

УДК 811.111'42

DOI <https://doi.org/10.32782/tps2663-4880/2021.19.1.23>

КЛЮЧОВІ СЛОВА АНГЛІЙСЬКОМОВНОГО НОВИННОГО ДИСКУРСУ У 2021 РОЦІ: КОРПУСНИЙ ПІДХІД

KEYWORDS OF THE ENGLISH ONLINE NEWS DISCOURSE IN 2021: A CORPUS APPROACH

Ділай М.П.,

orcid.org/0000-0001-5182-9220

кандидат філологічних наук, доцент,

доцент кафедри прикладної лінгвістики

Національного університету «Львівська політехніка»

Статтю присвячено розгляду сучасного англійськомовного новинного дискурсу з позицій корпусної лінгвістики. Корпусні дані та корпусні процедури проливають світло на мову в її реальному вживанні, тобто розкривають лінгвістичні патерни, які дозволяють зрозуміти, якими способами мова використовується для побудови дискурсів. Мета цієї праці – здійснити корпусний аналіз ключових слів англійськомовного новинного онлайн-дискурсу у 2021 році, використовуючи дані та інструменти корпусу NOW (News on the Web). Дослідження дасть змогу зробити статистично обґрунтовані висновки про тематику та основні концептуальні центри англійськомовних новин, а отже, виявити ключові концептуальні сфери, які є значимими сьогодні в суспільстві. У межах досліджень медійного дискурсу аналіз ключових слів є важливим для вивчення та інтерпретації ідеологічного значення, спрямованого на цільову аудиторію.

Ключові слова трактуємо як слова, які виділяються своєю частотністю в тексті порівняно з іншим референтним текстом/корпусом. Застосовано корпусокеровану методологію (corpus-driven methodology) вилучення ключових слів із текстів новин, яка базується на чітких статистичних мірах без жодних заздалегідь сформованих гіпотез. Для пошуку ключових слів у корпусі NOW створено три дослідницькі віртуальні корпуси англійськомовних новин 2021 року (період 01/01/2021 – 30/09/2021), а саме загальний корпус новин різних англійськомовних країн (2021_news), корпус новин США (2021_US) та корпус новин Великобританії (2021_GB). Автоматично визначено і проаналізовано списки специфічних ключових іменників у створених корпусах. Застосовуючи семантичний та контекстуальний аналізи, встановлено основні концептуальні домени найчастотніших ключових іменників (PANDEMIC, MEDIA, POWER, ELECTIONS, WORK, BUSINESS AND INDUSTRY, LAW AND ORDER, STATE AND PEOPLE, PROBLEMS). Здійснено статистичний аналіз колокацій та кластерів ключового слова *vaccine* у новинах США та Великобританії 2021 року, розглянуто особливості семантичної преференції та дискурсивної просодії. Разом із тим показано словотвірну продуктивність *vaccine*.

Ключові слова: англійськомовний новинний дискурс, корпусна лінгвістика, ключові слова, NOW корпус, колокації, словотвірна продуктивність.

The paper addresses modern English news discourse from the standpoint of corpus linguistics. Corpus data and corpus procedures shed light on real life language use, that is, they reveal linguistic patterns that allow us to understand the ways in which language is used to construct discourses. The purpose of this study is to carry out a corpus analysis

of the keywords of the English online discourse in 2021, using the data and tools of the NOW (News on the Web) corpus. The research will make it possible to draw statistically sound conclusions concerning the aboutness and the main conceptual centers of English news, and thus identify key conceptual domains that are significant in society today. Within media discourse research, keyword analysis is important for the study and interpretation of ideological meaning aimed at the target audience.

Keywords are interpreted as words that stand out for their frequency in the text compared to other reference text/corpus. A corpus-driven methodology is used to extract keywords from news texts. It is based on clear statistical measures without any pre-formed hypotheses. In order to identify the keywords, three virtual corpora of English news of 2021 were created in the NOW corpus (period 01/01/2021 – 30/09/2021), namely the general corpus of news of different English-speaking countries (2021_news), the corpus of the US news (2021_US) and the corpus of the UK News (2021_GB). Lists of specific key nouns of the created corpora have been automatically determined and analyzed. Using semantic and contextual analysis, the main conceptual domains of the most frequent keywords have been established (PANDEMIC, MEDIA, POWER, ELECTIONS, WORK, BUSINESS AND INDUSTRY, LAW AND ORDER, STATE AND PEOPLE, PROBLEMS). The statistical analyses of collocations and clusters of the keyword *vaccine* in the news of the US and Great Britain in 2021 have been carried out and specificity of semantic preference and discursive prosody has been considered. In addition, the word-building productivity of *vaccine* has been described.

Key words: English news discourse, corpus linguistics, keywords, NOW corpus, collocations, word-building productivity.

Постановка проблеми. Новинний дискурс розглядають як форму масової комунікації, яка публічно доступна та представлена періодично або регулярно за допомогою технологій та медіа [1, с. 587]. Це дискурс, з яким ми щодня стикаємося у випусках новин, програмах новин, на новинних веб-сайтах, або в газетах, який містить повідомлення про важливі поточні події та проблеми [2] або, за словами Т. ван Дейка, в якій нова інформація подана про нещодавні події [3, с. 4]. Важливим і актуальним залишається дослідження способів подання інформації в новинах, того, як власне фреймуються події в текстах новин і як це впливає на їхню інтерпретацію, формування в адресатів певного ставлення до прочитаного/почутого, поведінки. Дослідженням влади та ідеології в новинному дискурсі займається критичний аналіз дискурсу (Т. ван Дейк, Р. Фаулер, Н. Ферклафа). У зв'язку з розвитком інформаційних технологій, вільним доступом до Інтернету та його масовим користуванням онлайн-новини стали основним засобом ідеологічного впливу.

