
СОЦІОЛІНГВІСТИКА ЛІНГВОКУЛЬТУРОЛОГІЯ КОГНІТИВНА ЛІНГВІСТИКА

УДК 811.111:81-13

DOI: <https://doi.org/10.18524/2414-0627.2024.31.309443>

Ю. І. Дем'янчук,

канд. екон. наук,

Львівський державний університет безпеки життєдіяльності,

викладач кафедри іноземних мов та перекладознавства

ORCID: <https://orcid.org/0000-0001-8722-5568>

НОВА ПАРАДИГМА КОРПУСНОГО АНАЛІЗУ ГАЗЕТНИХ СТАТЕЙ ВІЙСЬКОВО-ПОЛІТИЧНОГО СПРЯМУВАННЯ

Стаття присвячена вивченню корпусу газетних статей The Guardian, The Washington Post, The New York Times та виявленню сучасних методик корпусного аналізу.

Наголошено, що корпуснокерований підхід є релевантним для удосконалення технології балансування корпусу текстів. Такий підхід вирізняється способами представлення корпусних даних, уніфікацією декодованої інформації. Із цією метою проаналізовано новини, репортажі на міжнародні теми, що висвітлюють політичну стратегію низки країн в контексті російсько-української війни.

Зауважено, що для розуміння сутності визначеної проблематики важливим було окреслити специфіку вивчення мовних патернів воєнно-політичного спрямування у межах корпуснокерованого підходу. У процесі дослідження з'ясовано, що масив різнотипних даних здатний впливати на формування оточення мовних конструкцій, що роз'яснюють ймовірнісні варіації моделей мовної системи у текстових вимірах. У статті розглянуто випадкові лінії конкордансу, виокремлено найчастотніші одиниці публіцистичної системи мови, які розкривають мовні зміни в період війни. Окрім того, у науковій праці визначено кінцеві змінні в статистичних моделях, домінантно-семантичні позиції мовних конструкцій та виявлено їх лінгвістичний потенціал, статус і функції у корпусі. На додачу, обґрунтовано важливість використання технології Text Mining, тесту ANNOVA, програмного забезпечення Cortical.io. під час обробки масиву даних.

Резюмовано, що нова парадигма корпусного аналізу полягає в удосконаленні методики корпусного аналізу — визначенні атрибуції текстів, ієрархічного розподілу, встановлення статистично значущих відмінностей між середніми значеннями змінних методом дисперсійного аналізу.

Результати проведеного дослідження уможливають висновувати, що механізм опрацювання текстів публіцистичного стилю мовлення відіграє важливу роль у розвитку нової парадигми корпусного аналізу.

Ключові слова: корпус текстів, корпуснокерований підхід (corpus-driven approach), корпусний аналіз, The Guardian, The Washington Post, The New York Times, Text Mining, ANNOVA.

Yu. I. Demyanchuk,

PhD in Economics,

Lviv State University of Life Safety,

*Lecturer of English language at the Foreign Languages
and Translation Studies Department*

ORCID: <https://orcid.org/0000-0001-8722-5568>

NEW PARADIGM OF CORPUS ANALYSIS OF MILITARY-POLITICAL NEWSPAPERS ARTICLES

This study using the corpus of newspaper articles from The Guardian, The Washington Post, and The New York Times attempts to explore contemporary methods of corpus analysis. It emphasizes that a corpus-driven approach is relevant for improving the technology of balancing text corpora. This approach is characterized by ways of representing corpus data and unifying decoded information based on the corpus. To this end, urgent news and international reports covering the political strategies of various countries in the context of the Russia-Ukraine war were analyzed. It is noted that understanding the essence of the defined issue required outlining the specifics of studying military-political language patterns within the corpus-driven approach. The research found that a variety of data types can influence the formation of the linguistic environment, which explains variations of language system models in textual dimensions. The results also showed random concordance lines, the semantic field of lexemes, and the most frequent units of the journalistic language system that reveal linguistic changes during the war period.

