ЛІНГВІСТИКА XXI СТОЛІТТЯ: ПРОБЛЕМАТИКА І МЕТОДОЛОГІЯ

КИЇВ – 2026

Міністерство освіти і науки України
Український державний університет
імені Михайла Драгоманова
Факультет іноземної філології

ЛІНГВІСТИКА XXI СТОЛІТТЯ: ПРОБЛЕМАТИКА І МЕТОДОЛОГІЯ

Колективна монографія

Київ – 2026

УДК 81’1-11/-13
ББК 81 Укр
Л 78

Рекомендовано Вченою радою факультету іноземної філології
Українського державного університету імені Михайла Драгоманова
(протокол № 9 від 23 квітня 2026 р.)

Рецензенти:
Куварова Олена Костянтинівна, доктор філологічних наук, професор, професор кафедри загального мовознавства та слов’янознавства Дніпровського національного університету імені Олеся Гончара.
Леміш Наталія Євгенівна, доктор філологічних наук, професор, декан факультету іноземної філології Українського державного університету імені Михайла Драгоманова.
Пономаренко Володимир Панасович, доктор філологічних наук, професор, завідувач відділу романських, германських і балтійських мов Інституту мовознавства імені О. О. Потебні НАН України.

Л 78 Лінгвістика XXI століття: проблематика і методологія: колективна монографія / за наук. ред. проф. Ю. В. Кравцової. Київ: Вид-во УДУ імені Михайла Драгоманова, 2026. 250 с. Мова укр. Електронне видання

ISBN 978-966-931-354-6

У колективній монографії розглядаються актуальні проблеми та методологічний інструментарій провідних галузей сучасної лінгвістики – лінгвокогнітології, лінгвоконцептології, лінгвокомунікативістики, інтернет-лінгвістики, медіалінгвістики, дискурсології, лінгвопоетики, комп’ютерної лінгвістики, корпусної лінгвістики.

Призначено для науковців, викладачів, здобувачів вищої освіти.

УДК 81’1-11/-13
ББК 81 Укр

© Автори розділів монографії, 2026
© Український державний університет імені Михайла Драгоманова, 2026

ISBN 978-966-931-354-6

РОЗДІЛ 3.

МОВНА ПАМ'ЯТЬ НАЦІЇ ЯК ОБ'ЄКТ КОРПУСНОЇ ЛІНГВІСТИКИ: МЕТОДОЛОГІЧНІ ВИМІРИ ЦИФРОВОЇ ГУМАНІТАРИСТИКИ

DOI: https://doi.org/10.31392/UDU-mono-fif-2026.03

Світлана Матвєєва
https://orcid.org/0000-0002-8357-9366
Рамунє Каспере
https://orcid.org/0000-0003-0782-3758

Вступ. У сучасних умовах тотальної цифровізації гуманітарного знання та глибоких соціальних і політичних трансформацій мова дедалі частіше осмислюється не лише як інструмент комунікації, а як фундаментальний механізм збереження, трансляції та переосмислення культурної пам'яті. У цьому вимірі вона постає водночас як архів колективного досвіду та як динамічний простір його актуалізації й інтерпретації. Особливої концептуальної ваги такий підхід набуває в контексті національної ідентичності, де мовна пам'ять виконує структуротвірну функцію, забезпечуючи тяглість історичного досвіду, його символічну легітимацію та відтворення в сучасному дискурсі (Margalit, 2004; Mianowski et al., 2019).

Сьогодні важливим є з'ясування того, яким чином мовна пам'ять нації фіксується, трансформується та інституціоналізується в цифрових лінгвістичних ресурсах, насамперед у корпусах текстів, тому фокусом аналізу стає взаємодія між колективною пам'яттю, мовною системою та цифровими методами дослідження, що дозволить перейти від умоглядних міркувань до емпірично обґрунтованого вивчення семантичних, дискурсивних і ціннісних зрушень. Теоретичним підґрунтям аналізу слугують концепції колективної та культурної пам'яті, розроблені в соціології, культурній антропології та дослідженнях пам'яті, тоді як методологічною основою виступає інструментарій корпусної лінгвістики і цифрової гуманітаристики. Такий міждисциплінарний

підхід уможливлює аналіз мовної пам’яті як емпірично доступного феномена, закарбованого в масивах реального мовного вжитку.

Сучасна наукова парадигма демонструє зсув від статичного розуміння мови як замкненої системи до її динамічного осмислення як «живого архіву» культури. Цей зсув стає можливим завдяки синергії цифрової гуманітаристики, що забезпечує концептуальний та епістемологічний контекст, і корпусної лінгвістики, яка надає точний емпіричний інструментарій для аналізу мовної пам’яті. Фундаментальним імпульсом для такого підходу стала праця Ж.-Б. Мішеля, Й.-К. Шеня, А. П. Айдена, А. Вереша та М. К. Ґрея, в якій було запропоновано кількісний аналіз культури на матеріалі масиву Google Books. Автори фактично легітимізували ідею дослідження культурної пам’яті на рівні великих даних, продемонструвавши, що динаміка мовного вжитку може слугувати індикатором колективної уваги, забування та символічної значущості (Michel et al., 2011).

3.1. Корпус як репрезентація мовної пам’яті нації.

Теоретичне підґрунтя для розуміння корпусу як репрезентації пам’яті знаходимо у працях В. Тейберта, який радикально переосмислює природу значення, трактуючи його як результат сукупності попередніх актів мовлення. У цьому сенсі корпус постає як матеріалізована пам’ять мовної спільноти, а корпусний аналіз – як інструмент її реконструкції (Teuber, 2010).

Подальший розвиток цієї логіки демонструють дослідження Е. Макенері та Х. Бейкер (McEnery, Baker, 2016), а також когнітивно-дискурсивні студії А. Музольффа (Musolff, 2016), які показують, як стійкі метафоричні сценарії, дискурсивні шаблони та оцінні моделі закарбовуються в мовній пам’яті й відтворюються в сучасному дискурсі.

Поняття мовної пам’яті органічно пов’язане з класичною теорією колективної пам’яті, сформульованою М. Альбваксом, відповідно до якої індивідуальні спогади завжди конструюються в межах соціальних рамок, однією з ключових серед яких є мова (Halbwachs, 1950). Подальше розрізнення комунікативної та культурної пам’яті, запропоноване Я. та А. Ассманами, дозволяє інтерпретувати корпуси як простір фіксації переходу від плинного мовного вжитку до інституціоналізованих форм пам’яті (Assmann,

1992; Assmann, 1999). К. А. Кроулі показує, що наративні структури формують тривалу пам'ять через тематичні та емоційно виразні елементи, що кореспондує з результатами корпусного аналізу семантичних зрушень у національному дискурсі (Crowley, 2018).

Корпусна лінгвістика уможливлює емпіричне простеження цього переходу: стабілізація колокацій, дискурсивних ролей і моделей оцінювання сигналізує про входження мовних одиниць до культурної пам'яті. В. Любінене стверджує, що цифрові контексти формують нові мовні практики, що важливо для розуміння того, як сучасні цифрові корпуси фіксують і трансформують колективну пам'ять (Liubinienė, 2012). Таким чином, корпуси постають не лише як технічний інструмент аналізу, а як цифрове втілення мовної пам'яті спільноти.

