

УДК 811.111:34(045)

DOI <https://doi.org/10.24919/2308-4863/55-3-25>

Олена ТКАЧИК,

orcid.org/0000-0001-6544-3046

кандидат філологічних наук, доцент,
доцент кафедри теорії, практики та перекладу англійської мови
Національного технічного університету України
«Київський політехнічний інститут імені Ігоря Сікорського»
(Київ, Україна) *helen.tkachyk@gmail.com*

Анастасія ЧЕРНИЖ,

orcid.org/0000-0002-2849-5995

студентка II курсу магістратури факультету лінгвістики
Національного технічного університету України
«Київський політехнічний інститут імені Ігоря Сікорського»
(Київ, Україна) *chernyzh.anastasia@gmail.com*

КОРПУСНЕ ДОСЛІДЖЕННЯ АНГЛОМОВНИХ ЗАКОНОДАВЧИХ АКТІВ ЄВРОПЕЙСЬКОГО СОЮЗУ

Стаття присвячена корпусному дослідженню англomовних текстів законодавчих актів Європейського Союзу (далі ЄС), які включають регламенти, директиви та рішення і є основними юридичними документи ЄС, що регулюють життєдіяльність держав-членів. Оскільки раніше корпусний аналіз англomовних законодавчих актів ЄС не проводився, він надав змогу детально дослідити юридичні документи під іншим кутом. У ході дослідження застосовувались методи корпусної лінгвістики з метою виявлення особливостей функціонування лексичного складу текстів законодавчих актів ЄС. Було створено корпус текстів LAEU – Legislative Acts of the European Union, який має три підкорпуси: CR – Commission Regulation, DEPC – Directive of the European Parliament and of the Council та CID – Commission Implementing Decision рівнозначного обсягу – біля 16 тисяч слів кожний.

Використано корпусний менеджер AntConc та онлайн-платформу Voyant Tools для фіксування найбільш уживаних клішованих слів та словосполучень, а також ключових лексем кожного з актів ЄС. За своїм лексичним складом вони містять не лише термінологію, а також клішовані слова та словосполучення з інших сфер, які було об'єднано у семантичні групи. Проведене дослідження показало, що законодавчі акти ЄС за своїм лексичним складом містять юридичну термінологію, що відповідає нормам права. Було зафіксовано, що найбільша кількість ключових слів та словосполучень у регламенті стосуються тематики «Джерела світла», директиві – тематики «Облігації з покриттям», а у рішенні – тематики «Інфекційні захворювання». Найбільшу кількість частотності та ключових слів було зафіксовано у рішенні, а найменше – у директиві. З'ясовано, що найкоротшими реченнями за кількістю слів є речення у рішенні, тому що вони мають меншу інформативність на відміну від речень у регламенті, що є найдовшими серед трьох типів текстів. Окрім того, сформовано «Хмару Слів» з ключових лексем у кожному з актів.

Ключові слова: корпусна лінгвістика, корпусний аналіз, корпусний менеджер, корпус, законодавчі акти, Європейський Союз.

Olena TKACHYK,

orcid.org/0000-0001-6544-3046

Candidate of Philological Science, Associate Professor,
Associate Professor at the Department of Theory, Practice and Translation of English
National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»
(Kyiv, Ukraine) *helen.tkachyk@gmail.com*

Anastasiia CHERNYZH,

orcid.org/0000-0002-2849-5995

2nd Year Master's Student at the Faculty of Linguistics
National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»
(Kyiv, Ukraine) *chernyzh.anastasia@gmail.com*

CORPUS STUDY OF ENGLISH LEGISLATIVE ACTS OF THE EUROPEAN UNION

The article is devoted to a corpus study of English-language texts of legislative acts of the European Union (hereinafter EU), which include regulations, directives and decisions and are the main legal documents of the EU regulating the life

activities of member states. Since corpus analysis of English-language EU legislation had not been conducted before, it provided an opportunity to examine legal documents in detail from a different angle. In the course of the study, the methods of corpus linguistics were used in order to identify the peculiarities of the functioning of the lexical composition of the texts of EU legislative acts. A corpus of LAEU texts – Legislative Acts of the European Union was created, which has three sub-corpus: CR – Commission Regulation, DEPC – Directive of the European Parliament and of the Council, CID – Commission Implementing Decision of equal volume – about 16 thousand words each.

