

РОЗДІЛ 8

СТРУКТУРНА, ПРИКЛАДНА ТА МАТЕМАТИЧНА ЛІНГВІСТИКА

УДК 81'33

DOI <https://doi.org/10.32782/tps2663-4880/2022.25.1.43>

ТОНАЛЬНИЙ АНАЛІЗ ЗАГОЛОВКІВ BBC ПРО ВІЙНУ В УКРАЇНІ (ЗА РЕЗУЛЬТАТАМИ IBM WATSON TONE ANALYZER)

TONE ANALYSIS OF BBC HEADLINES ABOUT THE WAR IN UKRAINE (USING IBM WATSON TONE ANALYZER)

Бадецька О.Й.,

*orcid.org/0000-0003-2325-7557**магістр мистецтв кафедри прикладної лінгвістики
Львівської національної політехніки*

Таран О.С.,

*orcid.org/0000-0001-6312-9563**кандидат філологічних наук,
доцент кафедри прикладної лінгвістики
Львівської національної політехніки*

У статті описано методи аналізу тональності, який останніми роками набув широкого визнання не лише серед дослідників, але й серед компаній, урядів та організацій. Проаналізовано декілька важливих наукових праць із цього питання.

Дослідження присвячене комп'ютерному аналізу тональності заголовків новин про війну в Україні на сайті BBC News. База даних новинних заголовків була створена за допомогою розширення HeadingsMap у браузері Google Chrome, що дозволило автоматизувати цей процес. Загалом було вилучено з сайту 1017 заголовків, які охоплювали період війни з березня по жовтень 2022 року. Подальша робота відбувалася із залученням IBM Watson: IBM Watson™ Tone Analyzer. Цей аналізатор базується на теорії психолінгвістики та використовує лінгвістичний аналіз для виявлення емоційних і мовних тонів у письмовому тексті. Сервіс може аналізувати тон як на рівні документа, так і на рівні речення: n-грами, знаки пунктуації, смайли, лайливі слова, вітання і полярність настроїв вводяться в алгоритми машинного навчання для класифікації категорій емоцій.

Автори описали порядок роботи з програмою, проаналізували заголовки й ті проблеми, які виникають у визначенні тонів. Заголовки-речення, для яких програма визначила емоційні тони, було погруповано відповідно до тональності. У результаті виявлено 6 тонів: аналітичний, впевненості, страху, радості, смутку і невпевненості.

Успішність роботи програми IBM Watson™ Tone Analyzer залежить від того, в прямому чи переносному значеннях ужито слово, від ясності й повноти контексту. Заголовки є мінімальним текстом, а тому програма не може ідентифікувати їх у ширшому контексті, не може врахувати соціокультурний, історичний чи політичний фон. У перспективі – здійснити аналіз тональності заголовків, але в контексті новинних текстів.

Ключові слова: медійний дискурс, аналіз тональності, IBM Watson™ Tone Analyzer, BBC, заголовки.

The article describes the methods of tone analysis, which in recent years has gained wide acceptance not only among researchers, but also among companies, governments, and organizations. Several important scientific works on this issue are analyzed.

The study is devoted to the computer tone analysis of news headlines about the war in Ukraine on the BBC News website. The database of news headings was created using the HeadingsMap extension in the Google Chrome browser, which made it possible to automate this process. A total of 1,017 headlines covering the war period from March to October 2022 were extracted from the site. Further work took place with the IBM Watson: IBM Watson™ Tone Analyzer. This analyzer is based on the theory of psycholinguistics and uses linguistic analysis to detect emotional and linguistic tones in written text. The service can analyze the tone both at the document level and at the sentence level: n-grams, punctuation marks, emoticons, expletives, greetings and mood polarity are entered into machine learning algorithms to classify emotion categories.

The authors described the procedure for working with the software, analyzed headlines and also the problems that arise in determining tones. Headlines, for which the program identified emotional tones, were grouped according to tonality. As a result, 6 tones were revealed: analytical, confidence, fear, joy, sadness and uncertainty.

The success of the IBM Watson™ Tone Analyzer analysis depends on whether the word is used literally or figuratively, and on the clarity and completeness of the context. Headlines are minimal text, so the software cannot identify them in a wider context, cannot take into account the socio-cultural, historical or political background. In the future, it is necessary to analyze the tonality of headlines, but in the context of news texts.

Key words: media discourse, tone analysis, IBM Watson™ Tone Analyzer, BBC, headlines.

Постановка проблеми. Медійний дискурс є об'єктом різноманітних досліджень, зважаючи на специфіку лексики, жанрове розмаїття, вплив на інші дискурси, специфічні синтаксичні структури, зв'язок ЗМІ з соціальними, соціокультурними та економічними відносинами. Як зазначає М. Теравсевић, мова медійного дискурсу використовує широкий спектр риторичних засобів на фонологічному, синтаксичному, лексичному, семантичному, прагматичному та текстовому рівнях [11, с. 76]. Універсальним і, напевне, найважливішим компонентом у структурі медійного дискурсу є заголовок – це початок діалогу між автором і читачем. Новинні заголовки справедливо називають передвісниками новин, адже вони охоплюють значно ширшу аудиторію людей, які переглядають заголовки, ніж тих, хто читає новини, а тому заголовки мають привернути увагу читача та змусити його перейти до повного тексту новин. Саме це, як уважає R. Taiwo, є причиною використання маніпулятивних технологій у заголовках [10, с. 222]. Тональність і експлуатація емоцій у цьому відіграють важливу роль.