Сучасна цифрова гуманітаристика орієнтована не просто на фіксацію мовних змін, а на їх пояснення в культурному та соціальному контексті. Обчислювальні моделі семантичного аналізу дають змогу виявляти різні типи змін значення слів, зокрема розмежовувати поступовий внутрішньомовний дрейф і різкі зрушення, спричинені екстралінгвістичними подіями.

Глобальні зміни в дистрибутивній семантиці слова, як правило, пов'язані з довготривалими процесами, такими як граматикалізація чи семантичне знебарвлення. Натомість локальні зміни в колі найближчих семантичних сусідів сигналізують про культурні злами, коли слово починає функціонувати в нових контекстах і набуває нових конотацій (Hamilton et al., 2016; Kutuzov et al., 2018). Таке розмежування дозволяє використовувати діахронічні векторні моделі не лише як лінгвістичний інструмент, а й як засіб вимірювання суспільних трансформацій (Garg et al., 2018; Gulordava, Baroni, 2011). Саме через аналіз таких змін стає можливим дослідження мовної пам'яті як колективного емоційного та ціннісного досвіду. Корпусний підхід дозволяє «оцифрувати» не лише лексичні одиниці, а й реакції спільноти на історичні події, зафіксовані в мовному вжитку.

3.2. Концепт УКРАЇНЕЦЬ як маркер мовної пам'яті нації.

Одним із показових маркерів мовної пам'яті є еволюція лексем, що позначають національну ідентичність. Аналіз

корпусних даних свідчить, що одиниця *українець* наприкінці XIX століття закріплюється як публічний ідентифікатор, витісняючи альтернативні номінації (див. Рис. 3.1.).

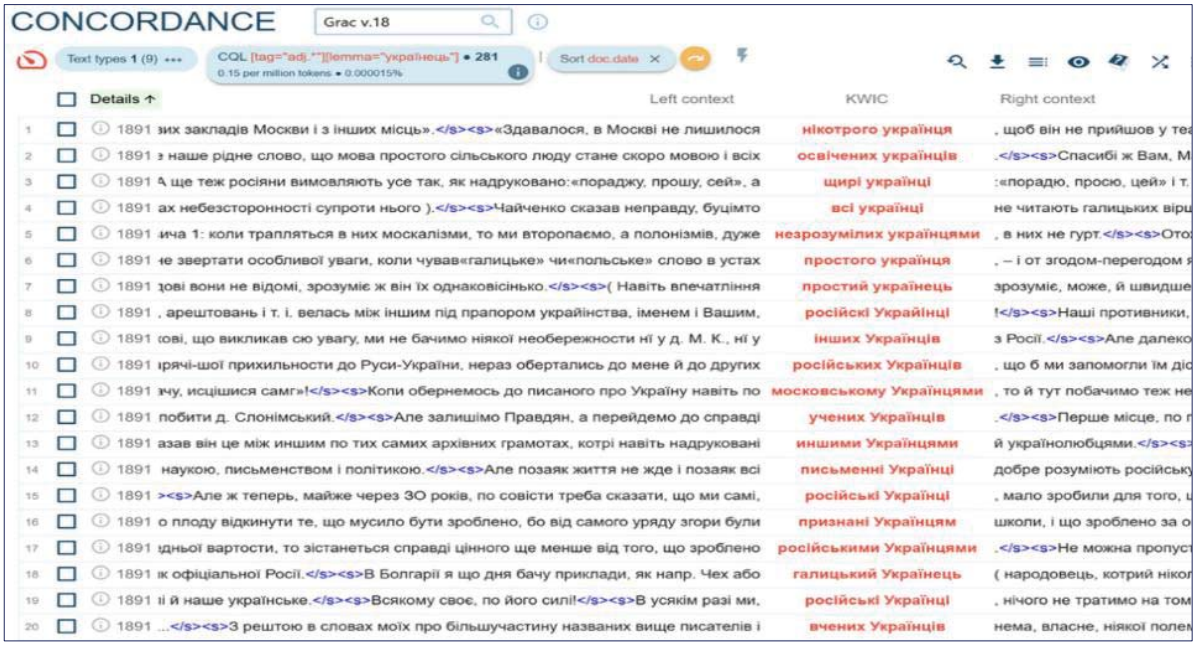


Рис. 3.1. Конкорданс лексеми *українець*, 1891 рік (сформовано за допомогою корпусу ГРАК, версія 18).

Після 1991 року спостерігається повна інституалізація цього ідентифікатора в публічному дискурсі. Корпусний аналіз його колокацій демонструє формування типізованого та нормативного образу носія ідентичності. Препозитивні означення на кшталт *кожен*, *пересічний*, *середньостатистичний* вказують на сприйняття української ідентичності як базової соціальної норми (див. Рис. 3.2.).



Рис. 3.2. Результати аналізу препозитивних колокацій лексеми *українець* (сформовано на основі даних Leipzig Corpora Collection (LCC), версія 2014).

Водночас атрибутивні прикметники *свідомий*, *справжній*, *етнічний*, *щирий* засвідчують етичну та ціннісну наповненість цієї ідентичності. Правобічне сусідство з модальними дієсловами (*повинен*, *має*, *мусить*) формує дискурс громадянської відповідальності та активної позиції (див. Рис. 3.3.).



Рис. 3.3. Результати аналізу атрибутивних колокацій лексеми *українець* (сформовано на основі даних корпусу LCC, версія 2014).

Додатковим маркером інституалізації є включення лексеми *українець* до соціоекономічних і медійних контекстів, а також її персоніфікація через імена публічних осіб і наративи успіху. Цей високий рівень частотності свідчить про зміцнення наративу громадянської відповідальності, обов'язку та потенціалу, які асоціюються з національною належністю. Сучасний *українець* у дискурсі не лише є, але й активно діє (дієслова *знає*, *живе*, *вважає*, *любить*). Крім того, значна кількість слів, пов'язаних зі споживанням (*споживає*, *витрачає*, *випиває*, *з'їдає*) вказує на використання терміна в контексті соціоекономічних та маркетингових досліджень, що також є маркером інституалізації.

Показовим є часте сусідство слова *українець* з іменами відомих особистостей (*Володимир*, *Віталій*, *Сергій*, *Олександр*, *Іван*, *Андрій* та ін.) та дієсловами, пов'язаними зі спортивними та публічними досягненнями (*виграв*, *здобув*, *переміг*, *став*). Це вказує на те, що публічний образ українця великою мірою формується через успішних медійних осіб, які виступають як символічні репрезентанти нації на міжнародній арені.

Таким чином, корпусний аналіз демонструє рух до стабільного, інституалізованого стану української ідентичності. Дискурс закріплює типізований ідеальний образ (*свідомий*

пересічний українець), набір обов'язків та потенціалів (висока модальність), персоніфікацію через успішних публічних діячів. У сукупності ці характеристики свідчать про стабілізацію та нормативне закріплення української ідентичності в мовній пам'яті.

Отже, українська ідентичність після 1991 року перестала бути лише декларативною, набувши конкретних лінгвістичних маркерів у масовому дискурсі. Проте ідентичність не існує у вакуумі. Вона завжди пов'язана з ключовими державницькими цінностями.