The corpus manager AntConc and the Voyant Tools online platform were used to capture the most used clichéd words and phrases, as well as key tokens of each of the EU acts. According to their lexical composition, they contain not only terminology, but also clichéd words and phrases from other fields, which were combined into semantic groups. The conducted research showed that the legal acts of the EU contain legal terminology that corresponds to the norms of law. It was recorded that the largest number of key words and phrases in the regulation relate to the topic «Light sources», the directive – to the topic «Covered bonds», and in the decision – to the topic «Infectious diseases». The highest number of frequencies and keywords was recorded in the decision, and the lowest in the directive. It was found that the shortest sentences in terms of the number of words are the sentences in the decision, because they have less informativeness in contrast to the sentences in the regulation, which are the longest among the three types of texts. In addition, a «Word Cloud» was formed from key lexemes in each of the acts.

Key words: corpus linguistics, corpus analysis, corpus manager, corpus, legislative acts, European Union.

Постановка проблеми. Корпусні дослідження є невід'ємною частиною сучасної лінгвістики. Завдяки корпусному аналізу можна не лише з легкістю уточнювати результати вже проведених досліджень мовних явищ, а й розширювати діапазон і масштаби аналізованого мовного матеріалу за рахунок значного збільшення обсягу текстів та аспектів вивчення мови. Завдяки автоматичній обробці текстів корпусні методи відкривають нові можливості для лінгвістики.

Однією з малодосліджених сфер, які ще не підлягали обробці корпусними методами є англійські законодавчі акти ЄС. З інтеграцією України до ЄС виникає потреба у детальному вивченні правових міжнародних документів з лінгвістичної точки зору. Тому цікавим і актуальним вбачається визначення специфіки функціонування лексичного складу цих документів, встановлення ключових лексем та контексту їх вживання.

Аналіз досліджень. Корпусна лінгвістика є однією з найперспективніших галузей лінгвістичних наук, яка постійно розвивається, а провідні науковці США та Європи активно працюють над укладанням не лише загальномовних, а й специфічних корпусів.

Корпусна лінгвістика зацікавлює своїми можливостями, швидкістю отриманих результатів та ефективністю багатьох вчених з різних країн світу, серед них Д. Бібер, А. Уїлсон, В. В. Жуковська, К. Майер, М. Кіто, О. М. Карпова, А. Люделінг, В. Крофт, В. П. Захаров та інші. Новизна дослідження полягає у тому, що вперше буде здійснено корпусний аналіз мовних кліше на матеріалах англійських законодавчих актів (регламентів, директив, рішень) ЄС.

Мета статті – з'ясувати та описати специфіку функціонування лексичного складу текстів англійських законодавчих актів ЄС за допомогою корпусного менеджера AntConc та Voyant Tools.

Виклад основного матеріалу. Термін «корпусна лінгвістика» почав набувати популярності починаючи з ХХ століття, у той час, коли активно розвивалася комп'ютерна індустрія. Визначною науковою подією була публікація збірника «Corpus Linguistics: Recent Development in the Use of Computer Corpora in English Language Research» 1983 року (Скобнікова, 2018: 116). Корпусна лінгвістика є розділом і напрямом прикладної лінгвістики, тому вона займається побудовою корпусів, а також здійснює аналіз текстів за допомогою комп'ютерних технологій (Viber, 2004: 4). З активним технологічним розвитком та збільшенням обсягу інформації, яку потрібно було систематизувати, почали з'являтися корпуси текстів та корпусні менеджери, які дозволяли автоматизувати багато процесів обробки природної мови. Саме ці лінгвістичні корпуси надали змогу швидкого пошуку інформації та розмежування результатів відповідно до різних лінгвістичних критеріїв та параметрів пошуку (Яницька, 2016: 160).