Additionally, the study identifies final variables in statistical models, dominant-semantic positions of linguistic constructions, the valency of marked units, and their linguistic potential, status, and functions within the corpus. The importance of using Text Mining technology, the ANNOVA test, and Cortical.io software for data processing is also justified. It is concluded that improving corpus analysis methods depends on text attribution, hierarchical clustering, determining statistically significant differences between mean values of variables through analysis of variance. Moreover, the study enable the conclusion that the text processing mechanism plays an important role in the development of new paradigm of corpus analysis.

Key words: text corpus, corpus-driven approach, corpus analysis, The Guardian, The Washington Post, The New York Times, Orange Text Mining, ANNOVA.

Постановка проблеми та обґрунтування актуальності її розгляду. Існує чимало досліджень специфіки корпусної лінгвістики, здебільшого з точки зору інтерференції методів формування корпусів текстів, заразом розроблення комп'ютерних систем для їхнього опрацювання [3]. За допомогою корпусних технологій можна удосконалити ці методи, визначити технології балансування корпусів, різних за обсягом, жанрово-стильовим розмаїттям, контекстуальним наповненням, що вирішить низку завдань, як-от: фіксацію лексичних розбіжностей між різними мовами, формування оточення вищих дистрибутивних конструкцій, з'ясування контексту, що впливає на активізацію семантичного потенціалу одиниць системи мови та окреслення емоційно забарвленої лексики.

Оскільки опрацювання текстів відіграє важливу роль у розвитку нових методів корпусного аналізу, корпусної системи даних, методик лінгвістичних розвідок, питання відбору текстів до корпусу завжди є релевантним. Для оптимізації корпусу і його продуктивного застосування, науковці пропонують досліджувати випадково вибрані лінії конкордансу, контури семантичного поля лексем, спонтанне мовлення на кожному етапі експлуатації програмного забезпечення.

Зауважимо, що обробка корпусних даних це — сукупність операцій (збирання, введення, перетворення, зчитування, зберігання), пов'язаних із дослідженням одиниць надсегментної фонетики, морфологічних норм/помилки під час визначення граматичних особливостей частин мови, синтаксичних одиниць у всіх аспектах їх вияву, системи правил передавання звукової мови на письмі, неморфологічного способу словотворення. Відповідно корпус є цінним джерелом не лише якісної інформації, а й кількісної. Наша прагматична мета це — удосконалення методу корпусного аналізу на основі корпуснокерованого підходу.

Наразі досі не окласифіковано різнотипні дані газетних видань, які здатні впливати на формування оточення мовних конструкцій, не визначено змінні статистичних моделей, що демонструють ймовірнісні варіації моделей мовної системи у текстових вимірах й не виокремлено найчастотніші одиниці, що розкривають мовні зміни в період війни. Це дослідження **актуальне**, через те, що корпуснокерований підхід, на якому ґрунтується обробка газетних статей, заразом сприяє удосконаленню корпусного аналізу забезпечує повноту представлення мовних явищ, уточнення дистрибутивних конструкцій, конкретизацію емоційно забарвленої лексики, визначення атрибуції текстів, що входять до політичної стратегії низки країн у висвітленні російсько-української війни.

Аналіз останніх досліджень і публікацій. Питання корпусного аналізу лексики суспільно-політичного характеру досліджували А. Mohammed та А. Morshid, M. Lórinicz, S. Hadi, H. Triana, T. Reffinaldi. Корпуснокерований підхід у наукових розвідках застосовують R. Lu, A. Mohamadi, A. Che, D. Biber. У працях P. Bański, M. Fischer, E. Frick, E. Ketzan, M. Kupietz, C. Schnober, O. Schonefeld, A. Witt приділено увагу новій платформі корпусного аналізу, зміні перцепції лінгвістичної теорії.

Значний внесок у розвиток корпусної лінгвістики, її теоретичної бази, корпусного аналізу тексту зробили й вітчизняні дослідники, такі як М. Шведова, С. Бук, І. Біскуб, В. Жуковська, О. Таран, О. Демська-Кульчицька, Т. Анохіна та ін.

