

РОЗДІЛ 2 ГЕРМАНСЬКІ МОВИ

УДК 811.111'373.4

DOI <https://doi.org/10.32782/tps2663-4880/2025.40.1.6>

СТВОРЕННЯ ЧАСТКОВОГО СЛОВНИКА СЛЕНГУ В СОЦІАЛЬНИХ МЕРЕЖАХ (НА ОСНОВІ TWITTER (X) АБО REDDIT)

CREATING A PARTIAL DICTIONARY OF SLANG IN SOCIAL NETWORKS (BASED ON TWITTER (X) OR REDDIT)

Вороновська А.О.,

*orcid.org/0009-0006-5062-9052**Студентка III курсу факультету іноземної філології**Волинського національного університету імені Лесі Українки*

Кауза І.Б.,

*orcid.org/0000-0002-3030-2523**кандидат філологічних наук, доцент,**доцент кафедри прикладної лінгвістики**Волинського національного університету імені Лесі Українки*

Стаття присвячена аналізу створення частотного словника сучасного сленгу на основі глибокого аналізу текстів, отриманих із соціальних мереж Twitter (X) та Reddit. Ці платформи обрано як джерело матеріалу через їхню активну комунікативну природу, високу щоденну активність користувачів та широке використання неформальної мови. Аналіз доповнено вивченням особливостей зазначених соціальних мереж, того, як вони відображають мовні тенденції в реальному часі, як швидко реагують на події й дозволяють відстежувати динаміку мовних змін, зокрема появу та закріплення нових сленгових одиниць.

Дослідження ґрунтується на корпусному підході, який передбачає обробку великого масиву мовного матеріалу за допомогою цифрових інструментів. Такий підхід забезпечує об'єктивність і репрезентативність результатів, дозволяє виявляти закономірності в уживанні слів, їхню частотність, контекстуальну варіативність тощо. У межах роботи було здійснено низку етапів: збирання текстових даних (твітів, коментарів, дописів), їх очищення від шумів (посилань, технічних символів, емодзі тощо), а також токенизація – розбиття тексту на окремі мовні одиниці. Подальша лематизація дозволила звести слова до їх початкових форм, що є необхідним кроком для частотного аналізу.

Особлива увага приділялася формуванню чітких критеріїв відбору сленгових одиниць. Серед них – неформальний характер слова, його популярність серед молодіжної або онлайн-аудиторії, експресивне забарвлення та відповідність певним семантичним або функціональним ознакам. Зібрані одиниці класифікувались за частотою вживання, що дало змогу визначити ядро сучасного сленгу та провести його порівняльний аналіз між платформами. Отримані результати мають практичну цінність як для лексикографії, так і для подальших лінгвістичних, соціолінгвістичних та прикладних досліджень.

Ключові слова: сленг, частотний словник, соціальні мережі, Twitter, Reddit, цифрова лінгвістика, онлайн-комунікація, токенизація, лексикографія, інтернет-дискурс.

This study presents an attempt to develop a frequency dictionary of contemporary slang based on an in-depth analysis of texts collected from the social networks Twitter (X) and Reddit. These platforms were chosen as sources due to their dynamic communicative nature, high daily user activity, and the prevalence of informal language. One of their distinctive features is the reflection of real-time linguistic trends, which allows for the observation of how new slang units emerge, evolve, and become entrenched in digital discourse.

The research is grounded in a corpus-based approach, which involves processing a large volume of linguistic material using digital tools. This approach ensures objectivity and representativeness of the results and allows for the identification of regularities in word usage, frequency, and contextual variation. The study included several stages: data collection (tweets, comments, posts), noise removal (links, symbols, emojis, etc.), and tokenization – splitting texts into individual language units. Further lemmatization was applied to reduce words to their base forms, an essential step for frequency analysis.

Special attention was paid to defining clear criteria for selecting slang units. These included the informal nature of the word, its popularity among youth or online communities, expressive coloration, and correspondence to specific semantic or functional features. The collected units were classified according to frequency of use, enabling the identification of a core slang vocabulary and its comparative analysis across platforms. The results have practical value for lexicography, as well as for further linguistic, sociolinguistic, and applied research.