3.3. Концепт НЕЗАЛЕЖНІСТЬ як маркер мовної пам'яті нації.

Аналіз концепту НЕЗАЛЕЖНІСТЬ демонструє, як мовна пам'ять реагує на історичні виклики та трансформує ціннісні орієнтири. Корпусні дані різних часових зрізів фіксують перехід від реактивного дискурсу боротьби до інтерналізованої цінності, що потребує постійного підтримання.

Як бачимо, в 1940-х роках лексика зосереджена на темі зовнішньої загрози та опору. Переважають такі дієслова, як: *вимагати*, *відбирати*, *загрожувати*, *обстоювати*, *порушувати*. Ключовим фокусом є незалежність як об'єкт боротьби, що постійно перебуває під зовнішнім тиском і за який платять найвищу ціну (*сконати*). Дискурс є реактивним, оборонним (див. Рис. 3.4.).

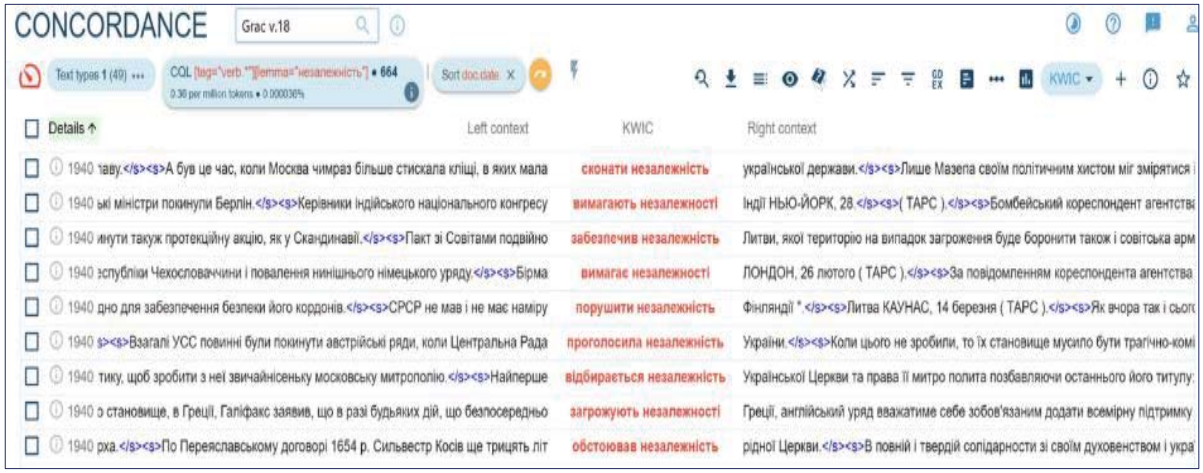


Рис. 3.4. Конкорданс лексеми *незалежність*, 1940 рік (сформовано за допомогою корпусу ГРАК, версія 18).

Протягом 1960-х і 1980-х років спостерігаємо етап *здобуття*, *збереження* та *ризику*. Тут ми бачимо розширення фокусу. Так, у 1960-ті роки з'являється тема збереження (*берегти*, *зберігати*) та

легітимізації (*визнавати*, *надавати*) (див. Рис. 3.5.). 1980-ті роки (напередодні незалежності) є періодом максимальної активізації, що охоплює весь цикл: від внутрішнього бажання (*бажати*) та боротьби (*добиватися*, *завойовувати*, *здобувати*, *проголошувати*) до усвідомлення крихкості та необхідності захисту (*втрачати*, *підривати*, *рятувати*). Ключовий фокус – незалежність перетворюється на прагматичну національну мету, що перебуває у зоні високого ризику (див. Рис. 3.6.).

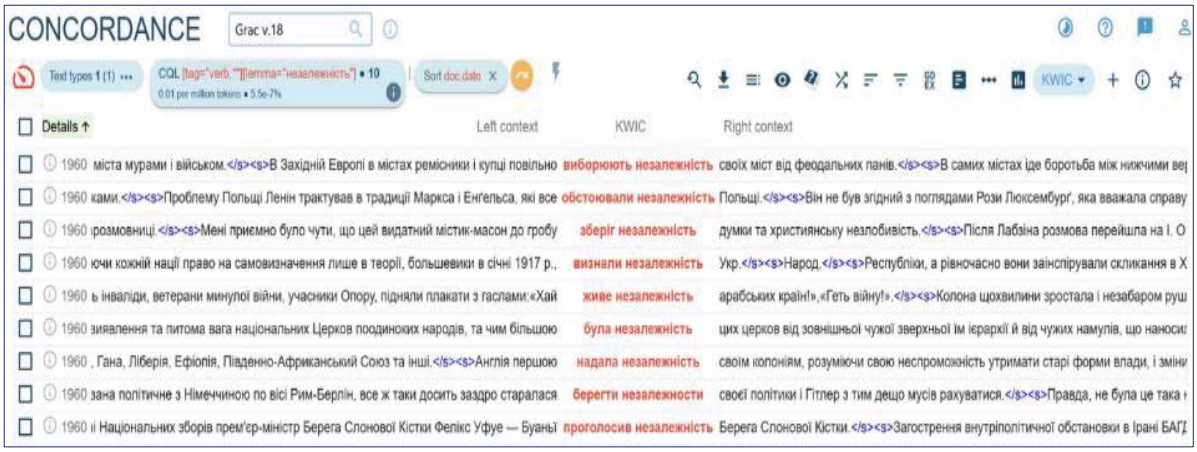


Рис. 3.5. Конкорданс лексеми *незалежність*, 1960 рік (сформовано за допомогою корпусу ГРАК, версія 18).

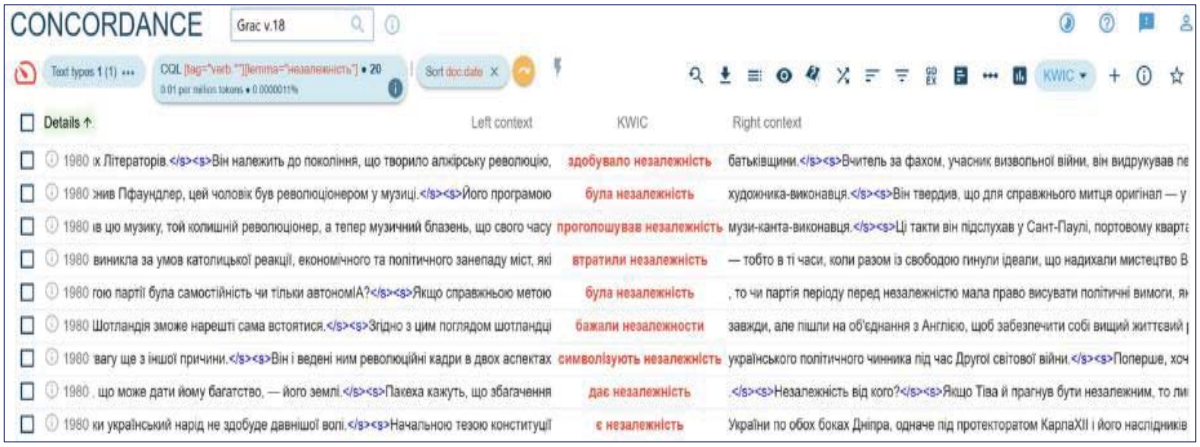


Рис. 3.6. Конкорданс лексеми *незалежність*, 1980 рік (сформовано за допомогою корпусу ГРАК, версія 18).

