

УДК 81'33

DOI <https://doi.org/10.32782/2410-0927-2023-19-3>

Світлана ДРУЖБЯК

кандидат філологічних наук, доцент кафедри прикладної лінгвістики, Національний університет «Львівська Політехніка», вул. С. Бандери, 12, м. Львів, Україна, 79000

ORCID: 0000-0001-8971-9852

Аліна ГАЙМАХЕР

студентка магістратури кафедри прикладної лінгвістики, Національний університет «Львівська Політехніка», вул. С. Бандери, 12, м. Львів, Україна, 79000

ORCID: 0009-0000-1793-7789

Бібліографічний опис статті: Дружбяк, С., Гаймахер, А. (2023). Тональний аналіз колокацій із компонентом «перемога». *Актуальні питання іноземної філології*, 19, 14–19, doi: <https://doi.org/10.32782/2410-0927-2023-19-3>

ТОНАЛЬНИЙ АНАЛІЗ КОЛОКАЦІЙ ІЗ КОМПОНЕНТОМ «ПЕРЕМОГА»

У статті розглянуті методи аналізу тональності, які в останні роки стали дуже популярними, і цей інтерес охопив не лише науковців, але й компанії, уряди та різноманітні організації. В рамках дослідження проаналізовано кілька суттєвих наукових праць, які стосуються цього питання.

Дослідження присвячене аналізу тональності колокацій із компонентом «перемога», які були вилучені з корпусу сучасних періодичних видань з ГРАК на базі програмного забезпечення Sketch Engine. Загалом було зібрано 70 словосполук, які охоплюють період війни з лютого 2022 року. Тональний аналіз проводився за допомогою Google Cloud Natural Language API. Це потужний та дуже простий у використанні інструмент, який використовує передові можливості машинного навчання, що розроблені Google, для аналізу та розуміння структури та значення тексту. Він має ряд функцій з обробки природної мови, включаючи аналіз настрою, розпізнавання об'єктів та синтаксичний аналіз. Це дає можливість розробникам створювати додатки, які здатні обробляти та аналізувати великі обсяги тексту в режимі реального часу. API надає оцінку тональності та величину для кожного фрагмента тексту, причому оцінка варіюється від -1 (негативна) до 1 (позитивна).

Було виявлено домінування позитивної тональності на рівні 66%. Основні фрази, які відображають позитивну оцінку, включають «наша перемога», «велика перемога», «нова перемога». Це свідчить про оптимізм та віру в перемогу серед українського населення. Також були виявлені словосполучення, які отримали нейтральну оцінку та рідкісні випадки, коли слово «перемога», пов'язане з негативними емоціями. Також нами було проведено класифікацію колокацій за допомогою функції Text Moderation.

Це дослідження допомагає краще розуміти сприйняття та відношення до поняття «перемога» в українському суспільстві під час російсько-української війни.

Ключові слова: аналіз тональності, медійний дискурс, колокації, ГРАК, Google Cloud Natural Language API.

Svitlana DRUZHBIAK

Candidate of Philological Sciences, Associate Professor at the Department of Applied Linguistics, Lviv Polytechnic National University, 12 Stepana Bandery str., Lviv, Ukraine, 79000

ORCID: 0000-0001-8971-9852

Alina HAIMAKHER

Master's student at the Department of Applied Linguistics, Lviv Polytechnic National University, 12 Stepana Bandery str., Lviv, Ukraine, 79000

ORCID: 0009-0000-1793-7789

To cite this article: Druzhbiak, S., Haimakher, A. (2023). Tonalnyi analiz kolokatsii iz komponentom «peremoha» [Tone analysis of collocations with the component «перемога»]. *Current Issues of Foreign Philology*, 19, 14–19, doi: <https://doi.org/10.32782/2410-0927-2023-19-3>

TONE ANALYSIS OF COLLOCATIONS WITH THE COMPONENT «ПЕРЕМОГА»

This article discusses methods of tone analysis, which have become very popular in recent years, and this interest has spread not only to academics but also to companies, governments, and various organizations. The study analyzes several significant scientific works related to this issue.

The study is devoted to the tone analysis of collocations with the component "victory", which were extracted from the corpus of modern periodicals from the GRAC based on the Sketch Engine software. A total of 70 collocations were collected, covering the period of war since February 2022. Tone analysis was conducted using the Google Cloud Natural Language API. This is a powerful and very easy-to-use tool that uses advanced machine learning capabilities developed by Google to analyze and understand the structure and meaning of text. It has a number of natural language processing features, including sentiment analysis, object recognition, and parsing. This enables developers to create applications that can process and analyze large amounts of text in real time. The API provides a tone score and magnitude for each piece of text, with the score ranging from -1 (negative) to 1 (positive).