Key words: slang, frequency dictionary, social networks, Twitter, Reddit, digital linguistics, online communication, tokenization, lexicography, internet discourse.

Постановка проблеми. У сучасному інформаційному суспільстві соціальні мережі стали не лише засобом комунікації, а й потужним джерелом формування нових мовних явищ. Одним із таких явищ є сленг – динамічна, варіативна лексика, що швидко змінюється та часто залишається поза увагою традиційної лексикографії. Зважаючи на високий темп появи нових слів та мовних моделей, постає необхідність їхнього системного вивчення, зокрема шляхом створення частотних словників на основі великих масивів цифрового тексту.

Попри численні дослідження в галузі соціолінгвістики та комп'ютерної лінгвістики, феномен інтернет-сленгу ще недостатньо охоплений у межах академічної лексикографії. Особливо відчувається брак структурованих баз, які б відображали актуальні мовні тенденції в онлайн-дискурсі.

Аналіз останніх досліджень і публікацій. Останні дослідження свідчать про зростання наукового інтересу до сленгу в соціальних мережах. Наприклад, у роботі Айзенштейн Д. (2022) розглядаються зміни у частотності сленгових слів та їх стабільність у порівнянні зі стандартною лексикою [6]. Дослідники Нітхія Р. та Арок М. (2023) у своєму огляді акцентують на ролі сленгу у соціальних мережах як засобі емоційного вираження та групової ідентифікації [7]. Дослідження Рахман О. С. та ін. (2022) показує, як тематичне моделювання дозволяє виявити структуру й особливості інтернет-сленгу [9]. Також варто згадати дослідження Вах'юні С. (2023), яка аналізує жіночий сленг серед інфлюенсерів покоління Z [10].

Мета і завдання дослідження. Метою роботи є створення та аналіз частотного словника сучасного сленгу на основі аналізу текстів із соціальних мереж Twitter (X) або Reddit. Для досягнення цієї мети було здійснено попередню обробку текстів: очищення, токенизацію, лематизацію, визначено сленгові одиниці за допомогою маркування, розраховано частотність уживання кожної сленгової одиниці, створено частотний словник із супровідними коментарями та проведено аналіз виявлених одиниць за лексико-семантичними та соціолінгвістичними ознаками.

Виклад основного матеріалу. Сленг як лінгвістичне явище відіграє важливу роль у сучасній комунікації, особливо в інтернет-просторі, де нові слова та вирази виникають і поширюються з надзвичайною швидкістю. Цей феномен є особливим пластом нелітературної лексики, що характеризу-

ється експресивністю, емоційністю та швидкою змінністю. Сленг виникає та функціонує переважно в усному мовленні, хоча з поширенням соціальних мереж набуває значного поширення і в писемній комунікації, зберігаючи при цьому характеристики усного мовлення. Дослідження сленгу набуває особливої актуальності в сучасній лінгвістиці через його тісний зв'язок із соціальними процесами та його роль як індикатора мовних змін і тенденцій у суспільстві. Сленг також відображає особливості культури та світогляду певної соціальної групи, виконуючи функцію групової ідентифікації та встановлення меж між «своїми» та «чужими».

Аналіз лінгвістичних особливостей сленгу в соціальних мережах дозволяє виокремити певні тенденції, які відображають специфіку цього типу комунікації. Однією з найпомітніших рис є змішування кодів та гібридизація мови, що проявляється в активному використанні англіцизмів, їхньому поєднанні з елементами інших мов, створенні нових лексичних одиниць на основі англословних коренів. Цей процес пов'язаний з глобалізацією інтернет-простору та домінуванням англійської мови в технологічній сфері, що впливає на розвиток сленгу в інших мовах. Ще однією характерною рисою є тяжіння до економії мовних засобів, що виявляється в поширенні скорочень, аббревіацій, акронімів та усічених форм слів. Ця тенденція зумовлена технічними обмеженнями соціальних платформ, а також прагненням до швидкої передачі інформації. Креативні словотворчі процеси, які часто порушують традиційні мовні норми, також становлять визначальну рису сленгу в соціальних мережах, відображаючи мовну гру, іронію та прагнення до оригінальності [7, с. 80].