2000-ні роки ознаменовують етап інституалізації та зміцнення (див. Рис. 3.7.). Після проголошення незалежності фокус зміщується з отримання на підтримку та розвиток існуючого статусу. У цей період переважають дієслова активного збереження та розбудови: *відстоювати*, *втримувати*, *зберігати*, *зміцнювати*,

забезпечувати. Також зберігається мотив *виборювати* та *відвойовувати*, що свідчить про усвідомлення незавершеності процесу. Ключовий фокус – незалежність є активним, безперервним процесом, який вимагає політичної волі та інституціональних зусиль.

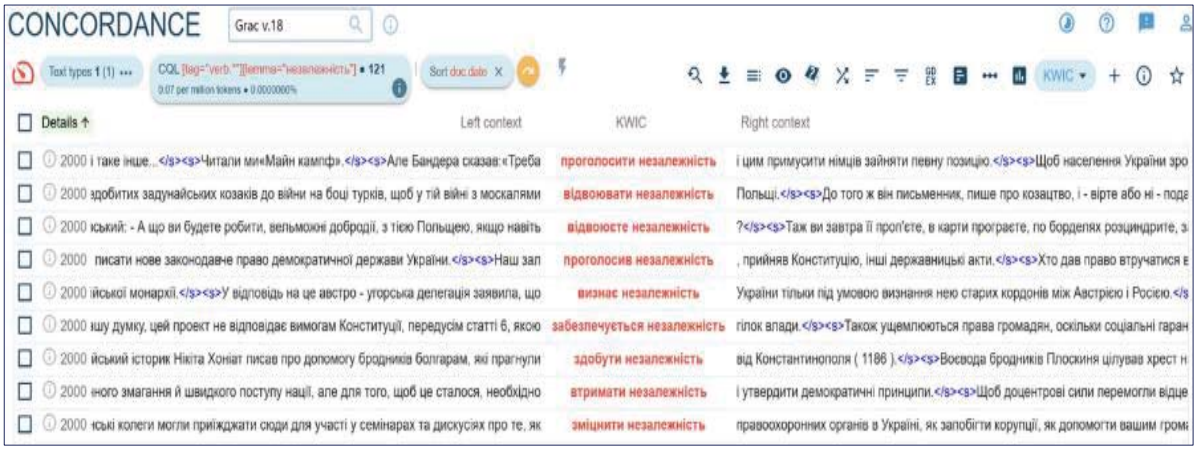


Рис. 3.7. Конкорданс лексеми *незалежність*, 2000 рік (сформовано за допомогою корпусу ГРАК, версія 18).

2020-ті роки – це незалежність як фундаментальна цінність. У найновішому зрізі значно зростає частка лексики, що відбиває суб'єктивне, емоційне ставлення до незалежності: *любити*, *цінувати*, *прагнути*, *хотіти*, *дочекатися*. Вона стає особистою, внутрішньою цінністю, а не лише політичним актом. Водночас, дієслова *виборювати*, *здобувати* свідчать про те, що цей ціннісний вимір співіснує з триваючою боротьбою в умовах зовнішньої агресії (див. Рис. 3.8.).

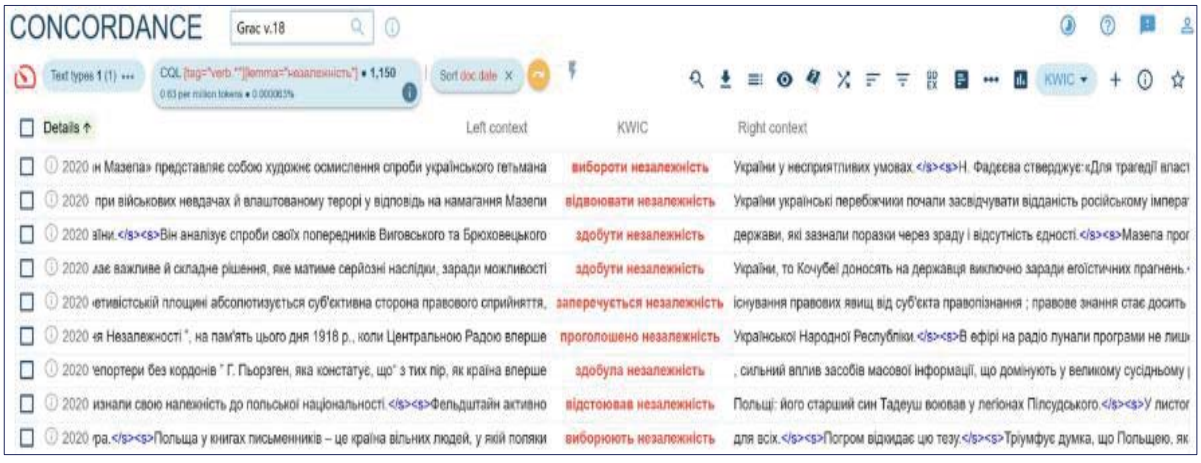


Рис. 3.8. Конкорданс лексеми *незалежність*, 2020 рік (сформовано за допомогою корпусу ГРАК, версія 18).

Слово *незалежність* через колокації демонструє, як культурно-історичний зсув відбивається в семантичному наповненні. Так, у 1940-х роках незалежність осмислюється як об’єкт зовнішнього тиску та спротиву; у другій половині XX століття – як прагматична політична мета; у XXI столітті – як активний процес інституційного захисту; у найновішому періоді – як глибоко особиста та емоційно насичена цінність (див. Рис. 3.9.).

	1940	1960	1980	2000	2020
Бажати					
Берегти					
Боронити					
Бути					
Виборювати					
Визнавати					
Визначати					
Виказувати					
Вимагати					
Відбирати					
Відбуватися					
Відвойовувати					
Відстоювати					
Втрачати					
Втримувати					
Давати					
Добиватися					
Досягати					
Дотримуватися					
Дочекатися					
Забезпечувати					
Завойовувати					
Загрожувати					
Зберігати					
Здобувати					
Зміцнювати					
Любити					
Надавати					
Обстоювати					
Отримувати					
Підривати					
Порушувати					
Прагнути					
Проголошувати					
Просити					
Рятувати					
Символізувати					
Сконати					
Хотіти					
Цінувати					

Рис. 3.9. Діахронічна матриця колокацій лексеми *незалежність* (1940–2020 рр.).

Отже, аналіз дієслів, які передують іменнику *незалежність* у різних часових зрізах, демонструє чітку еволюцію сприйняття цього концепту – від боротьби та зовнішньої загрози до внутрішньої цінності, досягнення та активного збереження (Diez, 1999; Wilson, 2015; Wodak, Meyer, 2016). Ця цінність, зміцнена у мовній пам'яті (Skinner, 2002), стала об'єктом найбільшої загрози.