Корпусна лінгвістика є сьогодні однією з основних методологічних платформ досліджень дискурсу. Корпусні дані й корпусні процеси (комп'ютерні процедури маніпулювання даними в різні способи) проливають світло на мову в її реальному вживанні (дискурс), тобто розкривають лінгвістичні патерни, які дозволяють зрозуміти, якими способами мова використовується в побудові дискурсів (або якими способами конструюється реальність) [4]. Сьогодні за допомогою репрезентативних, великих за обсягом електронних корпусів текстів можна отримати інформацію про динаміку лексичного складу мови, функціонування граматичних та синтаксичних одиниць, здійснювати аналіз вживання мовних одиниць у різних жанрах, за різний період, у різних авторів тощо. У межах критичного ана-

лізу корпусні дослідження дискурсу спрямовані на виявлення певних закономірностей використання мови в тексті на тлі соціального контексту, а критичний характер аналізу спрямований на «розкриття і піддавання сумніву усталених припущень про мову і соціальні феномени, а також визнання дискурсу як потенційно потужного чинника соціальних змін» [5, с. 55].

Явище ключовості (keyness) пов'язують з явищем тематики тексту (aboutness), тобто розумінням основних концептів, тем та поглядів, які обговорюються в тексті чи корпусі [6, с. 7–10, 26, 53–54]. Відповідно, аналіз ключових слів розглядають як спосіб установаження тематики тексту (aboutness) [7, с. 71]. Аналіз ключових слів базується на тому, що повторювані (рекурентні) способи представлення концептів та ідей дозволяють виявити, як ми сприймаємо соціальний світ [8, с. 597]. У межах досліджень медійного дискурсу аналіз ключових слів є важливим для вивчення та інтерпретації ідеологічного значення (ideological meaning), спрямованого на цільову аудиторію [9, с. 565]. Як зазначає Т. ван Дейк, вибір лексики є надзвичайно важливим аспектом новинного дискурсу, саме в лексиці втілюються приховані думки та ідеології [3, с. 176].

Застосування сучасних методик та інструментів корпусної лінгвістики дозволяють сформулювати якісні висновки про використання лексики в сучасному новинному дискурсі, базуючись на великому за обсягом текстовому матеріалі та статистичних даних.

Аналіз останніх досліджень і публікацій. У межах корпусної лінгвістики **ключове слово** (keyword) трактують як слово, яке виділяється своєю частотністю в тексті (у корпусі) порівняно з іншим корпусом [7; 10]. Тобто ключові слова – це слова, які є статистично більш значущі (salient) у тексті або певних текстах, ніж у великому рефе-

рентному корпусі. Для визначення ключових слів у корпусі застосовуються спеціальні статистичні міри (наприклад, *log likelihood value*). Відповідно, такого типу ключові слова отримали назву **корпусозіставні статистичні ключові слова** (*corpus-comparative statistical keywords*) [9, с. 566]. Ключовість (*keyness*) розглядають як міру частотності, з якою слово трапляється нерівномірно в певному типі тексту, і, як правило, визначають, порівнюючи відносну частотність слова у фокусному корпусі (*focus corpus*) та референтному корпусі (*reference corpus*), використовуючи статистичні тести. У межах корпусної лінгвістики розглядають два типи аналізу ключовості: експланаторний та фокусований [11]. Задля досягнення цілей корпусобазованого дискурсного дослідження варто застосовувати експланаторний аналіз ключовості, оскільки він ґрунтується на аналізі конкордансу, який показує контекст лексичних одиниць. Фокусований аналіз ключовості передбачає порівняння нормалізованої частотності певних одиниць у двох корпусах задля вирішення певної дослідницької мети. Сьогодні існує низка корпусних менеджерів, які дають змогу автоматично видобути ключові слова в тексті (найпопулярніші це AntConc, WordSmith, SketchEngine, UAM, LancesBox).

Постановка завдання. Мета статті – застосувати корпусний підхід для визначення ключових слів англійськомовного новинного дискурсу у 2021 році та дослідити їхні семантичні, контекстуальні особливості та прагматичне навантаження. Завдання цього дослідження: автоматизоване визначення ключових слів англійськомовних новин за 2021 рік з використанням корпусу онлайн-новин NOW; встановлення основних концептуальних доменів ключових слів; аналіз колокацій та кластерів ключового слова *vaccine* у новинах США та Великобританії; аналіз словотвірної продуктивності ключового слова *vaccine*.

Виклад основного матеріалу. Для встановлення ключових слів англійськомовного новинного дискурсу у 2021 році в цьому дослідженні використано дані та інструменти корпусу текстів англійськомовних онлайн-новин NOW (*News on the Web*) [12]. Корпус NOW доступний на сайті <https://www.corpusdata.org/> [13]. Він містить понад 13,6 мільярдів слів – даних з інтернет-газет та журналів з 2010 року і дотепер. Корпус поповнюється приблизно на 180–200 млн слів кожного місяця. Як зазначають розробники, корпус NOW є структурованим корпусом, який показує, що реально відбувається в мові на даний момент [12]. Функціонал корпусу дає можливість здійснювати

порівняння за датами та країнами, а також у віртуальних корпусах, які дозволяють створювати спеціальні колекції текстів відповідно до вибраного реєстру, веб-сайту, слова. Є можливість здійснювати дослідження варіативності англійської мови.

У цій праці застосовано корпусокеровану методологію (*corpus-driven methodology*), тобто представлено спробу вилучення ключових слів із текстів новин, базуючись на чітких статистичних мірах без жодних заздалегідь сформованих гіпотез. Така методологія вимагає щонайменше два порівняльних корпуси. Один – це дослідницький, або фокусний корпус (*study (focus) corpus*), тобто корпус, який досліджуємо. Інший – це референтний корпус (*reference corpus*), з яким ми порівнюємо дослідницький корпус для виявлення значущих ознак (*salient features*). Для пошуку ключових слів створено три дослідницькі віртуальні корпуси англійськомовних новин 2021 року (період 01/01/2021 – 30/09/2021), а саме загальний корпус новин різних англійськомовних країн (2021_news), корпус новин США (2021_US) та корпус новин Великобританії (2021_GB). Кожен містить приблизно 1000 текстів. Референтним корпусом є весь корпус NOW.