Особливістю застосування корпусних методів аналізу є те, що вони дають автоматизовані, репрезентативні та валідні результати, максимально зменшуючи похибку результатів дослідження. Важливою властивістю корпусних методів є вивчення лексичних одиниць в контекстах, отримання результатів щодо частотності лексем та граматичних категорій, їхньої сполучуваності та управління у тексті (Жуковська, 2013). Корпусний аналіз дозволяє простежити наявність чи відсутність сталих виразів, фразеологізмів, здійснювати пошук словоформ за допомогою лем чи набору морфологічних ознак тощо.

Для проведення корпусного аналізу текстів необхідним є оволодіння спеціальними комп'ютерними програмами, які надають можливість подальшої обробки лінгвістичних даних та проведення дослідження. Така робота виконується за допомогою

спеціальних програм, таких як коркндасери чи корпусні менеджери, яких на сьогоднішній день є велика кількість (Скобнікова, 2018: 117).

Для дослідження було використано корпусний менеджер *AntConc*, розробленим Ентоні Лоуренсом. Цей корпусний менеджер є однією з найкращих пошукових систем, оскільки він здійснює пошук даних та надає користувачеві статичну інформацію у досить зручному вигляді. Вказаний корпусний менеджер надає широкий спектр можливостей і є однією з найпростіших програм, яку можна легко та безкоштовно встановити на комп'ютер. *AntConc* також полегшує дослідницьку роботу завдяки простому та зручному у користуванні інтерфейсу. Він є надзвичайно потужним конкордансером, який надає широкий спектр функціональних можливостей, генеруючи список ключових слів, список частотності уживаних словосполучень, їхню сполучуваність, а також дозволяє здійснювати пошук, використовуючи одне слово чи словосполучення (Скобнікова, 2018: 117).

Ще одна програма, що була застосована у дослідженні, – це *Voyant tools*. На відміну від *AntConc*, вона є онлайн-платформою та не потребує жодних завантажень на комп'ютер. *Voyant tools* здатна демонструвати результати дослідження за допомогою діаграм та схем різних типів, а також дозволяє створити хмару слів з даного корпусу.

Матеріалом для дослідження слугували три типи англomовних фахових текстів права ЄС, а саме регламент, директива та рішення. Основним критерієм відбору текстів було наявність у них нетипової термінології, тобто не лише юридичних термінів. Створений корпус було названо LAEU – Legislative Acts of the European Union, він складається з 3 поданих актів, з них регламент (COMMISSION REGULATION (EU) 2019/2020 of 1 October 2019 laying down eco-design requirements for light sources and separate control gears), директива (DIRECTIVE (EU) 2019/2162 OF THE EUROPEAN PARLIAMNET AND OF THE COUNCIL of 27 November 2019 on the issue of covered bonds and covered bond public supervision) та рішення (COMMISSION IMPLEMENTING DECISION (EU) 2018/945 of 22 June 2018 on the communicable diseases and related special health issues to be covered by epidemiological surveillance as well as relevant case definitions). Загальний обсяг корпусу становить 48 807 слів. В межах загального корпусу було виокремлено три підкорпуси

CR – Commission Regulation, DEPC – Directive of the European Parliament and of the Council та CID – Commission Implementing Decision, що містять рівнозначну кількість слів – близько 16 000.

У кожному підкорпусі аналізувалися частотність вживання слів та словосполучень та ключові слова. Під частотністю розуміємо кількість вживань слів чи словосполучень у певному підкорпусі, а під показником «ключові слова» – слова, що використовуються для опису тематики у кожному з підкорпусів.