The dominance of a positive tone was found to be 66%. The main phrases reflecting a positive assessment include «науа перемога», «велика перемога», «нова перемога». This indicates optimism and belief in victory among the Ukrainian population. We also identified word combinations that received a neutral assessment and rare cases when the word «перемога» is associated with negative emotions. We also classified collocations using the Text Moderation function.

This study helps to better understand the perception of and attitudes toward the concept of «перемога» in Ukrainian society during the Russian-Ukrainian war.

Key words: tone analysis, media discourse, collocations, GRAC, Google Cloud Natural Language API.

Актуальність проблеми. У сучасній епосі розвитку інформаційних технологій спостерігається значне збільшення різноманітних джерел інформації для суспільства. Вони ставлять перед собою завдання не лише повідомляти про новини і надавати повну картину подій у світі, але й активно привертати увагу громадськості, частково впливати на свідомість і формувати уявлення про них. Повномасштабна війна між Україною та Росією супроводжується активною інформаційною війною, в якій ЗМІ мають ключову роль. Слово «перемога» є одним з центральних понять і виступає в якості ключового тематичного елемента у матеріалах, які публікуються в періодичних виданнях. Одним із способів вивчення сприйняття цього поняття є аналіз колокацій. Мовознавчі джерела свідчать, що вивчення цього лінгвістичного аспекту має довгу історію, яка триває понад століття. Результати досліджень колокацій використовують в розробці сучасних лексикографічних систем, автоматичного аналізу тексту, в машинному перекладі (Бобкова, 2014). Вони можуть бути як нейтральними, так і мати позитивний або негативний відтінок. Відповідно до цього, аналіз мовної тональності колокацій зі словом «перемога» в контексті повномасштабної війни набуває великої актуальності, оскільки він допомагає краще розуміти вплив ЗМІ на сприйняття подій та визначати, як це ключове слово відображає реалії воєнного конфлікту.

Аналіз останніх досліджень і публікацій. Питанням виділення колокацій з текстового матеріалу та встановлення критеріїв їхньої

ідентифікації присвячені праці таких авторів, як Т. Бобкова, М. Хохлова, В. Захаров, В. Гладка.

Ми були зацікавлені у методиці тонального аналізу (сентимент-аналіз), основним завданням якого є класифікація полярності тексту, тобто визначення, чи є виражена в тексті думка позитивною, негативною або нейтральною.

Протягом останнього десятиріччя сентимент-аналіз став дуже популярним у галузі обробки природної мови, що підтверджується великою кількістю проведених досліджень у цій галузі. На сьогодні він застосовується в різних галузях, таких як соціологія, політологія та маркетинг.

Коли дослідження в галузі сентимент-аналізу ще тільки зароджувалися, зміст і тематика опитувань в основному зосереджувалися на завданнях, деталізації та сферах застосування. Кумар розглянув основні терміни, завдання та рівні деталізації, пов'язані з аналізом тональності (Kumar, 2012).

В Україні спостерігаємо активний розвиток цієї галузі, перш за все для аналізу тональності україномовних відгуків. Крім цього, здійснюється робота над емоційно-смісловим аналізатором української мови. Також вже почато створення сентимент-анотованого корпусу, на основі якого розробляють тональний словник (Романишин, 2013). Українські дослідники, які в своїх роботах розглядають методи обробки текстів на природній мові, описують програмні застосунки та проводять сентимент-аналіз, базуючись на різних підходах та концентруючись на різних аспектах: О. Немеш (2015), Н. Шаховська (2018), А. Романюк (2013), Н. Дарчук (2019).

Мета пропонованого дослідження – за допомогою аналізатора Google Cloud Natural Language API здійснити аналіз тональності колокацій із компонентом «перемога», які були вилучені з корпусу періодичних видань.

Виклад основного матеріалу. Джерельною базою цього дослідження послуговував Генеральний регіонально анотований корпус української мови (ГРАК). За допомогою програмного забезпечення Sketch Engine було створено підкорпус інтернет-видань з лютого 2022 року та створено списки колокацій.

ГРАК – це велика, репрезентативна, структурована колекція текстів українською мовою у супроводі програми, яка дозволяє будувати на базі корпусу власні підкорпуси, шукати слова, граматичні форми та їх сполучення, а також обробляти результати пошуку, сортувати, робити збалансовані вибірки і одержувати різну статистичну інформацію. Корпус призначений для здійснення лінгвістичних досліджень з граматики, лексики, історії української літературної мови, а також для використання під час укладання словників та граматик (ГРАК).