Засоби корпусу NOW дозволяють автоматично видобути ключові іменники, дієслова, прикметники, прислівники, словосполучення іменники+іменник та прикметник+іменник. Можемо отримати два типи ключових слів – частотні та специфічні (*FREQ, SPECIFIC*). Специфічні ключові слова – це слова, які є більш частотні у створеному віртуальному корпусі в порівнянні з усім корпусом новин. Так, у таблиці 1 зображено найчастотніші специфічні ключові іменники у віртуальному корпусі 2021_news (обсяг корпусу: 697 884 слів, 966 текстів; 5,1% від загального); кількість токенів для слова і текстів; на скільки більш частотним є слово, ніж очікувано (формула підрахунку подібна до *Log Likelihood*); загальна кількість токенів слова у всьому корпусі й очікувана кількість токенів у віртуальному корпусі.

У результаті семантичного та контекстуального аналізу виокремлено концептуальні домени найчастотніших специфічних ключових слів: PANDEMIC (*coronavirus, vaccine, virus, (quarantine) facility, test, case, health, death*); MEDIA (*playlist, platform, material, content, ad, statement, news, media, comment, fact, information, report, story, video*); POWER (*governor, administration, president, government, impact, leader, force, power*); ELECTIONS (*election, president*); WORK (*worker, staff, work, job*); BUSINESS AND INDUSTRY (*money, company, market, service, business,*

Таблиця 1

Найчастотніші специфічні ключові іменники у віртуальному корпусі 2021_news

WORD (CLICK FOR CONTEXT)	FREQ	#TEXTS	+		ENTIRE CORPUS	EXPECTED
			SPECIFIC			
			FREQ			
			90	96		
			TEXTS			
CORONAVIRUS	168	98	3.9		846	43.4
VACCINE	710	102	0.3		53,393	2,739.9
VIRUS	193	106	0.0		79,316	4,070.1
GOVERNOR	314	99	0.0		198,636	10,193.0
ADMINISTRATION	222	100	0.0		239,087	12,268.7
FACILITY	274	119	0.0		345,113	17,709.5
ELECTION	364	120	0.0		485,904	24,934.2
RIGHT	764	504	0.0		1,050,410	53,901.8
START	188	127	0.0		261,577	13,422.8
WORKER	366	124	0.0		527,909	27,089.7
TEST	327	131	0.0		477,415	24,498.6
SECURITY	524	186	0.0		790,49	40,564.0
STATEMENT	431	235	0.0		651,981	33,456.4
CASE	864	232	0.0		1,332,000	68,351.6
INVESTIGATION	233	115	0.0		361,629	18,557.0

industry); LAW AND ORDER ((human) rights, permission, security, case, investigation, incident, officer; control, rule, law, police, court, charge); STATE AND PEOPLE (resident, town, community, state, person, country, public, region, nation); PROBLEMS (challenge, effort, concern, risk, support, matter; situation, issue, problem).

У таблиці 2 подано найчастотніші специфічні ключові іменники у створених віртуальних корпусах англійськомовних новин. Результати досить прогнозовані та свідчать, що центральною тематикою новин у 2021 році залишається пандемія (лексми, які реалізують цей концепт, є най-

частотнішими специфічними ключовими іменниками). З-поміж ключових іменників корпусу новин Великобританії спостерігаємо високу частотність одиниць, які реалізують значення заходів протидії пандемії. Окрім характерного для всіх корпусів іменника *vaccine*, у списку є *lockdown*, *restriction*, *recommendation*, *measure*. У списку ключових іменників у корпусі новин США високі позиції займають одиниці концептуального поля POWER, ELECTIONS, LAW AND ORDER.

Варто зазначити, що слово *coronavirus* менш статистично помітне в корпусі новин 2021 року, ніж у корпусі за 2020 рік, і впродовж 2021 року

Таблиця 2

Найчастотніші специфічні ключові іменники у створених віртуальних корпусах

2021_news	2021_US	2021_GB
CORONAVIRUS	CORONAVIRUS	CORONAVIRUS
VACCINE	VACCINE	LOCKDOWN
VIRUS	GOVERNOR	VACCINE
GOVERNOR	ADMINISTRATION	RESTRICTION
ADMINISTRATION	SECURITY	RECOMMENDATION
FACILITY	STATEMENT	VIRUS
ELECTION	ELECTION	VALUE
RIGHT	RIGHT	RISK
START	BILLION	RATE
WORKER	STATE	IMPACT
TEST	INVESTIGATION	AMOUNT
SECURITY	VEHICLE	MEASURE
STATEMENT	MEETING	MILLION
CASE	PERCENT	INVESTMENT

частотність його вживання поступово зменшувалася (рис. 1).

Статистично більш значущим у новинах 2021 року є іменник *vaccine* (рис. 2). З розподілу частотності за країнами видно, що в новинах США він з'являється більш ніж удвічі частіше, ніж у новинах Великобританії (рис. 3).

Аналіз ключових слів передбачає дослідження колокацій (семантичної сполучуваності), передусім особливостей семантичної преференції (сполучуваність лексичної одиниці з одиницями певного лексико-семантичного поля) та дискур-

сивної просодії (сполучуваність лексичної одиниці з одиницями, які мають певну конотацію [14, с. 141, 144–145]. Розгляд патернів колокацій (collocation patterns) дозволяє виявити систематичні семантичні асоціації [15], які пов'язані з такими поняттями критичного аналізу дискурсу, як топос [16, с. 74–76] та топік [17, с. 129–130]. Ці патерни дають змогу виявити елементи, які зумовлюють асоціації, пов'язані зі словом.

Найчастіше слово *vaccine* зустрічається в корпусі новин NOW за 2021 рік разом зі словом *covid-19* (межі пошуку – 4 одиниці зліва і 4 справа).