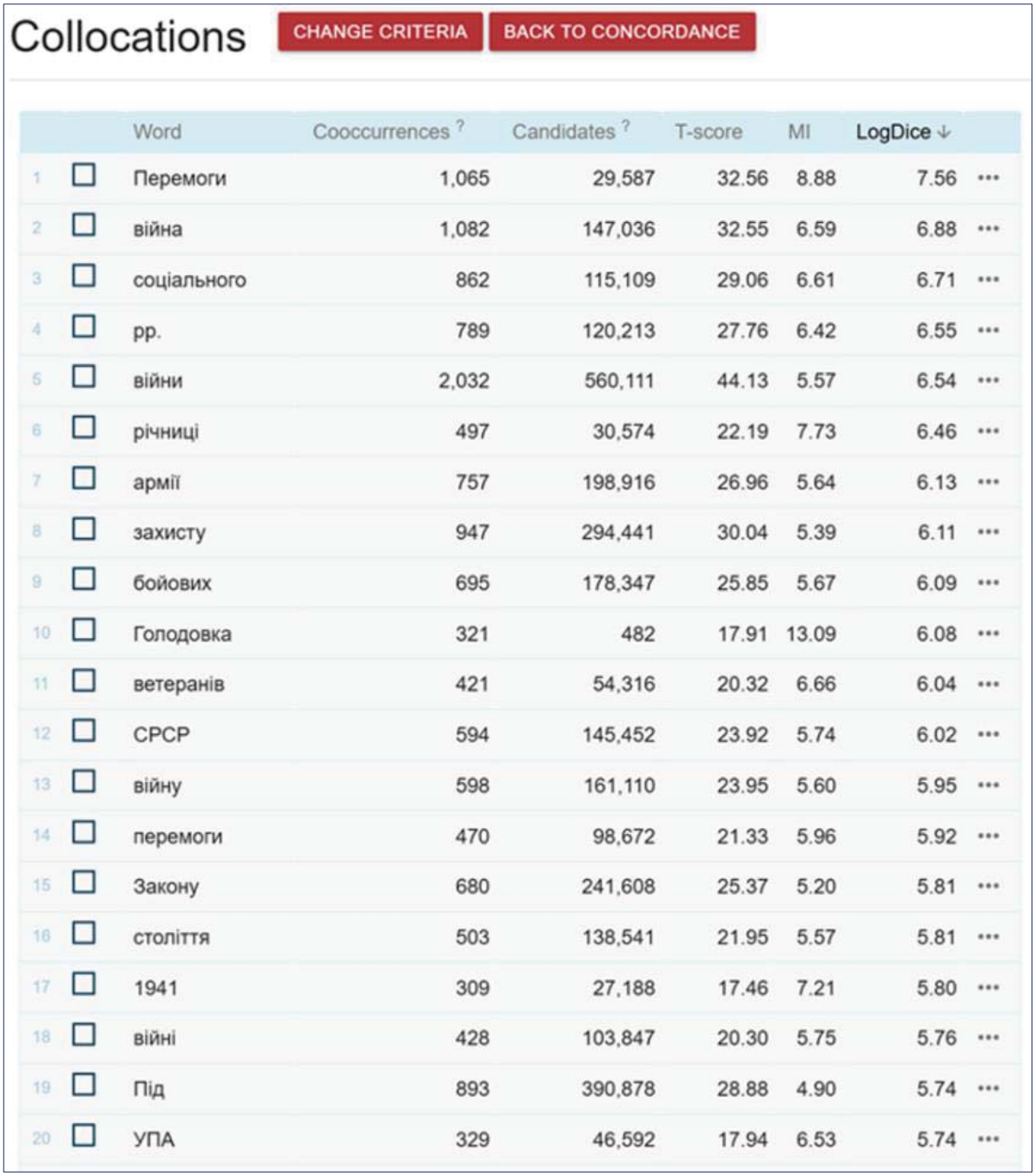
3.4. Концепт ВІЙНА як маркер мовної пам'яті нації.

Найбільш радикальні семантичні зрушення спостерігаються у функціонуванні концепту ВІЙНА. Мова віддзеркалює різке входження поняття *війна* у щоденний комунікативний простір із 2014 року. Результати корпусно-дискурсивного аналізу медійних джерел після 2014 року демонструють кардинальну семантичну трансформацію іменника *війна* в українському публічному дискурсі. Якщо раніше контексти могли бути абстрактно-історичними, то тепер вони асоціюються з *агресією*, *вторгненням*, *окупованими територіями*, *героями*, *волонтерами*. Це дозволило точно розрізнити український дискурс від того, що використовував хибний евфемізм *спеціальная воєнная операція*.

Аналіз топ-100 колокатів за показником LogDice чітко розмежовує сприйняття війни як історичного артефакту та як актуальної, внутрішньої, екзистенційної реальності.

У період 1991–2013 роки слово *війна* існувало переважно в контексті історичної пам'яті та зовнішніх подій. Спостерігається домінування історичного наративу: колокати *Другої*, *Великої*, *Вітчизняної*, *1941*, *1945*, *СРСР*, *Радянського*, *Союзу*, *ветеранів*, *перемоги* демонструють, що ключовим семантичним супутником була *Друга світова війна* (у радянському і пострадянському трактуванні). Присутність колокатів *УПА*, *ОУН*, *українці*, *німці*, *євреї*, *голодовка*, *1933* вказує на внутрішню боротьбу і конфлікт пам'ятей та переосмислення історичних подій. Колокати *Афганістан*, *Ірак*, *росія*, *Німеччина* свідчать про те, що інші конфлікти сприймалися як зовнішні або віддалені події з позицій географії та світового контексту (див. Рис. 3.10.).

Отже, для цього періоду *війна* – це переважно минуле, об'єкт меморіалізації та ідеологічної боротьби, а не актуальна загроза для країни.