Формування мети і завдання статті. З урахуванням того, що наукова парадигма корпусної лінгвістики передбачає вивчення корпусів текстів і підходів до їхнього дослідження, **метою** пропонованої праці є запропонувати нову парадигму корпусного аналізу корпусу періодичних видань The Guardian, The Washington Post, The New York Times. Для досягнення поставленої цілі необхідно розв'язати низку **завдань**: 1) укласти корпус, що складається з текстів газетних статей The Guardian, The Washington Post, The New York Times (обсягом 5121 текстових документів), які були написані після 2022 року (термінові новини, репортажі на міжнародні теми), 2) виявити мовні патерни, які найчастіше використовують публіцистичні видання в окресленні воєнно-політичної ситуації в Україні, 3) вибудувати стратегію розвитку корпусного аналізу.

Отже, **об'єктом** аналізу є статті публіцистичних видань, а **предметом** — окреслення специфіки корпуснокерованого підходу, інтелектуального аналізу текстів, дисперсного аналізу до вивчення мовних явищ під впливом російсько-української війни, сконденсовані результати якої виформовують нову парадигму корпусного аналізу.

Виклад основного матеріалу дослідження. Передусім важливо окреслити дефініційне поняття “корпуснокерований підхід”. Цей підхід функціонує досить тривалий час і є основою корпусного аналізу.

За визначенням L. Rong, A. Mohamad, A. An, A. Che це — об'єктивна перспектива, що охоплює нову методологію та повне використання сукупності доказів [6, с. 122]. Формування вибірки корпусу є індуктивним, відбувається внаслідок з'яви лінгвістичних конструктів й відрізняється від стандартної практики лінгвістів [2]. Коли ми говоримо про вибірку колокацій, високочастотну послідовність слів (фіксованих, змінних слотів), ключові слова на основі корпуснокерованого підходу, то слід пам'ятати про системні закономірності одиниць корпусного аналізу й асоціації між семантичними фреймами, що забезпечують повноту представлення мовних явищ. Інакше кажучи, вони розкривають нові закономірності шляхом індуктивного аналізу корпусу [Ibid].

Убачаємо доречним проаналізувати корпус газетних статей, важливими компонентами якого є лінгвістична, культурно-історична, воєнно-політична, економічна вербалізація в публіцистичних дискурсах. Матеріали, зібрані в текстах газетних статей, — це мовні патерни, які найчастіше використовують журналісти у висвітленні воєнно-політичної ситуації в Україні [5]. Неодмінною умовою такого напрямку дослідження є випрацювання методів статистичної лінгвістики, які зазвичай передбачають частотний й дисперсійний аналіз [7]. Наукове обґрунтування інтерпретації корпусних даних орієнтоване на статистичне моделювання і методи машинного навчання [8].

Доцільність створення корпусу на основі газетних статей визначалася такими передумовами:

- виокремлення публіцистичного стилю, що уможливорює всестороннє виявлення дистрибутивних контекстів;

- визначення найвпливовіших інформаційних ресурсів світу, які забезпечують повноту представлення мовних явищ в контексті російсько-української війни;
- класифікація масиву різнотипних даних, які здатні впливати на формування оточення мовних конструкцій;
- удосконалення технології балансування корпусу, що на пряму залежить від атрибуції текстів;
- виокремлення найчастотніших одиниць системи мови та ключових слів, які розкривають мовні зміни в період війни;
- визначення кінцевих змінних в статистичних моделях, що роз'яснюють ймовірнісні-варіації моделей мовної системи у текстових вимірах.

Окрім лінгвістичного аналізу даних щодо вживання мовних конструкцій в корпусі газетних статей, які асоціюються з воєнно-політичними патернами, з допомогою технології Text Mining (програме забезпечення Orange), тесту ANOVA, програмного забезпечення Cortical.io (обробка природньої мови) можливо визначити їхній статус і функцію у вибірці текстів, здійснити ієрархічний розподіл на кластери, визначити *t*-statistic в контексті мовних змін під впливом війни, виявити статистично значущі відмінності між середніми значеннями методом дисперсійного аналізу. Маємо зауважити, що виокремлені мовні конструкції із статистичною значимістю є імпульсом до виявлення контекстуальної подібності кластерних парадигм у площині політичної лінгвістики, території і ресурсів, історії і культури, ідеології, економічної політики.