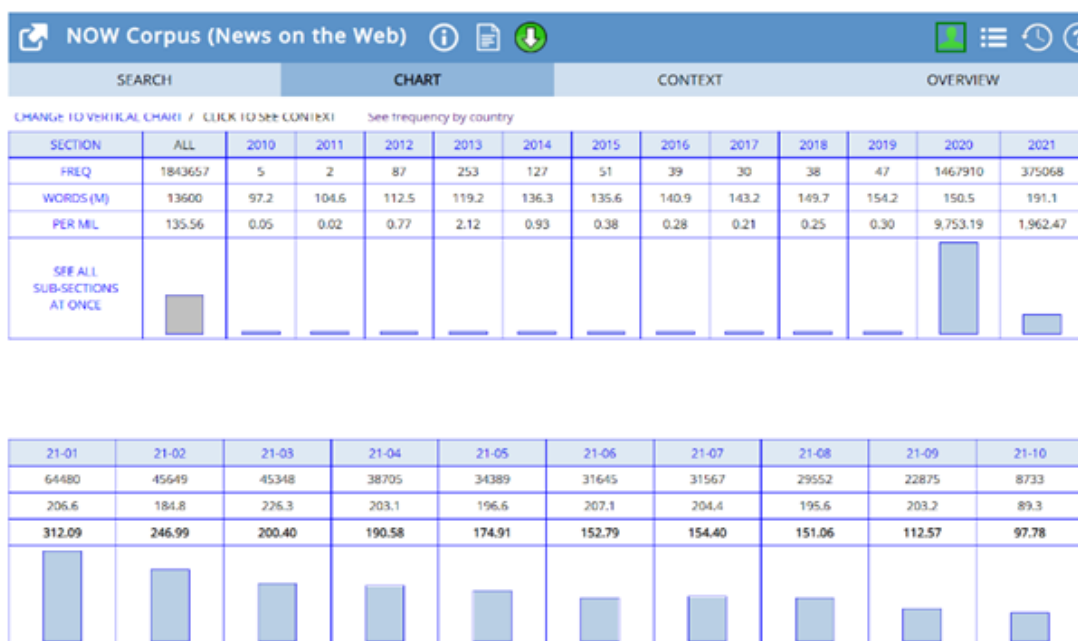


Рис. 1. Слово *coronavirus* у корпусі NOW

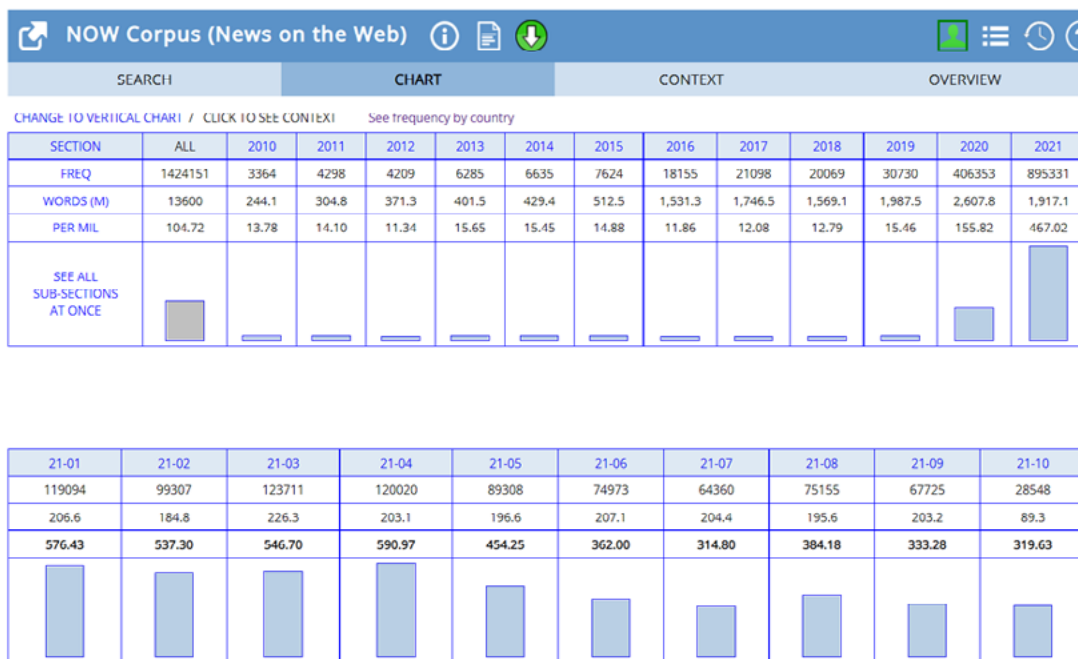


Рис. 2. Слово *vaccine* у корпусі NOW, частотність за роками

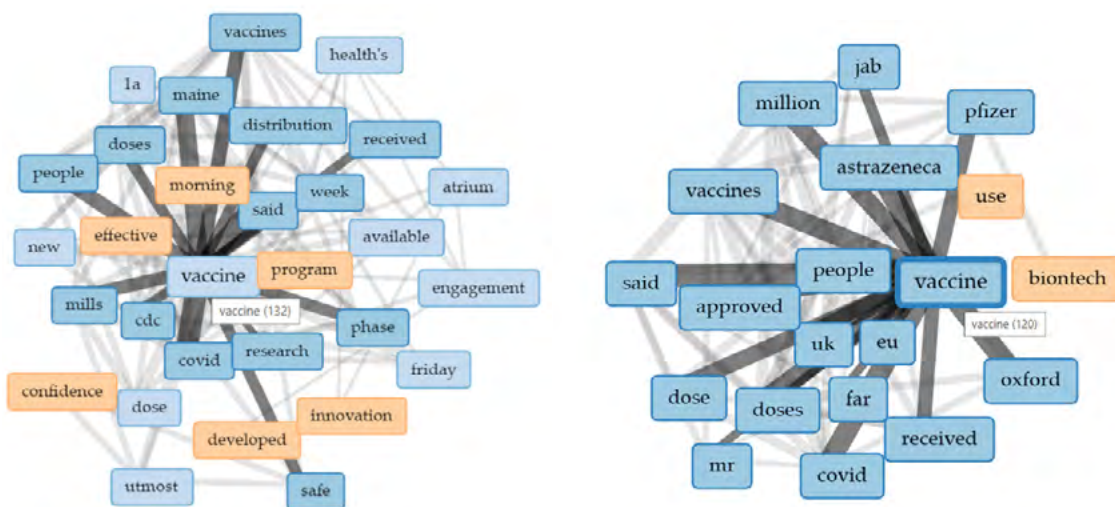
SECTION	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT (SEE ALL)
United States	732998	4,581.3	160.00	
Canada	108192	1,546.4	69.96	
Great Britain	92855	1,558.3	59.59	
Ireland	51923	968.1	53.64	
Australia	59951	955.9	62.71	
New Zealand	25789	500.5	51.53	
India	109748	1,351.6	81.20	
Sri Lanka	10647	105.2	101.17	
Pakistan	18756	332.3	56.45	
Bangladesh	15869	80.1	198.06	
Malaysia	28279	330.2	85.65	
Singapore	26356	515.0	51.18	
Philippines	35547	397.8	89.36	
Hong Kong	6965	66.5	104.74	
South Africa	37999	692.6	54.86	
Nigeria	29859	676.0	44.17	
Ghana	4853	155.0	31.30	
Kenya	12343	233.4	52.89	