Тональний аналіз, або сентимент-аналіз – це різновид аналізу текстів, метою якого є виявлення емоційного ставлення або суб'єктивних суджень автора щодо певного об'єкта в текстовому повідомленні. Основним завданням аналізу є автоматична оцінка певного об'єкта (людини, повідомлення в ЗМІ, події, організації тощо) в текстовому повідомленні з метою отримання числового або категоріального показника. В. Liu у своєму дослідженні зауважує, що існують синонімічні поняття та терміни для опису цього аналізу, наприклад: «*opinion mining*» (аналіз думок, розвідка думок), «*emotional analysis*» (аналіз емоцій), «*brand monitoring*» (бренд-моніторинг). На його думку, в академічних колах найчастіше використовують *opinion mining* та *sentimental analysis* (Liu, 2012). Термін «*sentiment analysis*» вперше був згаданий в роботі Nasukawa T. та Yi J., «*opinion mining*» (аналіз думок) в роботі Dave K.

Онищенко І. зазначає, що при аналізі тональності слід звернути увагу на чотири елементи: 1) суб'єкт тональності (автора повідомлення); 2) об'єкт тональності (про що йде мова); 3) аспект тональності (характеристику об'єкта); 4) оцінку тональності – сентимент (Онищенко, 2004). Дарчук Н. додає, що велике

значення для характеристики тональності має жанр текстів. Наприклад, в інформаційних (новинних) текстах можуть бути множинними як суб'єкти сентименту (автори та їхні оцінки), так і об'єкти, а також ситуації (аспекти), що значно ускладнює процес автоматичного аналізу тональності, як і багатозначність оцінної лексики (Дарчук, 2019).

Реалізація аналізу може бути складним і трудомістким процесом, який часто вимагає глибоких знань методів обробки природної мови (NLP) та алгоритмів машинного навчання. Google Cloud Natural Language API спрощує аналіз настроїв і робить його доступним для будь-кого. Google Cloud Natural Language API – це потужний, простий у використанні інструмент, який використовує передові можливості машинного навчання Google для аналізу та розуміння структури і значення тексту. Він надає ряд функцій NLP, включаючи аналіз настрою, розпізнавання об'єктів і аналіз синтаксису, що дозволяє розробникам створювати додатки, які можуть обробляти і аналізувати великі обсяги тексту в режимі реального часу. Однією з ключових переваг використання Google Cloud Natural Language API для аналізу настроїв є його простота. API надає оцінку настрою та величину для кожного фрагмента тексту, причому оцінка варіюється від –1 (негативна) до 1 (позитивна), а величина відображає загальну силу настрою.

Важливо зазначити, що API природної мови вказує на відмінності між позитивними і негативними емоціями в документі, але не ідентифікує конкретні позитивні і негативні емоції. Наприклад, «злий» і «сумний» вважаються негативними емоціями. Однак, коли API природної мови аналізує текст, який вважається «злим», або текст, який вважається «сумним», відповідь лише вказує на те, що почуття в тексті є негативним, а не «сумним» або «злим».

Текст з нейтральною оцінкою може вказувати на низький рівень емоцій, а може вказувати на змішані емоції, з високими позитивними і негативними значеннями, які нівелюють одне одного. Зазвичай, для розрізнення цих випадків можна використовувати значення величини емоцій, оскільки дійсно нейтральні документи матимуть низьку величину, тоді як змішані документи матимуть вище значення величини.

Нами було проаналізовано 70 колокацій. Наводимо результати, вказуючи індекс тональності в дужках:

1. Позитивна оцінка (66% колокацій): наша перемога (0,86), велика перемога (0,86), нова перемога (0,86), перемога добра (0,79), величезна перемога (0,78), перемога та мир (0,75), повна перемога (0,73), швидка перемога (0,70), майбутня перемога (0,69), перемога України (0,68), вірити в перемогу (0,67), тактична перемога (0,67), перемога українців (0,65), вклад у перемогу (0,65), перша перемога (0,64), перемога ЗСУ (0,63), маленька перемога (0,62), символ перемоги (0,6), вольова перемога (0,57), чекати перемогу (0,56), українська перемога (0,56), бажати перемоги (0,54), настане перемога (0,53), неминуча перемога (0,51), спільна перемога (0,505), остаточна перемога (0,485), принести перемогу (0,46), воля до перемоги (0,44), військова перемога (0,44), важлива перемога (0,42), разом до перемоги (0,40), святкувати перемогу (0,40), крок до перемоги (0,39), якнайшвидша перемога (0,38), впевненість у перемозі (0,38), перемога армії (0,37), працювати на перемогу (0,34), перемога над росією (0,34), шлях до перемоги (0,33), перемога на фронті (0,32), перемога у війні (0,29), пришвидшити перемогу (0,29), здобути перемогу (0,27), відзначати перемогу (0,27), отримати перемогу (0,26), прискорити перемогу (0,25).