Collocations

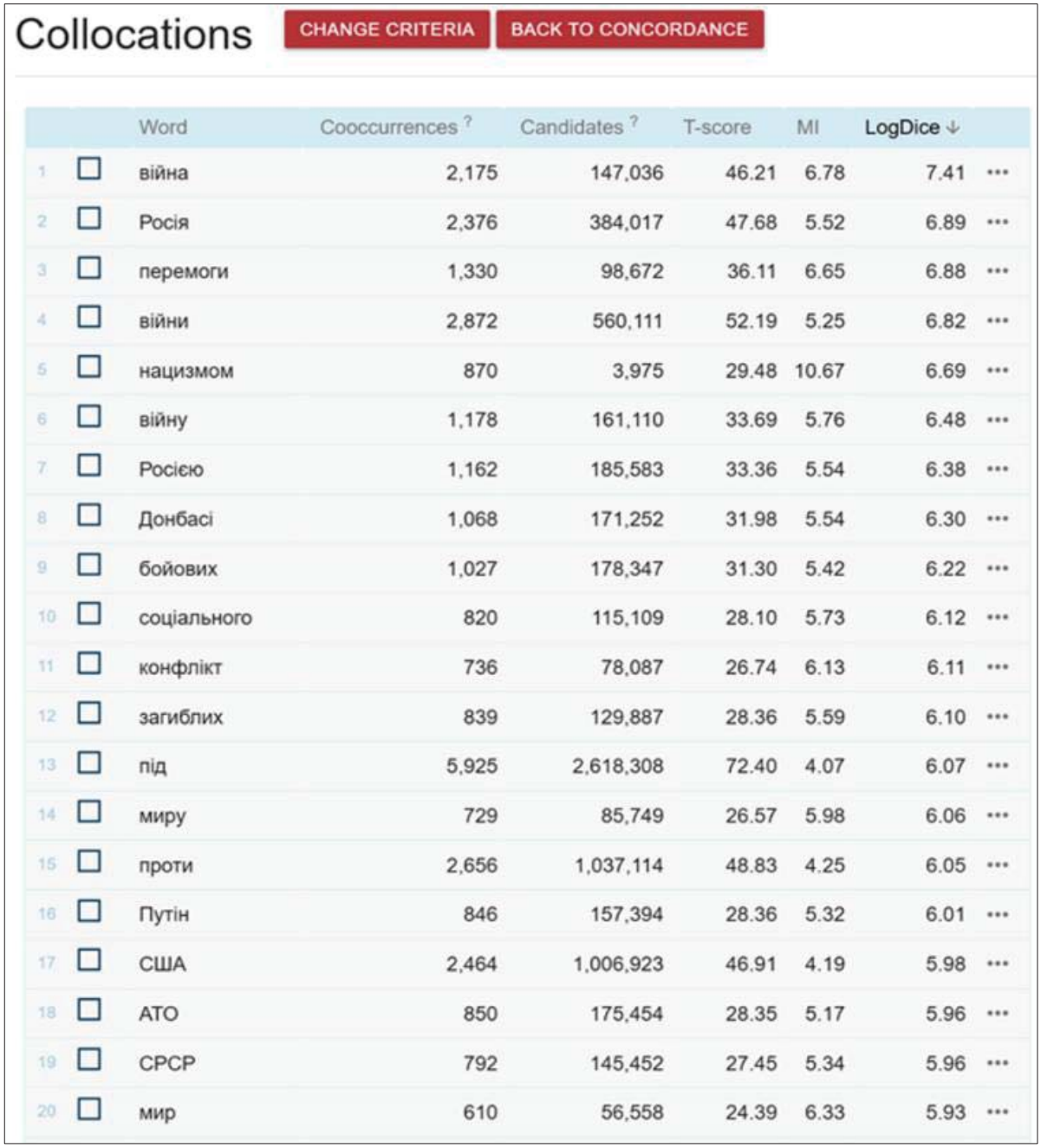
CHANGE CRITERIA | BACK TO CONCORDANCE

		Word	Cooccurrences	Candidates	T-score	MI	LogDice ↓	
1	☐	Перемоги	1,065	29,587	32.56	8.88	7.56	•••
2	☐	війна	1,082	147,036	32.55	6.59	6.88	•••
3	☐	соціального	862	115,109	29.06	6.61	6.71	•••
4	☐	рр.	789	120,213	27.76	6.42	6.55	•••
5	☐	війни	2,032	560,111	44.13	5.57	6.54	•••
6	☐	річниці	497	30,574	22.19	7.73	6.46	•••
7	☐	армії	757	198,916	26.96	5.64	6.13	•••
8	☐	захисту	947	294,441	30.04	5.39	6.11	•••
9	☐	бойових	695	178,347	25.85	5.67	6.09	•••
10	☐	Голодовка	321	482	17.91	13.09	6.08	•••
11	☐	ветеранів	421	54,316	20.32	6.66	6.04	•••
12	☐	СРСР	594	145,452	23.92	5.74	6.02	•••
13	☐	війну	598	161,110	23.95	5.60	5.95	•••
14	☐	перемоги	470	98,672	21.33	5.96	5.92	•••
15	☐	Закону	680	241,608	25.37	5.20	5.81	•••
16	☐	століття	503	138,541	21.95	5.57	5.81	•••
17	☐	1941	309	27,188	17.46	7.21	5.80	•••
18	☐	війні	428	103,847	20.30	5.75	5.76	•••
19	☐	Під	893	390,878	28.88	4.90	5.74	•••
20	☐	УПА	329	46,592	17.94	6.53	5.74	•••

Рис. 3.10. Топ-100 колокатів лексеми *війна* за показником LogDice: період «історичної дистанції» (1991–2013 рр.) (сформовано за допомогою корпусу ГРАК, версія 18).

Після 2014 року семантика змінюється на актуальну та географічно локалізовану, зберігаючи при цьому елементи старого наративу. Найбільш вагомими стають власні назви, що позначають місцевість і локальний характер конфлікту: *Донбас*, *Крим*, *схід*, *АТО*, *анексія*, *окупація*. Це перетворює війну на внутрішній географічно визначений феномен. Колокати *росія*, *путін*, *рф*

остаточно закріплюють пряме визначення ворога й агресора, хоча термін *конфлікт* все ще уживається поруч із *війна*. Дієслова *веде*, *триває* підкреслюють, що війна є актуальною, поточною, незавершеною подією. Елементи *перемога*, *Друга*, *світова*, *нацизм*, *Німеччина*, *СРСР* все ще високо ранжуються, вказуючи на те, що за інерцією пам'яті дискурс 2014–2021 років намагався вписати нову війну в історичні рамки (див. Рис. 3.11.).



Collocations

CHANGE CRITERIA BACK TO CONCORDANCE

	Word	Cooccurrences	Candidates	T-score	MI	LogDice
1	війна	2,175	147,036	46.21	6.78	7.41
2	Росія	2,376	384,017	47.68	5.52	6.89
3	перемоги	1,330	98,672	36.11	6.65	6.88
4	війни	2,872	560,111	52.19	5.25	6.82
5	нацизмом	870	3,975	29.48	10.67	6.69
6	війну	1,178	161,110	33.69	5.76	6.48
7	Росією	1,162	185,583	33.36	5.54	6.38
8	Донбасі	1,068	171,252	31.98	5.54	6.30
9	бойових	1,027	178,347	31.30	5.42	6.22
10	соціального	820	115,109	28.10	5.73	6.12
11	конфлікт	736	78,087	26.74	6.13	6.11
12	загиблих	839	129,887	28.36	5.59	6.10
13	під	5,925	2,618,308	72.40	4.07	6.07
14	миру	729	85,749	26.57	5.98	6.06
15	проти	2,656	1,037,114	48.83	4.25	6.05
16	Путін	846	157,394	28.36	5.32	6.01
17	США	2,464	1,006,923	46.91	4.19	5.98
18	АТО	850	175,454	28.35	5.17	5.96
19	СРСР	792	145,452	27.45	5.34	5.96
20	мир	610	56,558	24.39	6.33	5.93

Рис. 3.11. Топ-100 колокатів лексеми *війна* за показником LogDice: період «локалізованого конфлікту» (2014–2021 рр.) (сформовано за допомогою корпусу ГРАК, версія 18).

Таким чином, *війна* цього періоду – це актуальний, локалізований конфлікт із чітко ідентифікованим агресором, який усвідомлюється як загроза суверенітету, території, країні.