Рис. 3. Слово *vaccine* у корпусі NOW, частотність за країнами

До списку найчастотніших увійшли також колокати: *is, doses, dose, have, be, get, AstraZeneca, will, Johnson, has, rollout, Pfizer, are, said, coronavirus, not, was, first, received, covid*. На рисунках 4 за допомогою програми *Voyant tools* [18] візуалізовано зв'язки між найчастотнішими колокатами *vaccine* (а також інтерколокацію колокатів) у віртуальних корпусах новин США та Великобританії за 2021 рік. Цікаво, що в новинах США більш помітними є колокати *distribution, research, safe*, у новинах Великобританії – *approved* та конкретні назви вакцин від Covid-19.

Семантична чи дискурсна просодії можуть допомогти сформувати топос (певні висно-

вки без прямої аргументації), оскільки адресат більш схильний звертати увагу на конотативний чи оцінний рівень специфічних частотних колокацій або, за словами К. Габріелатос, прийняти метафору, а не критично аналізувати її [15]. Загалом, у корпусі новин спостерігаємо позитивну семантичну/дискурсну просодію колокацій з компонентом *vaccine* (наприклад, *effective vaccine, safe vaccine, approved vaccine*). Вочевидь, це прагматично зумовлено наміром викликати позитивні асоціації, сприяти підвищенню загального рівня схвалення вакцин та бажання вакцинуватися, тобто переконати вакцинуватися.

Рис. 4. Колокати *vaccine* у віртуальних корпусах новин США (зліва) та Великобританії (справа) за 2021 рік

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS) OR NUMBER (SPECIFIED SECTION)

[HELP...]

SEC 1 (United States): 4,586,502,035 WORDS

SEC 2 (Great Britain): 1,558,958,869 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO		WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	OHIO	319	1	0.1	0.0	108.4	1	GP	47	7	0.0	0.0	19.8
2	DEC	270	1	0.1	0.0	91.8	2	SCOTTISH	33	5	0.0	0.0	19.4
3	ALLOTMENT	232	1	0.1	0.0	78.9	3	NHS	359	57	0.2	0.0	18.5
4	PRIORITIZED	229	1	0.0	0.0	77.8	4	MHRA	41	7	0.0	0.0	17.2
5	GOV	433	2	0.1	0.0	73.6	5	BORIS	73	13	0.0	0.0	16.5
6	AD	208	1	0.0	0.0	70.7	6	PRIORITISED	72	13	0.0	0.0	16.3
7	HOTLINE	203	1	0.0	0.0	69.0	7	RECOGNISED	25	5	0.0	0.0	14.7
8	SCHEDULING	394	2	0.1	0.0	67.0	8	WALES	101	22	0.1	0.0	13.5
9	CVS	191	1	0.0	0.0	64.9	9	HANCOCK	99	22	0.1	0.0	13.2
10	TRACKING	545	3	0.1	0.0	61.7	10	NIGHTCLUBS	52	13	0.0	0.0	11.8
11	SUBMIT	181	1	0.0	0.0	61.5	11	SIR	27	7	0.0	0.0	11.3
12	STATEWIDE	356	2	0.1	0.0	60.5	12	BREXIT	15	4	0.0	0.0	11.0
13	EDUCATORS	178	1	0.0	0.0	60.5	13	HOSPITALISATIONS	18	5	0.0	0.0	10.6
14	P.M	178	1	0.0	0.0	60.5	14	SCOTLAND	114	33	0.1	0.0	10.2
15	JAN	323	2	0.1	0.0	54.9	15	PROGRAMME	1433	419	0.9	0.1	10.1
16	FDA-APPROVED	155	1	0.0	0.0	52.7	16	SKY	33	10	0.0	0.0	9.7
17	SYRINGES	153	1	0.0	0.0	52.0	17	MINISTERS	49	15	0.0	0.0	9.6
18	AUG	149	1	0.0	0.0	50.6	18	WHILST	26	8	0.0	0.0	9.6
19	ORGANIZATIONS	141	1	0.0	0.0	47.9	19	MID-SEPTEMBER	61	19	0.0	0.0	9.4
20	JERSEY	280	2	0.1	0.0	47.6	20	GPS	32	10	0.0	0.0	9.4
21	FRANCISCO	133	1	0.0	0.0	45.2	21	FANTASTIC	22	7	0.0	0.0	9.2

Рис. 5. Специфічні колокати vaccine у новинах США, які не є типовими для новин Великобританії, і навпаки (період січень–вересень 2021 року)

Встановлено специфічні колокати *vaccine* у новинах США, які не є типовими для новин Великобританії, і навпаки (рис. 5).