Аналіз таких структур є надзвичайно корисним для виявлення інформативного потенціалу активної лексики різного частиномовного статусу й компонентного складу, за допомогою якого пасивна лексика продукує оновлене значення в континуумі загального вжитку [4]. Механізм реалізації її домінантно-семантичних позицій в системі мови, різних формах і стилях репрезентує стильові патерни (типові варіації) у корпусі газетних статей. Вони можуть бути підґрунтям розвитку методології формування корпусів. Двоїстий характер корпусу газетних статей The Guardian, The Washington Post, The New York Times обумовлюється як позначенням домінантних довоєнних і воєнних реалій (до і після 24 лютого 2022 року), нового соціально-історичного та культурного середовища, так і необхідністю найменувати їх відповідними маркерами.

Окремий аспект складає інтелектуальний аналіз текстів, пов'язаний із визначенням валентності одиниць військово-політичної терміносистеми, разом визначення їх лінгвістичного потенціалу. Цей процес структурно відбито в тесті ANNOVA, тому залишається простежити мовні зміни, спричинені численною з'явою колокацій в газетних виданнях, які висвітлюють перебіг російсько-української війни.

Зазначимо, що численна з'ява колокацій залежить від контенту їх вияву.

Дисперсійний аналіз допомагає проаналізувати особливості певних груп й цілком здатний істотно змінити уявлення про статистичні змінні, разом варіації на основі них. Наявний масив корпусних даних за період з 2022-2024 рр. уможливує здійснення аналізу мовних конструкцій публіцистичного жанру, перевірку атрибуції текстів, компаративні розвідки, виявлення потенційних моделей частотних колокацій

та їх квантитативного навантаження, з'ясування динаміки мовних змін під впливом російсько-української війни. Ці процеси сприяють виокремленню надійних критеріїв для оцінки прийнятності певних варіацій моделей мовної системи у текстових вимірах. Статистичні відмінності це — вагомні показники, які певним чином тлумачать ймовірнісні варіації, виявлені в корпусі.

Статистична система даних істотно спрощує опрацювання лексем, на першому етапі корпусного аналізу ми здійснюємо пошук слів з низькою/високою частотністю появи в газетних статтях (див. Таблиця 1).

Таблиця 1. Список слів (часткові дані)

Rank	Word list	TF-IDF	Rank	Word list	TF-IDF
1	troops	0.008	18	Kharkiv	0.004
2	ambassador	0.008	19	Ukrenergo	0.001
3	war	0.008	20	countrywide	0.000
4	crises	0.007	21	Moscow	0.001
5	attack	0.007	22	chief	0.000
6	firefighters	0.005	23	disinformation	0.000
7	negotiations	0.003	24	impact	0.000
8	crime	0.003	25	authorities	0.000
9	civilian	0.003	26	reforms	0.000
10	threat	0.002	27	transparency	0.000
11	border	0.002	28	dialogue	0.000
12	Kyiv	0.002	29	deliberation	0.000
13	operation	0.002	30	process	0.000
14	spokesperson	0.001	31	parliament	0.000
15	security	0.001	32	direction	0.000
16	airbase	0.001	33	legislation	0.000
17	law	0.000	34	ombudsman	0.000

Слова з різним коефіцієнтом стабільності ми об'єднали у лексико-семантичну групу на позначення воєнно-політичних понять, що детермінує їх мовний потенціал (важливість) у колекції корпусу. Характерним приміром для екстракції мовних патернів послуговували виокремлені фрагменти текстів у вигляді конкордансу:

*<Fighting has been raging near the city of **Kharkiv** since Russian **troops** crossed the **border** to open a new front on 9 May.>...<...>*

