Press.
- McEnery, A., & Xiao, Z. (2007). Parallel and Comparable Corpora: What is Happening. In *Incorporating Corpora: Translation and the Linguist* (p. 18–31). Multilingual Matters.
- O’Keeffe, A., McCarthy, M., & Carter, R. (2007). *From Corpus to Classroom : Language Use and Language Teaching*. Cambridge University Press.
- Rauf, S. (2012). *Efficient Corpus Selection for Statistical Machine Translation: thèse de Doctorat*. Université du Maine. <https://tel.archives-ouvertes.fr/tel-00732984/>.
- Reppen, R. (2010). *Using Corpora in the Language Classroom*. Cambridge University Press.
- Resnik, P., & Smith, N. A. (2003). The Web as a parallel corpus. *Computational Linguistics*, 29(3), 349–380.
- Scarpa, F. (2010). *La traduction spécialisée: une approche professionnelle à l’enseignement de la traduction*. University of Ottawa Press.
- Zanettin, F. (1998). Bilingual Corpora and the Training of Translators. *Meta*, 4(43), 616–630.




Співфінансується Європейським Союзом. Проте висловлені думки належать лише авторці і не обов'язково збігаються з поглядами Європейського Союзу чи Європейського виконавчого агентства з питань освіти та культури. Ні Європейський Союз, ні орган, що надає гранти, не відповідають за наведені в статті погляди.

*Дата надходження до редакції 01.06.2023*

*Ухвалено до друку 22.06.2023*

### **Відомості про автора**

<p><b>Корольова Алла Валер'янівна,</b></p> <p>доктор філологічних наук, професор, проректор з наукової роботи Київського національного лінгвістичного університету</p> <p>e-mail: <a href="mailto:alla.korolyova@knlu.edu.ua">alla.korolyova@knlu.edu.ua</a></p>		<p><b>Сфера наукових інтересів:</b></p> <p>когнітивна лінгво- макрокомпаративістика, природні та штучні мови, реконструкція, корпусна лінгвістика</p>
--	--	---