У списку колокатів, звісно, містяться специфічні топоніми та антропоніми (передусім імена глав держав та чиновників, відповідальних за вирішення питань, пов'язаних з охороною здоров'я та реалізацію програми вакцинації). Семантичний та контекстуальний аналіз специфічних колокатів показав, що в американських новинах більшу увагу приділено рекламі та агітації, ведеться потужна інформаційна кампанія популяризації вакцинації, про що свідчить використання колокатів *ad*, *statewide*, разом із тим важливим є контроль за процесом вакцинації (*Gov*,

hotline, *CVS*, *tracking*, *scheduling*, *organizations*, *serve*, *box*, *database*, *map*, *Czar*, *underserved*), її розподілом (*allotment*, *prioritized*, *educators*, *allotments*, *exemption*) та якістю вакцини (*FDA-approved*, *endorsed*). Корпусні дані засвідчили також відмінності в орфографії американського та британського варіантів англійської мови (*prioritized (USE) :: prioritised (BE)*). У корпусі британських новин специфічні колокати *vaccine*, які вказують на контроль за здійсненням вакцинації, це *GP*, *NHS*, *MHRA*, *programme*, *ministers*, *roadmap*, *taskforce*; зустрічаємо конкретні назви вакцин *Oxford/Astrazeneca*, *Biontech/Pfizer*. Слово *vaccine* вживається у контексті з прикметником позитивної семантики *recognised*; прикмет-

SEC 1 (United States): 4,586,502,035 WORDS

SEC 2 (Great Britain): 1,558,958,869 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO		WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	PROVIDERS	1207	6	0.3	0.0	68.4	1	DIVIDEND	39	3	0.0	0.0	38.2
2	WEBSITE	148	1	0.0	0.0	50.3	2	SHIELD	11	1	0.0	0.0	32.4
3	ALLOTMENT	130	1	0.0	0.0	44.2	3	MEASURED	12	2	0.0	0.0	17.7
4	ELIGIBILITY	1818	14	0.4	0.0	44.1	4	CENTRE	173	43	0.1	0.0	11.8
5	PROTOCOLS	109	1	0.0	0.0	37.0	5	PROGRAMME	1173	308	0.8	0.1	11.2
6	INFO	108	1	0.0	0.0	36.7	6	FIGURES	38	11	0.0	0.0	10.2
7	REGIMEN	194	2	0.0	0.0	33.0	7	BUS	54	16	0.0	0.0	9.9
8	DASHBOARD	192	2	0.0	0.0	32.6	8	PATCH	10	3	0.0	0.0	9.8
9	SERIES	479	5	0.1	0.0	32.6	9	BOUNCE	52	16	0.0	0.0	9.6
10	CZAR	94	1	0.0	0.0	32.0	10	TASKFORCE	124	42	0.1	0.0	8.7
11	AUTHORIZED	184	2	0.0	0.0	31.3	11	PROGRAMMES	88	30	0.1	0.0	8.6
12	ALLOTMENTS	91	1	0.0	0.0	30.9	12	SCEPTICISM	40	16	0.0	0.0	7.4
13	SELPIES	90	1	0.0	0.0	30.6	13	CENTRES	107	43	0.1	0.0	7.3
14	AWARENESS	89	1	0.0	0.0	30.3	14	PRIORITISATION	11	5	0.0	0.0	6.5
15	CREDENTIALS	78	1	0.0	0.0	26.5	15	PRODUCING	24	11	0.0	0.0	6.4
16	OFFICIAL	77	1	0.0	0.0	26.2	16	TAKE-UP	75	37	0.0	0.0	6.0
17	RESPONSES	76	1	0.0	0.0	25.8	17	ROW	22	11	0.0	0.0	5.9

Рис. 6. Специфічні словосполучення vaccine+* (R1) у новинах США в порівнянні з новинами Великобританії (період січень–вересень 2021 року)

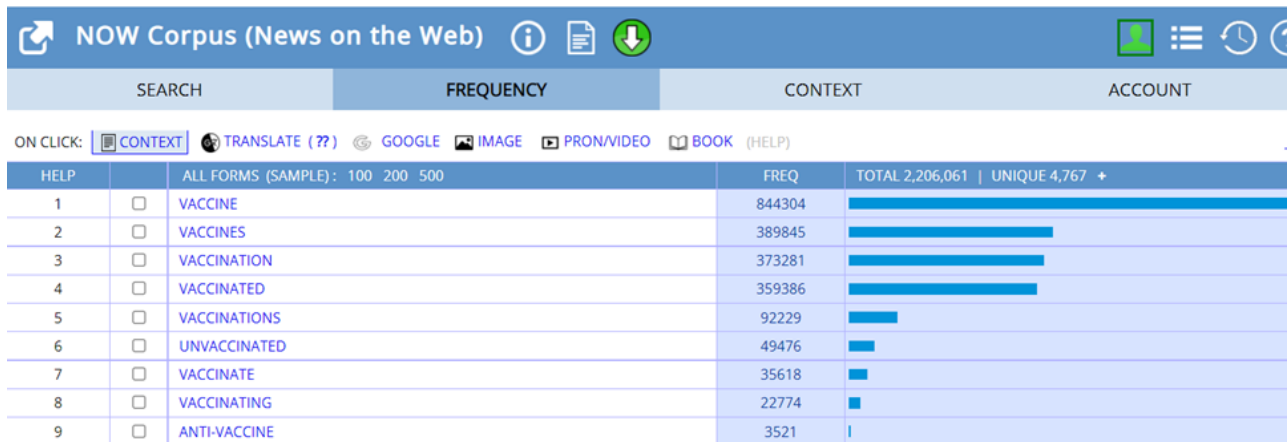


Рис. 7. Результати пошуку *vaccin*

никами, що позначають вищу міру позитивної суб'єктивної оцінки *phenomenal, fantastic*. На відміну від новин США, колокатами виступають *nightclubs* (у контексті документів про вакцинацію, які потрібні для відвідування, наприклад: *mandating mask use and vaccine passports for nightclubs; two doses of vaccine to enter nightclubs and other large events; introduce vaccine passports for nightclubs and large-scale gatherings*), *Brexit* (наприклад: *grateful to Johnson for Brexit and for the vaccine; hopefully with the vaccine programme Brexit is now in the rearview mirror; the obvious benefits of Brexit in enabling the vaccine development and rollout*), *hospitalisations* (у контекстах зменшення госпіталізації завдяки вакцині, наприклад: *lower hospitalisations due to the vaccine rollout, the vaccine will slow hospitalisations sufficiently, vaccine is preventing hospitalisations*).

Іншим способом дослідження значення слова в корпусі є дослідження кластерів (елементів) навколо або біля нього, послідовного групування слів. На рис. 6 показано специфічні словосполучення *vaccine+** (R1) для новин США в порівнянні з новинами Великобританії (період січень–вересень 2021 року), і навпаки.