<Negotiations “can’t be done in such a way that one side has gotten more or less,” Usov said. “Each side has to be able to prove they’ve won in some way.”>...<...>

*<The numbers include only those the United Nations has been able to corroborate, and the U.N. acknowledges that the full **civilian** toll is much higher.>...<...>*

<A bitter and bloody **war** in **Ukraine** has devastated the country, further isolated **Russia** from the West and fueled economic insecurity around the world.>...<...>

<The United States advocated a focused assault along that southern axis, but Ukraine's leadership believed its forces had to **attack** at three distinct points along the 600-mile front, southward toward both Melitopol and Berdyansk on the Sea of Azov and east toward the embattled city of Bakhmut.>...<...>

<**Firefighters** work to clear rubble and extinguish a fire by a heavily damaged building in **Kharkiv** in March 2022, and the same building in 2024.>...<...>

У нашій праці спостерігаємо посилення позиції однокомпокоментних лексем та широкі можливості експансії складноструктурованого суспільно-політичного явища війни. Знайдені вихідні величини формують оцінно-композиційне поле повномасштабного вторгнення Росії в Україну. Відтак, терміни з високим інформаційним навантаженням є смислотвірними термінологічними одиницями, що забезпечують онтологічний базис для трансформаційних видозмін. У такому ракурсі рефлексію змін у мові виражаємо у воєнно-політичній метамові, яка набуває особливої популярності в період війни. Саме опертя на корпусно-керований підхід уможлиблює розкрити потенціал закодованих термінологічних одиниць у вхідному тексті.

З огляду на репрезентативність виконання корпусного аналізу інтелектуальний аналіз тексту (в програмному забезпеченні Orange Text Mining та Cortical.io), на нашу думку, відкриває інноваційні перспективи для модернізації методології дослідження в корпусній лінгвістиці. Імплікованим є факт, що загальна дисперсія слів та загальна дисперсія символів формує оточення мовних конструкцій. Перегрупування мовно-стилістичних аспектів формування домінант публіцистичного дискурсу періодичних видань *The Guardian*, *The Washington Post*, *The New York Times* зміщує акценти тенденцій, віддаючи перевагу підкатегорії *російсько-українська війна*. Показовим щодо цього є встановлення ієрархічної кластеризації періодичних видань й відповідно рубрик.

До прикладу, зміну акцентів від розлогих до вузьких тем щодо російсько-українську війну спостерігаємо у наступних примірах (результат квалітативного аналізу висвітлено частково):

До війни 7–11 січня 2022 року:

You Can't Judge a District Just by Looking at It (The New York Times)

How Russia's Military Is Currently Positioned (The New York Times)

The Women Who Had Abortions Before Roe v. Wade (The New York Times)

Після 22-го лютого 2022 року:

Russia's Attacks on Civilian Targets Have Obliterated Everyday Life in Ukraine (The New York Times)

How Kyiv Has Withstood Russia's Attacks (The New York Times)

Russia's Shrinking War (The New York Times)

Виконуючи завдання дослідження за допомогою інструментів Text Mining здійснюємо ієрархічну кластеризацію видань. При цьому кожне видання співвідноситься з тим чи тим кластером (C1– *The Guardian*, C2 — *The Washington Post*, C3 — *The*

New York Times). У такому разі спостерігаємо флуктуацію якісно нового поняття ймовірнісних варіацій моделей мовної системи у публіцистичних статтях. Саме в такий спосіб загальний розподіл дисперсії слів та символів свідчить про більші відхилення значень між C1 та C3 відносно середнього значення розподілу й менші відхилення між випадковими величинами C2 і C3 (див. Рис. 1, Рис. 2).