Особливої уваги заслуговує аналіз мовної специфіки сленгу як лінгвокультурного феномену. На фонетичному рівні сленг часто характеризується порушенням традиційних норм вимови, що в письмовій формі відображається через специфічну орфографію, наприклад, навмисні спотворення слів (wut, ya, gonna), подвоєння або потроєння літер для передачі емоційного стану (sooooo, soooool), використання великих літер для емоційного виділення (AMAZING). На морфологічному рівні сленгові одиниці демонструють високу продуктивність певних словотворчих моделей, зокрема суфіксації (-er: gamer, streamer, influencer), префіксації (un-: unfriend, unfollow),

основоскладання (photobomb, doomscroll) та аббревіації (IMHO, LOL, BRB). На лексико-семантичному рівні для сленгу характерні процеси метафоризації, метонімії, а також розширення або звуження значення існуючих слів. Синтаксичні особливості сленгу проявляються в тенденції до спрощення конструкцій, еліптичності, використання емотиконів та інших графічних елементів, що замінюють цілі синтаксичні конструкції [8, с. 72].

Сучасні соціальні мережі оснащені різноманітними інструментами та функціями, які впливають на комунікативну поведінку користувачів та, відповідно, на їхній мовний вибір. Система хештегів у Twitter (X) та Instagram сприяє тематичній організації контенту та створенню нових сленгових виразів на основі популярних тегів. Функція репосту (ретвіту) забезпечує швидке поширення нових сленгових одиниць серед широкої аудиторії, а система коментарів створює простір для обговорення та колективного творення сленгу.

Аналіз корпусу текстів із соціальних мереж показав, що найбільш частотними є сленгові одиниці, пов'язані з вираженням емоційних реакцій та оцінок (LOL – laughing out loud, wow, cringe, facepalm), оскільки вони дозволяють компенсувати відсутність невербальних засобів комунікації в цифровому середовищі. Також високу частотність демонструють сленгізми, що позначають дії користувачів та функції соціальних мереж (to tweet, to post, to share), що пов'язано з необхідністю номінації нових реалій цифрового світу.

Процес збору та відбору мовного матеріалу зі соціальних мереж Twitter (X) та Reddit вимагає комплексного методологічного підходу, що передбачає визначення чітких критеріїв відбору, використання спеціалізованих технічних інструментів та дотримання етичних і правових норм дослідження. Формування репрезентативного корпусу текстів із вказаних соціальних мереж є фундаментальним етапом, що визначає надійність і валідність результатів усього дослідження мовних інновацій. Особливої уваги при плануванні збору даних потребує врахування структурних і функціональних відмінностей між Twitter та Reddit, які впливають як на характер комунікації в цих мережах, так і на технічні можливості отримання текстових даних. Ефективна стратегія формування дослідницького корпусу має забезпечувати збалансований відбір матеріалу, що відображає різноманітні дискурсивні практики, тематичні сфери та комунікативні ситуації, характерні для сучасного україномовного сегменту цих платформ. Розробка такої стратегії передбачає визна-

чення оптимального обсягу вибірки, часових меж дослідження та релевантних лінгвістичних характеристик, за якими здійснюватиметься аналіз текстових даних.

У контексті даної роботи основними критеріями відбору текстів із Twitter (X) та Reddit були визначені: мова повідомлень (англійська), хронологічні межі (повідомлення, опубліковані впродовж 2022–2025 років), тематична релевантність (повідомлення, що містять потенційні лексичні інновації в сферах технологій, соціальних явищ, політики та культури), популярність контенту (кількість взаємодій з повідомленням у формі лайків, репостів, коментарів), а також оригінальність тексту (перевага надавалася авторським дописам, а не автоматично згенерованим повідомленням чи рекламному контенту). Важливим критерієм також стало врахування різноманітності джерел – для забезпечення широкого охоплення різних дискурсивних спільнот було відібрано повідомлення від користувачів з різною кількістю підписників у Twitter та з різних тематичних сабредітів у Reddit, що дозволило дослідити як загальноживані неологізми, так і вузькоспеціалізовані лексичні інновації, характерні для окремих віртуальних спільнот. Додатковим критерієм стала наявність у повідомленнях хештегів, що полегшило категоризацію контенту за тематичними групами та відстеження поширення конкретних лексичних одиниць у різних контекстах [4, с. 52].