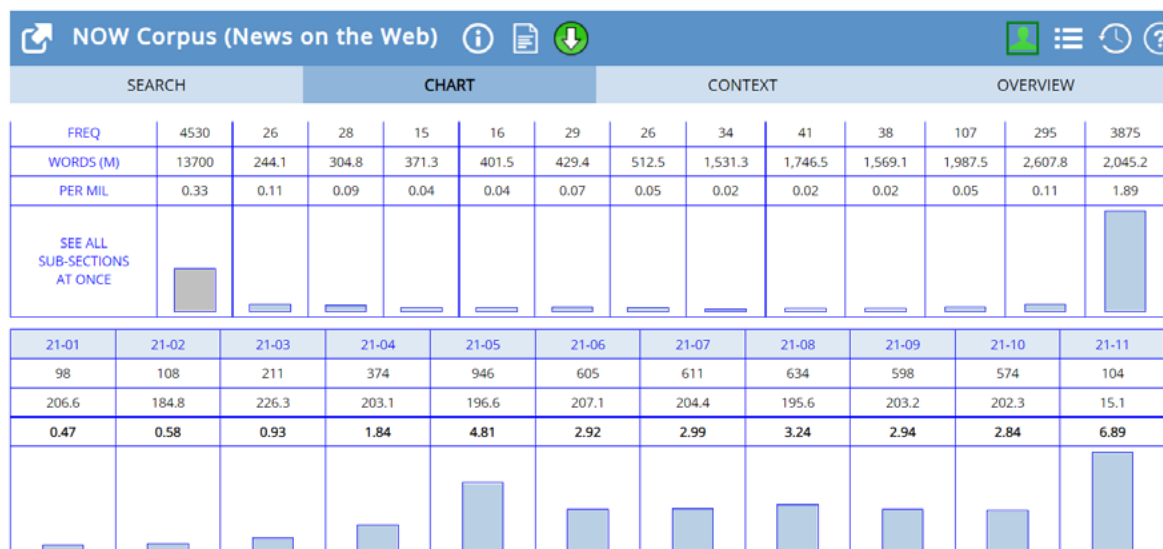
Специфічні вислови називають нові реалії своєї країни. У новинах США частотним є вислів *vaccine provider*, яким називають, зокрема, лікарні, медичні установи, відділи, центри, аптеки, які мають право здійснювати вакцинацію; вислів *vaccine credentials* вживається для позначення цифрових документів для підтвердження вакцинації. Цікавим є вислів *vaccine dividend*, який є типовим для британських новин і позначає додаткові свободи для вакцинованих (також відомий як *vaccine bonus*), наприклад: *opportunity to take advantage of the UK's world-leading vaccine dividend as countries across the globe begin to open up for international travel; The plan centres on*

a "vaccine dividend", which will allow vaccinated people to meet indoors at certain stages; to deliver the vaccine dividend for millions more people, for friends and family to reconnect.

Дослідивши форми слова *vaccine* у корпусі, виявлено розгортання словотвірної продуктивності слова (частково результати пошуку *vaccin* зображено на рис. 7).

Більшість похідних та складних слів, утворених від *vaccine*, позначають процес вакцинації, зокрема, його учасників, виробництво та розподіл вакцини (*vaccination, vaccinated, vaccinate, vaccinators, fully-vaccinated, vaccine-sharing, vaccinator, vaccinees, vaccinology, mass-vaccination, double-vaccinated, vaccinologist, vaccine-eligible, vaccine-producing, vaccine-makers, vaccine-maker, pre-vaccination, revaccination, pre-vaccine, vaccine-making, covid-vaccine, well-vaccinated, vaccine-distribution, double-vaccination, full-vaccination, covid-vaccinated, revaccinated, vaccine-manufacturing, re-vaccinated, revaccinate*).

Значна кількість похідних та складних слів мають негативне значення і вживаються передусім для позначення невакцинованих (*unvaccinated, non-vaccinated, un-vaccinated, nonvaccinated*), противників вакцинації та протестів проти примусової вакцинації (*anti-vaccine, anti-vaccination, vaccine-hesitant, vaccine-resistant, vaccine-skeptical, non-vaccination, antivaccine, vaccine-evading, vaccine-hesitancy, non-vaccine*), низького рівня вакцинації (*undervaccinated, least-vaccinated, under-vaccinated, low-vaccination, half-vaccinated, less-vaccinated*). Найбільш продуктивними є негативні префікси: *un-, non-, anti-*. Зустрічається використання префікса *pro-* у словах, які виражають підтримку вакцинації (*pro-vaccine, pro-vaccination*). Низка новотворів позначає наслідки вакцинації (*post-vaccination, vaccine-induced, vaccine-related, post-vaccine,*

Рис. 8. Слово *vax* у корпусі NOW

vaccine-derived, vaccination-related, vaccine-elicited, postvaccination, vaccine-associated), документи, які підтверджують вакцинацію (*proof-of-vaccination, vaccine-passport, proof-of-vaccine*), обмеження для невакцинованих (*vaccinated-only, vaccine-only, vaccine-or-test*).

У позитивному значенні вживаються слова *vaccine-preventable*, передусім у контексті захворювань, від яких існують вакцини (наприклад: *Implementing country plans to prevent and respond to outbreaks of vaccine-preventable diseases, and strengthen immunization systems as part of COVID-19 recovery efforts.*); *vaccine-driven, vaccine-led, vaccine-fuelled* – у контексті відновлення економіки завдяки вакцинації (наприклад: *A strong corporate earnings recovery and hopes of a vaccine-driven economic rebound have pushed US stocks to record levels in recent sessions. The British pound drew additional support from bets on a faster vaccine-led economic recovery. Investors are showing signs of increasing exuberance, reflecting optimism about a vaccine-fuelled global recovery and the changed economics of the post-coronavirus world*).

Варто згадати, що Оксфордський словник англійської мови 1 листопада 2021 року проголосив словом року 2021 слово *vax* – колоквіалізм, який як іменник означає «вакцина» або «вакцинація», як дієслово – «вакцинувати» [19]. Як зазначають дослідники Oxford Languages, які аналізували контент новин з метою відстежити зміни в англійській мові, у вересні 2021 *vax* вживалося у понад 72 рази частіше, ніж попереднього року. Корпусні дані NOW підтверджують різке зростання використання слова *vax* у 2021 році, особливо в листопаді (рис. 8).