Підкорпус ЄС CR (Commission Regulation) налічує 15 267 слів та 2 130 словосполучень. Середня кількість слів у речення становить 51, що вказує на найдовші речення серед трьох типів аналізованих текстів. Це може свідчити про те, що регламент ЄС містить більше лексичне та інформаційне наповнення речень, на відміну від директиви та рішення. Серед слів та словосполучень було виокремлено 62 ключові лексичні одиниці, що відносяться до тематики «Джерела світла». Серед найуживаніших термінів програма *AntConc* зафіксувала такі, як: *light* (354 вживання), *light source* (152), *control gear* (69), *a luminous flux* (29), *mains* (27), *containing product* (12), *chromaticity* (11), *colouring rendering index* (7), *gas discharge* (6).

Вдалося зафіксувати топ 5 ключових слів у даному корпусі – *light* (354), *source* (159), *control* (146), *shall* (130), *sources* (125) (рис. 1). Вони відрізняються своєю частотністю від попередніх 62.

Наступне дослідження було проведене з директивою ЄС (DIRECTIVE (EU) 2019/2162 OF THE EUROPEAN PARLIAMNET AND OF THE

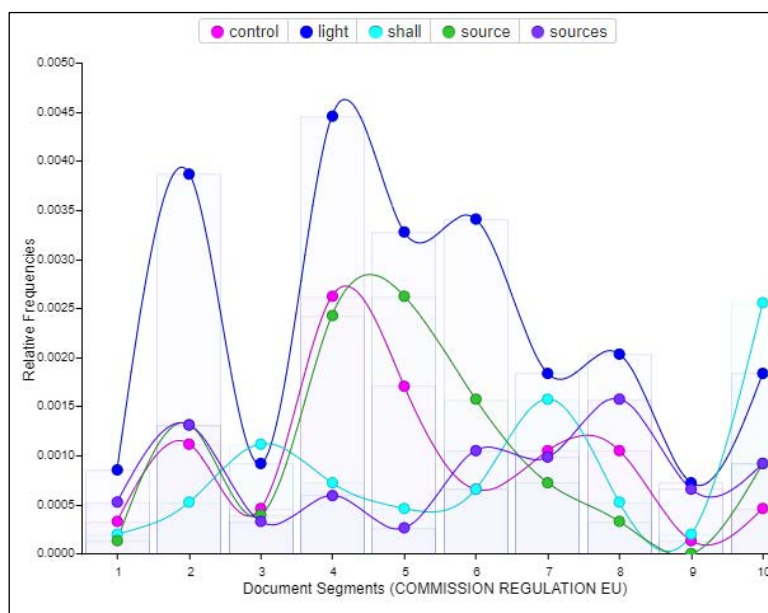


Рис. 1. Діаграма частоти вживання слів тематики «Джерела світла» англійською мовою

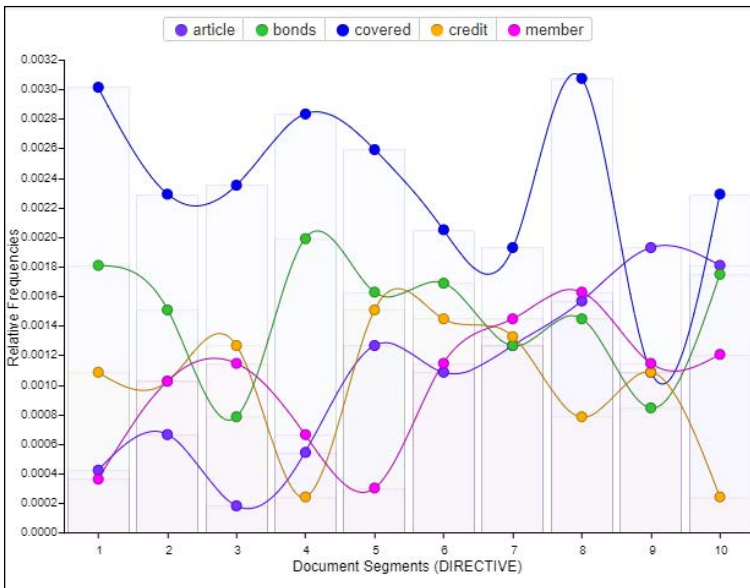


Рис. 2. Діаграма частоти вживання слів тематики «Облігації з покриттям» англійською мовою