Найвищий індекс в колокаціях «наша перемога», «велика перемога», «нова перемога», «перемога та мир», тому вважаємо доцільним навести приклади їх вживання в межах корпусу, наприклад:

*Дуже хочеться вірити, що завдяки всім вам скоро буде **наша перемога**.*

*Це, безумовно, **велика перемога** для нашої держави, для України.*

*Сьогодні всі ми віримо в **нову перемогу**.*

*Головне зараз – **перемога та мир**.*

*Нехай у цей день **перемога та мир** наближаються.*

2. Нейтральна оцінка (31%): внесок у перемогу (0,24), наближати перемогу (0,235), потрібна перемога (0,23), наближення перемоги (0,22), шанси на перемогу (0,19), заради перемоги (0,18), задля перемоги (0,18), перемога на полі бою (0,18), наблизити перемогу (0,18), закінчитися перемогою (0,16), необхідна перемога (0,16), забезпечити перемогу (0,13), перемога над ворогом (0,13), перемога над агресором (0,12), прагнення до перемоги (0,12), оголосити перемогу (0,12), вибороти перемогу (0,10), вирвати перемогу (0,10), прогнозувати перемогу (0,10), після перемоги (0,09), перемога Києва (0,06).

3. Негативна оцінка (3%): перемога росії (-0,30), перемога путіна (0,54), наприклад:

***Перемога росії** призведе до подальшої російської агресії.*

*Ми, українці, добре знаємо, як виглядатиме **перемога росії**, тому що знаємо, як виглядає сьогодні російська окупація.*

*Нарешті лукашенко розуміє, що **перемога путіна** в цій війні вже зовсім не гарантована, пише видання, тож білоруський диктатор намагається перевзуватись у повітрі.*

*Столтенберг: **перемога путіна** в Україні означатиме поразку НАТО (рис. 1).*



Рис. 1. Результати аналізу тональності колокацій

За допомогою функції Text moderation можна проаналізувати текст відповідно до категорій, які нам пропонують. Всі досліджувані колокації були віднесені до категорії War&Conflict (війна, військові конфлікти та великі фізичні конфлікти за участю великої кількості людей) та оцінено від 1 до 0. Найбільшу оцінку отримали такі колокації: військова перемога (0,99), перемога армії (0,99), перемога над росією (0,99), перемога у війні (0,99), перемога України (0,97), тактична перемога (0,97), перемога ЗСУ (0,97), українська перемога (0,97), перемога росії (0,97), здобути перемогу (0,96), перемога Києва (0,96), перемога путіна (0,96), прискорити перемогу (0,95), після перемоги (0,95).

Також деякі колокації мали високий бал в інших категоріях, наприклад Politics (політичні новини та медіа; обговорення соціальної, урядової та державної політики), Legal, Public Safety.

Висновки і перспективи подальших досліджень. Аналіз 70 колокацій із компонентом «перемога» на основі сучасних періодичних видань під час російсько-української війни виявив домінування позитивної тональності (66%). Основні ключові фрази, які відображають позитивну тональність, містять «наша перемога», «велика перемога», «нова перемога». Це свідчить про оптимізм і віру в перемогу серед українського населення. Існують також словосполучення, які відзначають нейтральну оцінку (31%) та рідкісні випадки, де перемога пов'язана з негативними емоціями (3%), зокрема, у відношенні до росії та путіна.

Тональний аналіз, який об'єднує теорії емоційних станів з конкретними завданнями обробки природної мови, робить комп'ютерне мислення більш схожим на людське. Цей підхід важливий для розвитку штучного інтелекту та відкриває значні можливості для подальших досліджень природної мови.