Для забезпечення збалансованого представлення різних тематичних сфер, у яких функціонують лексичні інновації, було визначено десять основних категорій контенту для цілеспрямованого збору даних: технології та цифровізація, суспільно-політичні процеси, культура та мистецтво, здоров'я та медицина, освіта та наука, економіка та бізнес, спорт та фізична активність, розваги та дозвілля, міжособистісні стосунки, навколишнє середовище та екологія. Для кожної категорії було сформовано перелік ключових слів та хештегів, за якими здійснювався пошук релевантних повідомлень через API платформ. Додатково для Reddit було визначено тематичні сабредіти, що відповідають кожній категорії, з яких збиралися дописи та коментарі. Такий підхід дозволив дослідити функціонування лексичних інновацій у різних дискурсивних контекстах, визначити тематичні сфери з найвищою неогенною активністю та порівняти особливості словотворення в різних фахових та інтересових спільнотах. Для кожної тематичної категорії було зібрано порівнянну кількість текстових даних, що забезпечило збалансованість корпусу та можливість здійснення

коректних кількісних порівнянь між різними дискурсивними сферами [2, с. 98].

У таблиці 1 представлено розподіл зібраного мовного матеріалу за платформами та тематичними категоріями, що демонструє збалансованість сформованого корпусу та його репрезентативність для дослідження лексичних інновацій у різних комунікативних сферах.

Завершальним етапом підготовки мовного матеріалу для дослідження стало створення спеціалізованої бази даних неологізмів, виявлених у зібраному корпусі. Для кожної лексичної інновації було створено окремий запис з детальною лінгвістичною характеристикою, що включає: графічну форму (варіанти написання), морфологічні характеристики, словотвірну модель, семантичне значення, приклади контекстуального вживання, дата першої фіксації в корпусі, частотність вживання в різних тематичних категоріях, динаміку поширення протягом досліджуваного періоду. Додатково для запозичень фіксувалася мова-джерело та ступінь адаптації до системи української мови. Така структурована база даних стала основою для подальшого комплексного аналізу лексичних інновацій, виявлення продуктивних словотворчих моделей та визначення основних тенденцій розвитку лексичної системи української мови в контексті цифрової комунікації. На підставі створеної бази даних було також укладено частотний словник неологізмів соціальних мереж, що відображає актуальний стан неогенних процесів в українському інтернет-дис-

курсі та може стати цінним ресурсом для подальших лінгвістичних досліджень [3, с. 59].

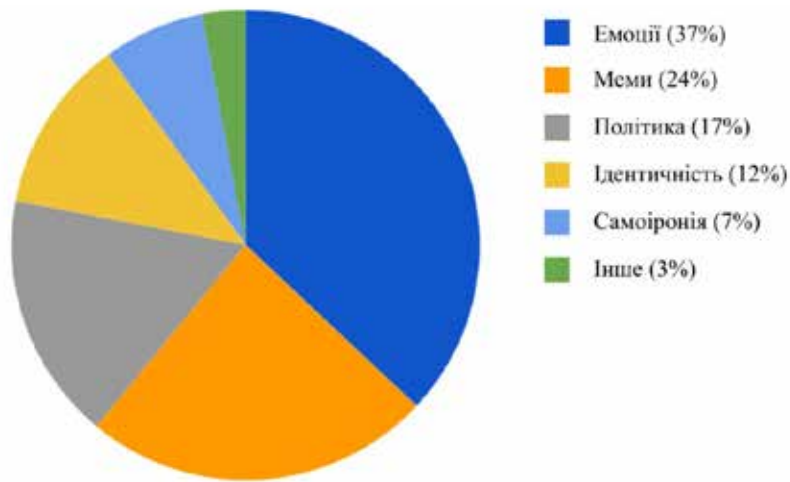
Аналіз частотного розподілу сленгових одиниць за тематичними групами, представлений на рис. 2, чітко демонструє домінування емоційно-орієнтованої лексики в досліджуваному корпусі. Суттєва перевага цієї категорії (37%) над іншими тематичними групами підтверджує первинну комунікативну функцію сленгу в соціальних мережах – експресивне висловлення особистого ставлення до подій, явищ та інших користувачів. Другу позицію посідає мем-орієнтована лексика (24%), значна частка якої свідчить про важливість гумористичного компонента в онлайн-комунікації та підкреслює роль соціальних мереж як середовища формування та поширення новітнього фольклору цифрової епохи. Суспільно-політична тематична група, що охоплює 17% від загального корпусу, відображає актуальність соціально-політичного дискурсу в українському сегменті соціальних мереж та демонструє тенденцію до вербалізації громадянської позиції через специфічні мовні засоби. Порівняно менша частка лексем, пов'язаних з ідентичністю (12%) та самоіронією (7%), корелює з обмеженістю відповідних комунікативних ситуацій у повсякденному спілкуванні користувачів.