У корпусі NOW виявлено низку похідних *vax*, зокрема *anti-vaxxers, vax, anti-vax, anti-vaxxer, vaxxed, anti-vaxx, vax-a-million, vaxx, unvaxxed, vaxathon, vaxed, anti-vaxers, vaxxers, antivaxxers, antivax* (результати пошуку **vax** у корпусі подано в порядку частотності, зі списку видалено власні назви виробників вакцин, компаній). Найчастотніші новотвори мають негативне значення, позначають противників вакцинації. Виявлено неологізми-словосполуки, які містять *vax*: *fully vaxxed, Covid vax, double vaxxed, vax card, vax mandate, Global Citizen VAX, vaxxed people (only), Vax-a-Million lottery, vax drive, mass vax, vax program, vax doses, vax rollout, vax rates, vax site, vax pass, vax campaign*.

Висновки. Корпусний підхід до дискурсу дає змогу виокремити ключові слова з великих масивів текстів та, відповідно, зробити статистично обґрунтовані висновки про тематику та основні концептуальні центри аналізованих текстів. Використовуючи дані та інструменти корпусу англійськомовних онлайн-новин NOW, встановлено частотні списки ключових слів англійськомовних новин за 2021 рік. З'ясовано, що найчастотніші ключові слова належать до концептуальних сфер PANDEMIC, MEDIA, POWER, ELECTIONS, WORK, BUSINESS AND INDUSTRY, LAW AND ORDER, STATE AND PEOPLE, PROBLEMS. Статистично виділяється в новинах 2021 року іменник *vaccine* (особливо в корпусі новин США). У результаті аналізу колокацій виявлено позитивну семантичну/дискурсну просодію колокацій з компонентом *vaccine*, що прагматично зумовлено наміром сформувати позитивні асоціації та переконати вакцинуватися. Виявлено специфічні колокації

в новинах США та Великобританії, які вказують на особливості сприйняття вакцинації в цих державах. Аналіз словоформ *vaccine* у корпусі показав підвищену словотвірну продуктивність слова

vaccine в англійськомовних новинах 2021 року. Показово, що значна кількість новотворів слова *vaccine* має негативне значення і вживається передусім для позначення противників вакцинації.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

1. Claridge C. News discourse. *Historical pragmatics* / eds. A.H. Jucker, I. Taavitsainen. Berlin and New York : Mouton de Gruyter, 2010. P. 587–620.
2. Bednarek M., Caple H. Introduction. *News Discourse. Bloomsbury Classics in Linguistics*. London : Bloomsbury Academic, 2012. P. 1–22. URL: <http://dx.doi.org/10.5040/9781350063747.0007> (дата звернення: 11.11.2021).
3. Van Dijk T.A. News as discourse. Hillsdale, NJ, US : Lawrence Erlbaum Associates, Inc., 1988. 200 p.
4. Baker P. Using Corpora in Discourse Analysis. London : Continuum, 2006. URL: <https://www.lancaster.ac.uk/staff/bakerjp/usingcorpora.htm> (дата звернення: 11.11.2021).
5. Кривенко Г.Л. Корпусні дослідження дискурсу: становлення, стан і перспективи. *Вісник Київського національного лінгвістичного університету. Серія «Філологія»*. 2017. Т. 20. № 1. С. 51–63. URL: http://nbuv.gov.ua/UJRN/Vknl_u_fil_2017_20_1_8 (дата звернення: 11.11.2021).
6. Phillips M. Lexical structure of text. Discourse analysis monograph, no. 12. English Language Research. University of Birmingham, 1989. 86 p.
7. Scott M. WordSmith Tools manual. Version 3.0. Oxford : Oxford University Press, 1998. 149 p.
8. Ädel A. How to use corpus linguistics in the study of political discourse. *The Routledge Handbook of Corpus Linguistics* / eds. A. O’Keeffe, M. McCarthy. London and New York : Routledge, 2010. P. 591–604.
9. O’Halloran K. How to use corpus linguistics in the study of media discourse. *The Routledge Handbook of Corpus Linguistics* / eds. A. O’Keeffe, M. McCarthy. London and New York : Routledge, 2010. P. 563–577.
10. Scott M., Tribble C. Textual patterns: Key words and corpus analysis in language education. Amsterdam : John Benjamins, 2006. 203 p.
11. Gabrielatos C. Keyness Analysis: nature, metrics and techniques. *Corpus Approaches to Discourse: A critical review* / eds. C. Taylor, A. Marchi. Oxford : Routledge. 2018. P. 225–258.
12. NOW Corpus (News on the Web) : веб-сайт. URL: <https://www.english-corpora.org/now/> (дата звернення: 11.11.2021).
13. Full-text corpus data : веб-сайт. URL: <https://www.corpusdata.org/> (дата звернення: 11.11.2021).
14. Sinclair J., Carter. R. Trust the text : language, corpus and discourse. London, New York, Routledge, 2004. 224 p.
15. Gabrielatos C. Collocational analysis as a gateway to critical discourse analysis. The case of the construction of refugees, asylum seekers and immigrants in the UK press. English Language Institute, University of Michigan, 10 March 2008. URL: <https://eprints.lancs.ac.uk/id/eprint/31109/1/RASIM-Collocations-ELI-2008.pdf> (дата звернення: 11.11.2021).
16. Reisigl M., Wodak R. Discourse and Discrimination: Rhetorics of racism and antisemitism. London and New York : Routledge, 2001. 298 p.
17. Sedlak M. You really do make an unrespectable foreign policy. *Racism at the Top: Parliamentary discourses on ethnic issues in six European states* / eds. R. Wodak, T.A. van Dijk. Klagenfurt, Austria : Drava-Verlag, 2000. P. 107–168.
18. Voyant tools : веб-сайт. URL: <https://voyant-tools.org/> (дата звернення: 11.11.2021).
19. Word of the Year 2021: Vax. *Oxford Languages* : веб-сайт. URL: <https://languages.oup.com/word-of-the-year/2021/> (дата звернення: 11.11.2021).