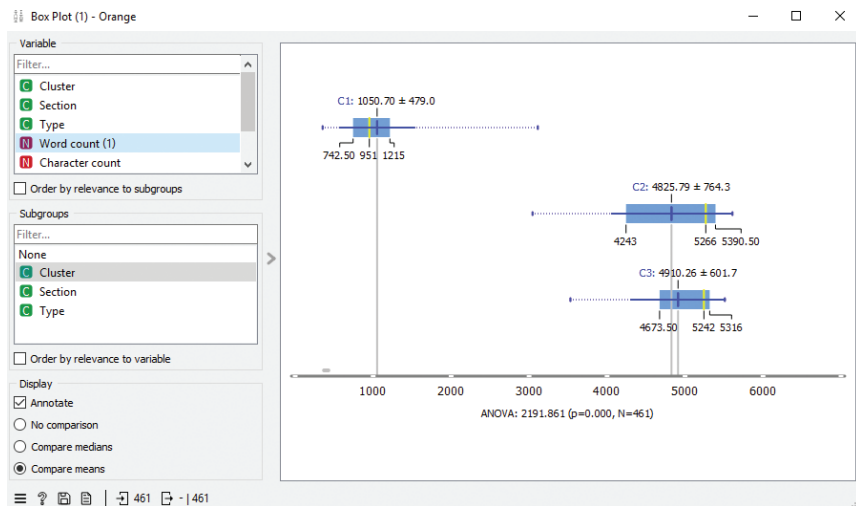


Рис. 1. Розподіл дисперсії слів

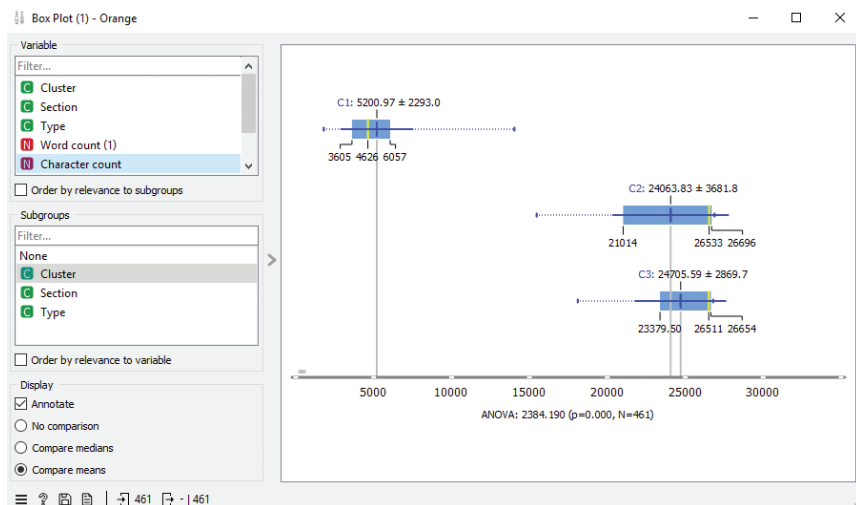


Рис. 2. Розподіл дисперсії символів

Ймовірнісні варіації моделей мовної системи увиразнено за рахунок функціонально-стилістичних елементів у певному контексті публіцистичної статті. Щобільше, розподіл дисперсії слів і символів впливають на рефлексію мовної специфіки військового поняття та відображення коефіцієнту стабільності термінів в корпусі.

Актуалізація експліцитного значення моделей мовної системи в корпусі притаманна передусім субстантивно-ад'єктивній воєнно-політичній термінології, як-от: *air superiority, asymmetric warfare, ballistic missile, civilian casualties, covert operation, cyber attack, defensive strategy, diplomatic mission, economic sanctions* та ін. [1].

Корпусний аналіз продовжує виформовуватись в тесті ANOVA, в якому Word List це список слів з високим, середнім та низьким коефіцієнтом стабільності (див. Таблиця 2).

Таблиця 2. ANOVA для порівняння кількісних показників у вигляді коефіцієнтів

Word Lists (coefficient)	N of documents	N of words in one doc (1000-6000)	N of character in one doc (5000-30000)	Cluster 1	Cluster 2	Cluster 3	Q_1	Q_2	Q_3	p-value
low stability coefficient	461	1 200	5 200	742	4 243	4 673	742	4 243	4 673	0.000
medium stability coefficient	461	5 200	25 500	951	5 266	5 242	951	4 410	4 830	0.000
high stability coefficient	461	5 400	28 000	1 215	5 390	5 316	1 050	4 825	4 910	0.000

З огляду на аналіз результатів побудова корпусу, опрацювання й експлуатація лінгвальних даних, пов'язана з розвитком методології формування корпусів.