Графічне представлення динаміки популярності традиційних та нових сленгових одиниць, наведене на Рисунку 3.2, візуалізує протилежні тренди в еволюції онлайн-лексики протягом останніх років. На діаграмі чітко прослідковується

Таблиця 1

Розподіл зібраного мовного матеріалу за платформами та тематичними категоріями

Тематична категорія	Twitter (X)		Reddit		Загалом
	Кількість повідомлень	Обсяг (слововживань)	Кількість дописів/коментарів	Обсяг (слововживань)	Обсяг (слововживань)
Технології та цифровізація	3850	92400	2250	112500	204900
Суспільно-політичні процеси	4120	98880	1870	93500	192380
Культура та мистецтво	2730	65520	1630	81500	147020
Здоров'я та медицина	2140	51360	1280	64000	115360
Освіта та наука	1960	47040	1520	76000	123040
Економіка та бізнес	2310	55440	1340	67000	122440
Спорт та фізична активність	1870	44880	1120	56000	100880
Розваги та дозвілля	2950	70800	1980	99000	169800
Міжособистісні стосунки	1580	37920	1240	62000	99920
Навколишнє середовище та екологія	1490	35760	770	38500	74260
Загалом	25000	600000	15000	750000	1350000



Джерело: за результатами авторського дослідження

Рис. 1. Частотний розподіл сленгових одиниць за тематичними групами

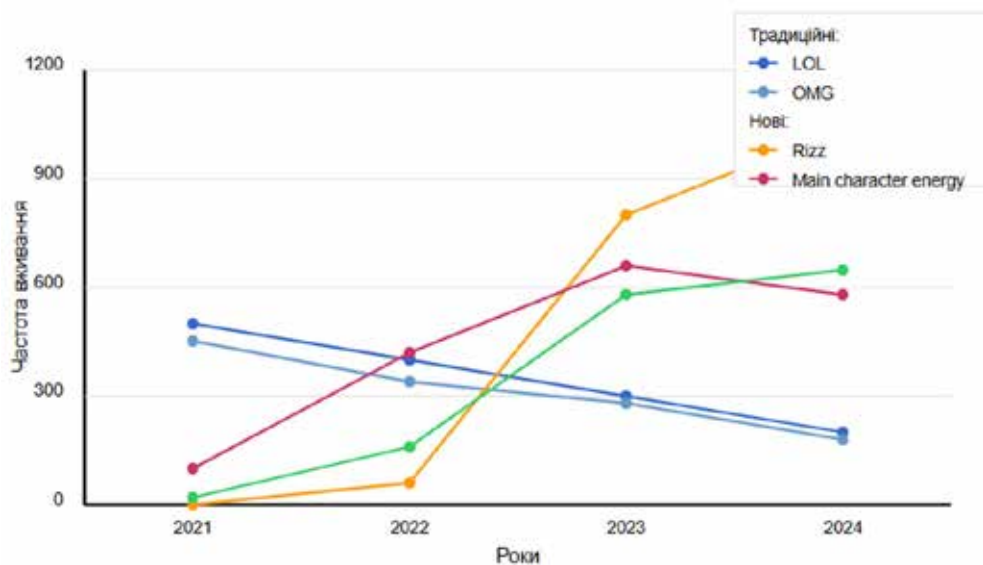


Рис. 2. Порівняльна динаміка популярності традиційних та нових сленгових одиниць

феномен «перехресної динаміки», коли зниження частотності традиційних сленгових акронімів (LOL, OMG) відбувається практично синхронно зі зростанням популярності нових лексичних одиниць (*“rizz”*, *“main character energy”*, *“soft launch”*). Особливий інтерес викликає форма кривої для виразу *“main character energy”*, яка після періоду активного зростання демонструє невелике зниження частотності вживання у 2024 році, що може свідчити про початок циклу «згасання» популярності цієї сленгової одиниці. Аналізуючи загальну конфігурацію графіків, можна спостерігати закономірну картину циклічності мовної моди, коли старі форми поступово витісняються новими, які, своєю чергою, також проходять певні

етапи популярності – від стрімкого зростання до стабілізації або поступового зниження частотності вживання [17, с. 94].