ЛІТЕРАТУРА:

1. Бобкова Т. Теоретико-методологічні підходи до вивчення колокацій у сучасному мовознавстві. *Вісник Київського національного лінгвістичного університету. Серія «Філологія»*. 2014. Т. 17. № 2. С. 14–22.
2. Генеральний регіонально анотований корпус української мови (ГРАК) / М. Шведова, Р. фон Вальденфельс, С. Яригін, А. Рисін, В. Старко, Т. Ніколаєнко та ін. Київ, Львів, Єна, 2017–2023. – uacorpus.org.
3. Дарчук Н. Лінгвістичні засади автоматичного сентимент аналізу українськомовного тексту. *Science and education a new dimension*. 2019. № 189. С. 10–13.
4. Немеш О., Романюк А., Теслюк В. Аналіз тональності тексту: основні поняття та приклади застосування // *Людина. Комп'ютер. Комунікація: зб. наук. праць*. Львів: Видавництво Львівської політехніки, 2015. С. 47–49.
5. Онищенко І. В. Категорія оцінки та засоби її вираження в публіцистичних та інформаційних текстах: автореф. дис. ... канд. філол. наук: 10.02.01. Дніпро, 2004. 22 с.
6. Романишин М., Романюк А. Тональний словник української мови на основі сентимент-анотованого корпусу. *Українське мовознавство*. 2013. Вип. 43. С. 63–74.
7. Шаховська Н., Шаховська Х. Метод аналізу відгуків клієнтів з природномовних текстів // *Штучний інтелект*. 2018. № 3 (81). С. 18–26.
8. Kumar A, Sebastian TM. Sentiment analysis: a perspective on its past, present and future. *Int J Intell Syst Appl*, 2012. № 4(10), 11–14.
9. Liu B. *Sentiment Analysis and Opinion Mining*. Williston-London: Morgan&Claypool Publishers, 2012. 168 p.
10. Liu B. *Sentiment Analysis: Mining Opinions, Sentiments and Emotions and Subjectivity*. Cambridge: Cambridge University Press, 2015. 384 p.

REFERENCES:

1. Bobkova, T. V. (2014). Teoretyko-metodolohichni pidkhody do vyvchennia kolokatsii u suchasnomu movoznavstvi [Theoretical and methodological approaches to the study of collocation in modern linguistics], *Visnyk Kyivskoho natsionalnoho lnhvistychnoho universytetu. Ser.: Filolohiia*, No. 2, 4–22. [in Ukrainian].
2. Heneralnyi rehionalno anotovanyi korpus ukrainskoi movy (HRAK) / M. Shvedova, R. fon Valdenfels, S. Yaryhin, A. Rysin, V. Stariko, T. Nikolaienko ta in. [General Regionally Annotated Corpus of Ukrainian (GRAC)] – Kyiv, Lviv, Yena, 2017–2023. – uacorpus.org.
3. Darchuk, N. (2019). Lnhvistychni zasady avtomatychnoho sentyment analizu ukrainskomovnoho tekstu [Linguistic approach for development of computer-based sentiment analysis in the Ukrainian language]. *Science and education a new dimension*, 189, 10–13. [in Ukrainian].

4. Nemes, O., Romaniuk, A., Tesliuk, V. (2015). Analiz tonalnosti tekstu osnovni poniattia ta pryklady zastosuvannia [Sentiment analysis: basic concepts and application examples]. Pro-ceedings from: Mizhnarodna naukovo-praktychna konferentsiia «Luiduna. Kompiuter. Komunikaciya» – International scientific and practical conference «Human, computers and communi-cation». (pp. 47–49). Lviv. [in Ukrainian].
5. Onyshhenko I. V. (2004). Kategoriya otsinky ta zasoby yii vyrazhennia v publicystychnykh ta informaciynykh tekstakh: avtoref. dys. na zdobuttia nauk. stupenya kand. filol. nauk: spets. 10.02.01. [The category of evaluation and the means of its expression in journalistic and informational texts] Dnipropetrovsk, 24. [in Ukrainian].
6. Romanyshyn M., Romaniuk A. (2013) .Tonalnyi slovnyk ukrainskoi movy na osnovi sentyment-anotovanoho korpusu.[A tonal dictionary of the Ukrainian language based on a sentiment-annotated corpus.] Ukrainske movoznavstvo. № 43, 63–74. [in Ukrainian].
7. Shakhovska N., Shakhovska Kh. (2018) Metod analizu vidhukiv kliientiv z pryrodnomovnykh tekstiv[A method for analyzing customer reviews from natural language texts] // Shtuchnyi intelekt. № 3 (81), 18–26. [in Ukrainian].
8. Kumar A, Sebastian TM .(2012) Sentiment analysis: a perspective on its past, present and future. Int J Intell Syst Appl. № 4(10), 11–14.
9. Liu B. (2012) Sentiment Analysis and Opinion Mining. Williston-London: Morgan&Claypool Publishers, 2012. 168 p.
10. Liu B. (2015). Sentiment Analysis: Mining Opinions, Sentiments and Emotions and Subjectivity. Cambridge: Cambridge University Press. 384 p.