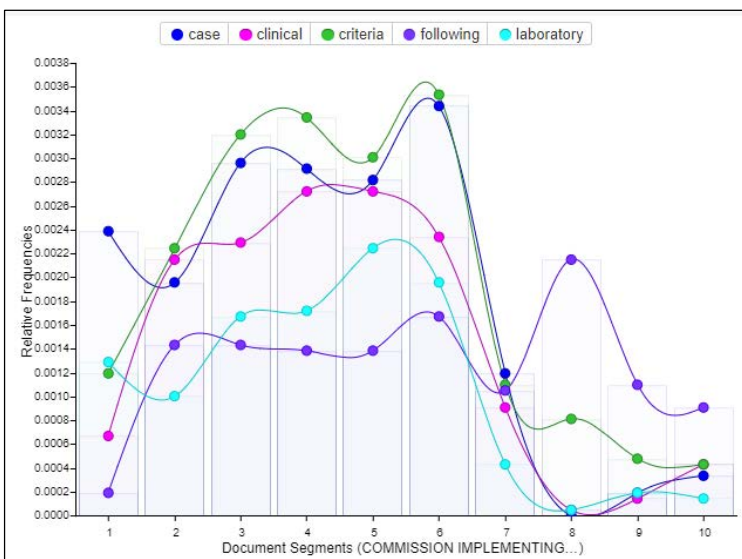


Рис. 3. Діаграма частоти вживання слів тематики «Інфекційні захворювання» англійською мовою

COUNCIL of 27 November 2019 on the issue of covered bonds and covered bond public supervision). Було зафіксовано, що даний корпус налічує 16 595 слів та 1565 словосполучень. Визначено, що кожне речення містить, в середньому, 48 слів. Даний корпус налічує 22 лексичні одиниці, які належать до тематики «Облігації з покриттям». Найбільш уживаними термінами у даному підкорпусі виявилися: *covered bonds* (240 вживань), *credit institution* (88), *cover pool* (74), *cover assets* (65), *resolution* (24), *collateral assets* (23), *mortgage* (14), *public undertakings* (13). Серед ключових слів даного корпусу є *covered* (390), *bonds* (244), *article* (178), *member* (167), *credit* (166) (рис. 2).

Також досліджено імплементаційне рішення ради ЄС (COMMISSION IMPLEMENTING DECISION (EU) 2018/945 of 22 June 2018 on the communicable diseases and related special health issues to be covered by epidemiological surveillance as well as relevant case definitions). Зазначений корпус містить 16 945 слів та 2 479 словосполучень. Середня кількість слів у реченні досягає 40. Серед найуживаніших словосполучень у даному корпусі є *clinical criteria* (213 вживань), *laboratory criteria* (181), *confirmed case* (123), *following symptoms* (39), *criteria confirmation* (34), *case transmission* (24), *infection evidence* (21). Було встановлено топ 5 ключових слів у даному корпусі: *criteria* (405), *case* (381), *clinical* (302), *following* (266), *laboratory* (224) (рис. 3).



Рис. 4. «Хмара слів»

Крім того, за допомогою онлайн-платформи *Voyant tools* проведено візуалізацію термінів трьох законодавчих актів, які найчастіше зустрічаються у корпусах текстів, у «Хмарі слів» (рис. 4). Таким чином, вдалося виділити термінологію кожного із підкорпусів у більш наочний спосіб. За допомогою «Хмари слів» отримано узагальнені дані, що стосуються ключових слів певної тематики.

Таким чином, у даному дослідженні корпусними методами було проаналізовано три типи законодавчих актів ЄС за критеріями ключових слів, частотності та довжини речень.