У нашій праці ми описуємо такі нові методологічні аспекти, як-от:

- вибір випадкових ліній конкордансу, контурів семантичного поля лексем, спонтанного мовлення;
- класифікація масиву різнотипних даних, які здатні впливати на формування оточення мовних конструкцій;
- виокремлення найчастотніших одиниць системи мови та ключових слів, які розкривають мовні зміни в різні періоди;

- визначення кінцевих змінних в статистичних моделях, що роз'яснюють ймовірнісні-варіації моделей мовної системи у текстових вимірах;
- реалізація домінантно-семантичних позицій мовних конструкцій, різних форм і стилів, що репрезентують типові варіації у корпусі;
- виявлення валентності маркованих одиниць, разом визначення їх лінгвістичного потенціалу, статусу і функції у корпусі;
- ієрархічний розподіл на кластери;
- встановлення статистично значущих відмінностей між середніми значеннями змінних методом дисперсійного аналізу;
- удосконалення технології балансування корпусу, що напряду залежить від атрибуції текстів.

Нова парадигма корпусного аналізу — це симбіоз корпусного, інтелектуального й дисперсійного аналізу текстів для моделювання іманентної реальності мовних одиниць в корпусному континууму.

Дослідження ґрунтоване на корпусних даних, разом даних одержаних внаслідок інтелектуальної і статистичної обробки колекції текстових документів спрямоване на вивчення широких можливостей корпусної та квантитативної лінгвістики з метою випрацювати практичні інструменти для вдосконалення корпусного аналізу.

Засадничими нової парадигми корпусного аналізу є: **побудова збалансовано-го корпусу** статей з The Guardian, The Washington Post і The New York Times, в якому акцентуація спрямована на міжнародні новини, політичні стратегії щодо російсько-української війни; **реалізація корпуснокерованого підходу** для уніфікації та декодування інформації (до прикладу, воєнно-політичних мовних патернів) — сприяє оптимізації корпусних даних; **імплементация Text Mining** в корпусні студії, у процесі параметризованого моделювання вихідні величини інтегрують у ієрархічні кластери; синергетичну цілісність забезпечує **інтеграція методів дисперсії** (ANOVA) до аналізу корпусу текстів, що уможливорює виявити ймовірнісні варіації у мовних конструкціях й визначити їх дистрибуцію. Нова парадигма корпусного аналізу необхідна для визначення як позитивної, так і негативної динаміки мовних змін під час війни.

Отже, корпуснокерований підхід є релевантним для удосконалення корпусного аналізу різних за обсягом текстів, жанрово-стильовим розмаїттям й контекстуальним наповненням. Позитивний результат полягає у визначенні атрибуції текстів на основі технології Orange Text Mining і програмного забезпечення Cortical.io., разом з'ясування

ймовірнісних варіацій моделей мовної системи та відображення коефіцієнту стабільності термінів, ґрунтуючись на результаті тесту ANOVA.

Висновки та перспективи досліджень. Проведене дослідження дає змогу діяти висновку, що механізм опрацювання текстів відіграє важливу роль у розвитку нових методів корпусного аналізу, корпусної системи даних, методик конкретизації лінгвістичних розвідок. Мовні патерни, які найчастіше використовують публіцистичні видання у висвітленні воєнно-політичної ситуації в Україні є цінним джерелом як

якісної, так і кількісної інформації. Удосконалення методу корпусного аналізу на основі корпуснокерованого підходу, інтелектуального і дисперсного аналізу спирається на принцип контекстуального уточнення дистрибутивних конструкцій, які забезпечують повноту представлення мовних явищ. Визначені мовні конструкції здатні впливати на формування оточення усіх одиниць терміносистеми, які, серед іншого, входять до політичної стратегії низки країн у висвітленні російсько-української війни. Перспективи подальшого дослідження полягають в ідентифікації *war-related collocations* в інших типах дискурсу [1], ґрунтуючись на корпуснокерованому підході.