Після повномасштабного вторгнення (період 2022–2025 роки) семантичне поле зазнає радикального зсуву в бік прямого звинувачення, правової оцінки та екзистенційної загрози. Колокати *24*, *лютий*, *2022*, *вторгнення*, *повномасштабний* чітко фіксують точну дату та характер події – пряме вторгнення. З'являються терміни *порушення*, *КК*, *ст.*, *умисний*, *вбивство*, *терор*, що відображає перехід від «конфлікту» до злочину та необхідності правової кваліфікації та судового переслідування. Зростає частотність колокатів *ЗСУ*, *солдатів*, *окупантів*, *російських*, *росіян*, підкреслюючи пряму участь та опозицію, процеси мілітаризації та персоніфікації. Згадки про *втрати*, *людей*, *дітей* підкреслюють гуманітарний та соціальний вимір. *НАТО*, *США*, *підтримка*, *допомога* свідчать про інтернаціоналізацію конфлікту, вихід у міжнародний контекст та залежність від зовнішньої підтримки. Так, війна в ці роки – це повномасштабна агресія, злочин, екзистенційна загроза, що вимагає правової оцінки та міжнародної допомоги для перемоги (див. Рис. 3.12.).

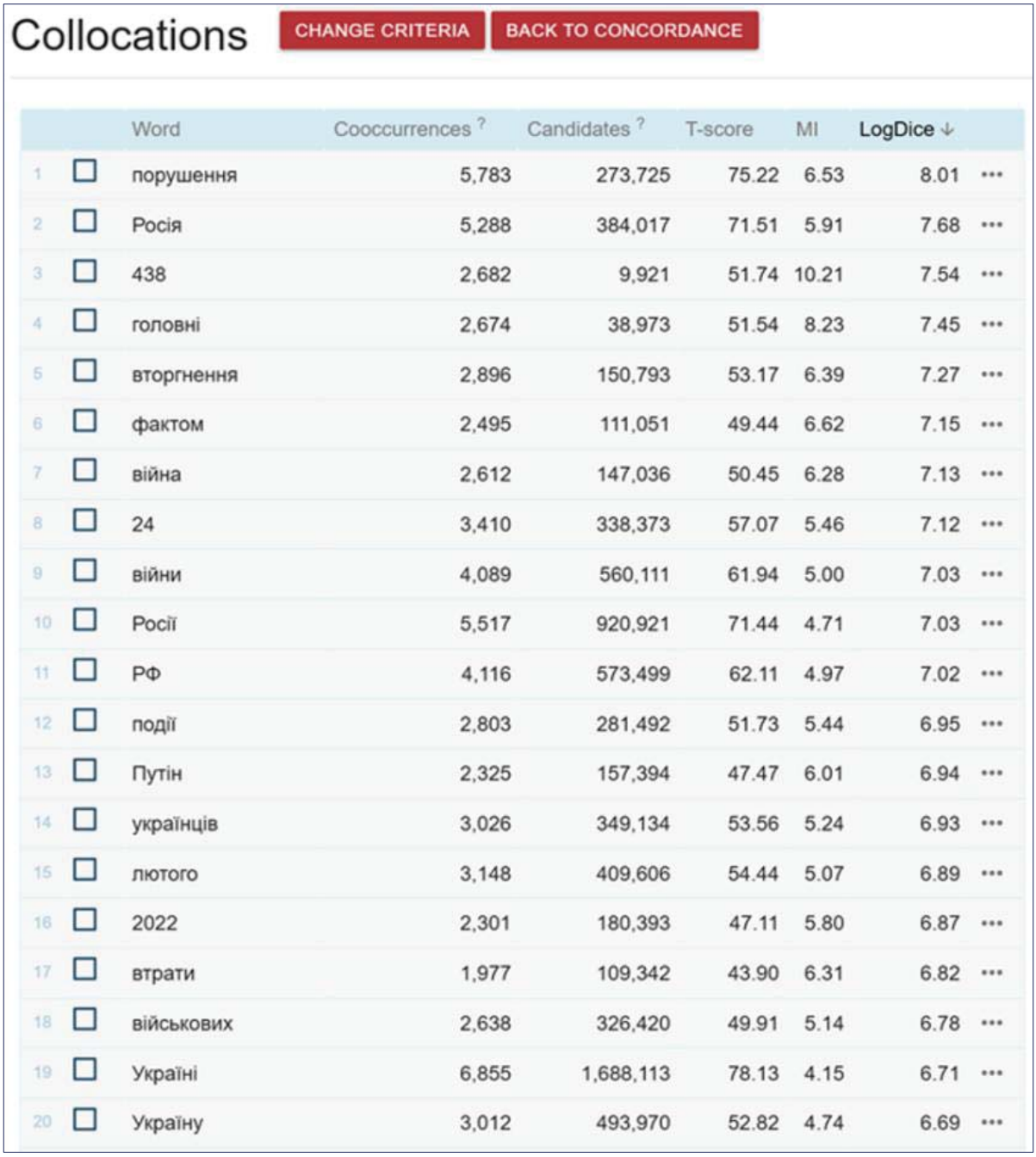
Отже, спостерігається певна загальна семантична динаміка. Трансформація семантичного поля слова *війна* в українському корпусі після 2014 року демонструє відмову від історичної абстракції та перехід до актуальної, конкретної та емоційно зарядженої реальності, а саме:

1) зсув від минулого до теперішнього: від *Великої Вітчизняної* до *повномасштабного вторгнення*;

2) звуження географії: від *Іраку* та *Афганістану* до *Донбасу* та *України*;

3) чітка ідентифікація: від *воюючої сторони* до *росії*, *рф*, *путіна*;

4) правова оцінка: включення юридичних термінів як ключових супутників.



	Word	Cooccurrences	Candidates	T-score	MI	LogDice
1	порушення	5,783	273,725	75.22	6.53	8.01
2	Росія	5,288	384,017	71.51	5.91	7.68
3	438	2,682	9,921	51.74	10.21	7.54
4	головні	2,674	38,973	51.54	8.23	7.45
5	вторгнення	2,896	150,793	53.17	6.39	7.27
6	фактом	2,495	111,051	49.44	6.62	7.15
7	війна	2,612	147,036	50.45	6.28	7.13
8	24	3,410	338,373	57.07	5.46	7.12
9	війни	4,089	560,111	61.94	5.00	7.03
10	Росії	5,517	920,921	71.44	4.71	7.03
11	РФ	4,116	573,499	62.11	4.97	7.02
12	події	2,803	281,492	51.73	5.44	6.95
13	Путін	2,325	157,394	47.47	6.01	6.94
14	українців	3,026	349,134	53.56	5.24	6.93
15	лютого	3,148	409,606	54.44	5.07	6.89
16	2022	2,301	180,393	47.11	5.80	6.87
17	втрати	1,977	109,342	43.90	6.31	6.82
18	військових	2,638	326,420	49.91	5.14	6.78
19	Україні	6,855	1,688,113	78.13	4.15	6.71
20	Україну	3,012	493,970	52.82	4.74	6.69

Рис 3.12. Топ-100 колокатів лексеми *війна* за показником LogDice: період «екзистенційного зламу» (2022–2025 рр.) (сформовано за допомогою корпусу ГРАК, версія 18).

Висновки. Корпусно-дискурсивний аналіз демонструє чітку динаміку від історизованого наративу (1991–2013) до локалізованого конфлікту (2014–2021) і далі – до екзистенційної та правово кваліфікованої реальності (2022–2025). Ці зміни колокаційного поля, зростання частотності юридичної, мілітарної та гуманітарної лексики свідчать про трансформацію національної

свідомості, зафіксовану у мовній пам'яті. Мова в цьому випадку постає як безпосередній індикатор переосмислення безпеки, ідентичності та місця держави у світовому контексті.