Аналіз даних частотних словників соціальних мереж розкриває унікальні особливості формування та функціонування сленгу в англомовному цифровому просторі. Таблиця 1 демонструє комплексне дослідження лексичного складу англійського сленгу в соціальних мережах за 2023–2024 роки.

Представлені дані ілюструють надзвичайно динамічну природу англійського сленгу в соціальних мережах, де значна частка лексики припадає на молодіжні неологізми та інтернет-жаргонізми. Особливо показовою є висока частка питомої анг-

Розподіл лексики за походженням у словнику сленгу англомовного сегменту соціальних мереж (2023–2024)

Походження лексики	Кількість одиниць	Відсоток від загального обсягу
Питома англійська лексика	423	42.3%
Молодіжні неологізми	389	38.9%
Інтернет-жаргонізми	112	11.2%
Запозичення з інших мов	76	7.6%
Загальна кількість	1000	100%

лійської лексики (42,3%), що свідчить про потужний внутрішньомовний потенціал словотворення.

Висновки з даного дослідження і перспективи подальших розвідок в цьому напрямку. Дослідження засвідчило високу продуктивність поєднання лінгвістичних методів з інструментами обробки великих обсягів даних. У майбутньому подібні підходи можуть бути розширені, напри-

клад, шляхом включення аналізу сленгу в інших соціальних платформах (Instagram, TikTok) або за рахунок вивчення візуального компоненту (меми, гіфки, емодзі) як носіїв сленгового значення. Робота відкриває нові перспективи для інтердисциплінарних досліджень, де мова є не лише засобом комунікації, а й ключем до розуміння цифрової культури.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

1. Булгакова О. В. Молодіжний сленг у сучасному українському медіапросторі. Наукові записки Національного університету “Острозька академія”. Серія: Філологія. 2020. Т. 8. № 76. С. 35–38.
2. Кочубей Л. І. Структурно-семантичні особливості сленгової лексики англомовного інтернет-дискурсу. Молодий вчений. 2018. № 1. С. 106–110.
3. Наконечна Т. Особливості використання сленгової лексики в інтернет-дискурсі. Молодий вчений. 2020. № 3. С. 327–330.
4. Crystal D. Language and the Internet. Cambridge : Cambridge University Press, 2001. 272 p.
5. Danet B., Herring S. C. The Multilingual Internet: Language, Culture, and Communication Online. Oxford : Oxford University Press, 2007. 384 p.
6. Eisenstein J. What to do about bad language on the internet. Proceedings of NAACL. 2013. P. 359–369.
7. Nithya R., Arock M. A systematic literature review on social media slang analytics in contemporary discourse. IJRIS. 2023. URL: <https://www.researchgate.net/publication/375735233> (дата звернення: 24.05.2025).
8. Keidar D. та ін. Slangvolution: a causal analysis of semantic change and frequency dynamics in slang. Proceedings of the 60th Annual Meeting of the ACL. 2022. URL: <https://aclanthology.org/2022.acl-long.101> (дата звернення: 24.05.2025).
9. Rahman O. S. та ін. Semantic Shift of English Internet Slangs Used in Social Media: Morphosemantic Analysis. Research and Innovation in Language Learning. 2022. URL: <https://jurnal.ugj.ac.id/index.php/RILL/article/view/6387> (дата звернення: 24.05.2025).
10. Wahyuni S. The use of slang by Gen Z female influencers on Instagram and Twitter (X). IJRIS. 2023. URL: <https://rsisinternational.org/journals/ijriss/articles/the-use-of-slang-by-gen-z-female-influencers-on-instagram-and-twitter-x> (дата звернення: 24.05.2025).