Література

1. Alfattah M., Morshid A. *Methods of translating war-related collocations: A case study of Yemeni undergraduates. Journal of Social Studies.* 2024. Vol. 30. Pp. 129–157.
2. Biber D. *A corpus-driven approach to formulaic language in English. International Journal of Corpus Linguistics.* John Benjamins Publishing Company. 2009. Vol.14. Pp.275–311.
3. Hadi S., Triana H., Tafati W. *Lexical cohesion analysis on articles of the Russia-Ukraine conflict in Arabic and English online newspapers. Studies in English Language and Education.* 2024. Vol. 11(1). Pp. 403–419.
4. Hae-Young K. *Analysis of variance (ANOVA) comparing means of more than two groups. Restorative dentistry & endodontics,* 2014. DOI: 39. 74-7. 10.5395/rde.2014.39.1.74.
5. Lőrincz M. *A corpus analysis of the representation of war in Ukraine in the international press. The Journal of V. N. Karazin Kharkiv National University. Series: Foreign Philology. Methods of Foreign Language Teaching.* 2022. Pp. 61–68.
6. Rong L., Mohamad A., An A., Che A. *A Comparative Study of Corpus-based and Corpus-driven Approaches. Linguistics International Journal.* 2022. Vol. 15. Pp. 119–132.
7. Бук С. *Основи статистичної лінгвістики : навч.-метод. посіб. Львів : Видавничий центр ЛНУ імені Івана Франка,* 2008. 124 с.
8. Таран О., Лебеденко І. *Семантико-дистрибутивний аналіз сленгізмів на основі даних корпусу ГРАК.* 2021. DOI: 21. 119-123. 10.31891/2415-7929-2021-21-22.

References

1. Alfattah, M., Morshid, A. (2024), *Methods of translating war-related collocations: A case study of Yemeni undergraduates. Journal of Social Studies.* Vol. 30, pp. 129–157. [in English].
2. Biber, D. (2009), *A corpus-driven approach to formulaic language in English. International Journal of Corpus Linguistics.* John Benjamins Publishing Company. Vol. 14, pp.275–311. [in English].
3. Hadi, S., Triana, H., Tafati, W. (2024), *Lexical cohesion analysis on articles of the Russia-Ukraine conflict in Arabic and English online newspapers. Studies in English Language and Education,* Vol. 11(1), pp. 403–419. [in English].

4. Hae-Young, K. (2014), *Analysis of variance (ANOVA) comparing means of more than two groups. Restorative dentistry & endodontics*. DOI: 39. 74-7. 10.5395/rde.2014.39.1.74. [in English].
5. Lőrincz, M. (2022), *A corpus analysis of the representation of war in Ukraine in the international press. The Journal of V. N. Karazin Kharkiv National University. Series: Foreign Philology. Methods of Foreign Language Teaching*, pp. 61–68. [in English].
6. Rong, L., Mohamad, A., An, A., Che, A. (2022), *A Comparative Study of Corpus-based and Corpus-driven Approaches. Linguistics International Journal*. Vol. 15, pp. 119–132. [in English].
7. Buk, S. (2008), *Osnovy statystychnoi linhvistyky : navch.-metod. posib. [Fundamentals of statistical linguistics: a textbook]*. Lviv: Vydavnychiy tsestr LNU imeni Ivana Franka. [in Ukrainian].
8. Taran, O., Lebedenko, I. (2021), *Semantyko-dystrybutyvnyi analiz slenhyzmiv na osnovi danykh korpusu HRAK [Slang units semantic and distributive analysis with GRAK corpus data]*. DOI: 21. 119-123. 10.31891/2415-7929-2021-21-22. [in Ukrainian].

* * *