Проведене дослідження підтверджує, що мовна пам'ять нації є динамічною системою, яка одночасно акумулює минулий досвід і активно формує способи осмислення сучасності. Корпусна лінгвістика та цифрова гуманітаристика дозволяють емпірично досліджувати ці процеси, виявляючи семантичні, дискурсивні та ціннісні зрушення на великомасштабному матеріалі.

Корпуси української мови постають як надійний архів культурної пам'яті та як простір творення нових інтерпретацій. Аналіз лексем *українець*, *незалежність* і *війна* демонструє, як мовні одиниці фіксують інституалізацію ідентичності, трансформацію державницьких цінностей і екзистенційний досвід спільноти. У цьому сенсі цифрова гуманітаристика виходить за межі технічного інструментарію і набуває етичного виміру, адже робота з корпусами означає відповідальність за репрезентацію пам'яті, досвіду та образу людини в цифровому просторі. Мовна пам'ять постає не як застиглий архів, а як активна історична сила, що формує горизонти майбутнього.

Література:

Assmann, A. (1999). *Erinnerungsräume: Formen und Wandlungen des kulturellen Gedächtnisses*. München: Verlag C.H. Beck.

Assmann, J. (1992). *Das Kulturelle Gedächtnis: Schrift, erinnerung und politische identität in frühen Hochkulturen*. München: Verlag C.H. Beck.

Crowley, K.A. (2018). Memorability in Narration: An Overview of Mnemonic Features in Oral and Written Tradition. *Kalbų Studijos*, 32, 77–93.

Diez, T. (1999). Speaking 'Europe': the politics of integration discourse. *Journal of European Public Policy,* 6 (4), 598–613.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences,* 115 (16), E3635–E3644.

Gulordava, K., & Baroni, M. (2011). A distributional similarity approach to the detection of semantic change in the Google Books Ngram corpus. *Proceedings of the GEMS 2011 Workshop on Geometrical Models of Natural Language Semantics*, 67–71.

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. *Proceedings of the*

54th Annual Meeting of the Association for Computational Linguistics, 1489–1501.

Halbwachs, M. (1950). *Les cadres sociaux de la mémoire*. Paris: Presses Universitaires de France.

Kutuzov, A., Øvrelid, L., Szymanski, T., & Velldal, E. (2018). Diachronic Word Embeddings and Semantic Shifts: a Survey. *Proceedings of the 27th International Conference on Computational Linguistics*, 1384–1397.

Liubinienė, V. (2012). Elements, Contexts and Users of New Media Language. *Kalbų Studijos*, 21, 74–83.

Margalit, A. (2004). *The Ethics of Memory*. Harvard University Press.

McEnery, A., & Baker, H. (2016). *Corpus linguistics and 17th-century prostitution: Computational linguistics and history*. Bloomsbury Academic.

Mianowski, J. (ed.), Borodoю M. (ed.), & Schreiber, P. (ed.) (2019). *Memory, Identity and Cognition: Explorations in Culture and Communication*. Springer.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., & Gray, M. K. (2011). Quantitative analysis of cutlure using millions of digitized books. *Science*, 331 (6014), 176–182.

Musolff, A. (2016). *Political Metaphor Analysis: discourse and scenarios*. London: Bloomsbury Academic.

Skinner, Q. (2002). *Visions of Politics: Regarding Method*. Cambridge University Press.

Teubert, W. (2010). *Meaning, Discourse and Society*. Cambridge University Press.

Wilson, A. (2015). *The Ukrainians: Unexpected Natio*. Yale University Press.

Wodak, R., & Meyer, M. (2016). *Methods of Critical Discourse Studies*. Sage Publications.

***Джерела ілюстративного матеріалу*:**

ГРАК – Генеральний регіонально анотований корпус української мови / Шведова, М., Вальденфельс, Р. фон, Яригін, С., Рисін, А., Старко, В., Ніколаєнко, Т., Лукашевський, А. та ін. Київ, Львів, Єна, 2017–2025. https://uacorpus.org

LCC – Leipzig Corpora Collection: Ukrainian mixed corpus based on material from 2014. Leipzig Corpora Collection. Dataset. https://corpora.uni-leipzig.de?corpusId=ukr_mixed_2014

Резюме. У центрі уваги цього розділу – феномен мовної пам'яті як ключова складова національної ідентичності та її репрезентація у цифрових ресурсах сучасної лінгвістики. Мовна пам'ять концептуалізується як динамічний культурний конструкт, у якому

поєднуються історичний досвід, ціннісні орієнтири та колективні уявлення спільноти. У цьому контексті корпусна лінгвістика постає не лише як технічний інструментарій, а як інноваційний дослідницький простір, що уможливлює вияв і аналіз глибинних культурних смислів, зафіксованих у мовному вжитку різних історичних періодів.

У роботі продемонстровано застосування методів аналізу великих мовних даних, зокрема корпусної семантики та корпусної лексикографії, як інструментів реконструкції колективної пам'яті, що дозволяють виявляти неочевидні семантичні зсуви та ідеологічні нашарування у дискурсі. На матеріалі корпусів української мови простежується еволюція ключових ціннісних доменів, які віддзеркалюють суспільні трансформації останніх десятиліть: від відновлення державності до воєнного досвіду і процесів європейської інтеграції.

Зміни у частотності, контекстах уживання та семантичних профілях ключових лексем інтерпретуються як маркери глибинних зрушень у національній свідомості. Український мовний корпус у цьому дослідженні постає як унікальний і методологічно значущий кейс для осмислення того, як лінгвальні ландшафти трансформуються під впливом геополітичних зламів, соціальних травм і переосмислення історичної спадщини, розширюючи теоретичні горизонти сучасної гуманітаристики.

THE LINGUISTIC MEMORY OF A NATION AS AN OBJECT OF CORPUS LINGUISTICS: METHODOLOGICAL DIMENSIONS OF DIGITAL HUMANITIES

Svitlana Matvieieva,
Ramunė Kasperė

Summary. At the center of this study is the phenomenon of linguistic memory as a key component of national identity and its representation in the digital resources of contemporary linguistics. Linguistic memory is conceptualized as a dynamic cultural construct that integrates historical experience, value orientations, and the collective imaginaries of a community. Within this framework, corpus linguistics is approached not merely as a technical methodology but as an epistemologically innovative research space that enables the identification and interpretation of deep cultural meanings embedded in language use across historical periods.

The study introduces and applies advanced methods of large-scale linguistic data analysis – particularly corpus semantics and corpus-based lexicography – as tools for reconstructing collective memory, revealing non-obvious semantic shifts and latent ideological layers within public discourse.

Drawing on Ukrainian language corpora, the analysis traces the evolution of key value domains that reflect major social transformations of recent decades, ranging from the restoration of statehood to the experience of war and processes of European integration.

Changes in frequency patterns, usage contexts, and semantic profiles of key lexemes are interpreted as indicators of profound transformations in national consciousness. The Ukrainian linguistic corpus thus emerges as a unique and critically important case for understanding how linguistic landscapes are reshaped under conditions of geopolitical rupture, social trauma, and the re-evaluation of historical heritage, contributing to broader debates in philosophy of language, memory studies, and identity research.