Висновки. Отже, порівнюючи три типи законодавчих актів ЄС (регламент, директива, рішення), можна зробити висновок, що вони рівнозначні за лексичним складом. Найбільшу кількість частотності

та ключових слів було зафіксовано у рішенні, а найменше – у директиві. Законодавчі акти ЄС містять не лише юридичну термінологію, а й лексеми відповідно до сфер застосування актів. Було зафіксовано, що найбільша кількість ключових слів та словосполучень у регламенті стосуються тематики «Джерела світла», у директиві – тематики «Облігації з покриттям», а у рішенні – тематики «Інфекційні захворювання». Найкоротшими реченнями за кількістю слів є речення у рішенні, тому що вони мають меншу інформативність на відміну від речень у регламенті, що є найдовшими серед трьох типів текстів.

Перспективою дослідження є подальше контрастивне вивчення англomовних та українськомовних законодавчих актів ЄС із використанням корпусних методів аналізу.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Білик К. М., Махачашвілі Р. К. Корпусне дослідження текстів рубрики «Надзвичайні новини» у французькій, англійській та українській мовах. Науковий вісник Міжнародного гуманітарного університету. 2019. № 43. С. 151-154
2. Демська-Кульчинська О.М. Базові поняття корпусної лінгвістики. Українська мова. 2003. № 1(6). С. 40-45
3. Жуковська В. В. Вступ до корпусної лінгвістики: навчальний посібник. Житомир: Вид-во ім. І. Франка, 2013. 142 с.
4. Скобнікова О. Аналіз концепту FAMILY на базі американських національних корпусів. Науковий вісник Херсонського державного університету. 2018. № 32. С. 115-120.
5. Ткачик О. Використання корпусних методів у гендерних дослідженнях. Наукові записки Національного університету «Острозька академія». 2018. № 69. С. 194-196.
6. Яницька О. М. Корпусний аналіз для визначення ключових лексем-номінантів концепту VIRA на матеріалі англійської, української та французької мов. Науковий вісник Міжнародного гуманітарного університету. 2016. № 25. С. 160-162.
7. Biber D. *Corpus Linguistics: Investigating language structure and use*. Cambridge University Press, 2004. 300p.

REFERENCES

1. Bilyk K. M., Makhachashvili R. K. *Corpusne doslidzhennia tekstiv rubryky «Nadzvychnaini novyny» u frantsuzkii, anhliiskii ta ukrainskii movakh.* [Corpus research of the texts of the "Extraordinary News" section in French, English and Ukrainian languages.] *Scientific Bulletin of the International Humanitarian University.* 2019. Nr 43. pp. 151-154 [in Ukrainian]
2. Demska-Kulchynska O.M. *Bazovi poniattia korpusnoi linhvistyky.* [Basic concepts of corpus linguistics.] *Ukrainian language.* 2003. Nr 1(6). pp. 40-45 [in Ukrainian]
3. Zhukovska V. V. *Vstup do korpusnoi linhvistyky: navchalnyi posibnyk.* [Introduction to corpus linguistics: a study guide.] *Zhytomyr: I. Franka publication,* 2013. 142 p. [in Ukrainian]
4. Skobnikova O. *Analiz kontseptu FAMILY na bazi amerykanskykh natsionalnykh korpusiv.* [Analysis of the FAMILY concept based on American national corpora.] *Scientific Bulletin of Kherson State University.* 2018. Nr 32. pp. 115-120 [in Ukrainian]
5. Tkachyk O. *Vykorystannia korpusnykh metodiv u hendernykh doslidzhenniakh.* [Using Corpus Methods to Analyze Gender] *Scientific notes of the National University «Ostroh Academy».* 2018. Nr 69. pp. 194-196. [in Ukrainian]
6. Yanytska O. M. *Korpusnyi analiz dlia vyznachennia kliuchovykh leksem-nominantiv kontseptu VIRA na materialu anhliiskoi, ukrainskoi ta frantsuzkoi mov.* [Corpus analysis to determine the key lexemes-nominants of the concept of BELIEF on the material of English, Ukrainian and French languages.] *Scientific Bulletin of the International Humanitarian University.* 2016. Nr 25. pp. 160-162 [in Ukrainian]
7. Biber D. *Corpus Linguistics: Investigating language structure and use.* Cambridge University Press, 2004. 300p.