Аналіз останніх досліджень і публікацій. Дослідженню ролі заголовків у медійному дискурсі присвячено праці Д. Реа, Р. Тайво, Б. Ендрю, М. Пфау та Е. Різель, з вітчизняних дослідників – А. Коваленко, Л. Грицюк. Заголовки вивчають з погляду сприйняття їх читачами (психолінгвістичний і когнітивний аспекти); за формою (традиційні та клікбейти; заголовки-накази / ствердження / питання / оклики) та змістом (структурні, мовні, текстові особливості); функції заголовків (інформативна, рекламна, підсумовування новин, переконання читача тощо). Характер і предмет дослідження зумовлює науково-методологічний інструментарій.

Нашу увагу привернула методика аналізу тональності (сентимент-аналізу), основним завданням якого є класифікація полярності тексту на рівні документа чи речення – незалежно від того, чи є висловлена думка в документі, реченні позитивною, негативною або ж нейтральною. Останніми роками аналіз тональності та настроїв набув широкого визнання не лише серед дослідників, але й серед компаній, урядів та організацій.

Перше дослідження заголовків у цьому аспекті було здійснене F. Allport & M. Lerpkin у 1943 році: вони вивчали, як заголовки вплинули на ставлення громадян до участі у військовій кампанії Другої світової війни. Так, заголовки про погані новини (наприклад, «Americans Lose 5 Troopships in African Occupation» та «Major Attack on Guadalcanal Opened by Japan») частіше викликали

в читачів новин бажання брати більш активну участь у війні, ніж заголовки про хороші новини (як-от: «Germans Battle Futilely at Stalingrad»). Підняття морального духу сприяли саме негативні заголовки [1, с. 217–19]. У 1990-х роках дослідники часто висловлювали занепокоєння щодо негативу в заголовках [6, с. 133] і побоювання щодо заохочування негативного ставлення до економіки США [2, с. 12–13]. Однак негатив у заголовках сприяв кращому запам'ятовуванню цих заголовків: коли негатив стимулює людей, таке збудження може збільшити кількість інформації, яку вони запам'ятовують [9, с. 260]. Отже, ці факти викликали в нас зацікавлення, що стимулювало здійснити дослідження тональності заголовків новин BBC про російсько-українську 2022 року. Це є метою нашої наукової розвідки.

Методологія дослідження. Методи аналізу тональності E. Cambria, B. Schuller, Y. Xia та C. Navasi поділяють на три основні категорії:

- методи, які базуються на знаннях (класифікують текст на основі наявності однозначних слів-емоцій, таких як *щасливий, сумний, наляканий, нудьгуючий*);
- статистичні методи (використовують елементи машинного навчання, як-от: латентний семантичний аналіз, опорні векторні машини, глибоке навчання);
- гібридні методи (використовують як машинне навчання, так і онтології та семантичні мережі, щоб виявити семантику, яка нечітко виражена) [3].

Останнім часом спостерігаємо зростання кількості завдань аналізу настроїв. Системний огляд методів аналізу тональності наведено в праці A. Lighthart, C. Catal та B. Tekinerdogan [7, с. 5029]. Розгляньмо декілька важливих розвідок щодо цього питання. Y. Volcani та D. Fogel запропонували ідентифікувати слова та фрази в тексті за емоційними шкалами. Система, заснована на їхній роботі, називається EffectCheck, представляє синоніми, які можна використовувати для збільшення або зменшення рівня викликаних емоцій у кожній шкалі [13]. P. Turney застосовував алгоритм для аналізу тональності PMI-IR, що базується на взаємній інформації та інформаційному пошуку для вимірювання подібності пар слів або фраз. Семантична орієнтація фрази обчислюється шляхом порівняння її з позитивним або негативним словом («добре» / «погано») [12]. A. Mudinas, D. Zhang і M. Levene розробили систему аналізу тональності pSenti на рівні понять, яка поєднує лексичний підхід з підходом на основі навчання [8]. H. Wang, Y. Lu, C. Zhai запропонували

модель латентної рейтингової регресії, що вимірює оцінки різних аспектів рецензії, а також визначає акцент автора у кожному аспекті [14]. Отже, методи керованого машинного навчання є широко використовуваною методикою тонального аналізу.

Для нашого дослідження ми обрали онлайн-сервіс IBM Watson™ Tone Analyzer, що використовує лінгвістичний аналіз для виявлення емоційних і мовних тонів у письмовому тексті. Сервіс може аналізувати тон як на рівні документа, так і на рівні речення: n-грами (уніграми, біграми та триграми), знаки пунктуації, смайли, лайливі слова, вітання і полярність настроїв вводяться в алгоритми машинного навчання для класифікації категорії емоцій. [5]. Ця комп'ютерна розробка базується на теорії психолінгвістики, зокрема твердженні про те, що мовлення відображає більше, ніж просто те, що люди хочуть сказати. Частота, з якою вони вживають певні типи слів, може вказувати на їхній стиль мислення, соціальні зв'язки й емоційний стан. Цікаве спостереження у цьому зв'язку висловили J. DiMiccio й D. Millen: у діловому спілкуванні електронною поштою люди швидше за все сприйматимуть негативні емоції з більшою інтенсивністю, ніж позитивні. А в соціальних мережах люди представляють різні онлайн-ідентичності, які впливають на створюване ними враження [4].

Виклад основного матеріалу. Джерельною базою цього дослідження послуговував англійськомовний новинний ресурс BBC News (<https://www.bbc.com/news>), а саме розділ “War in Ukraine”. База даних новинних заголовків була створена за допомогою розширення HeadingsMap у браузері Google Chrome, що дозволило авто-

матизувати цей процес. Загалом було вилучено з сайту 1017 заголовків, які охоплювали період війни з березня по жовтень 2022 року. Подальша робота відбувалася із залученням IBM Watson: IBM Watson™ Tone Analyzer.

Алгоритм роботи з аналізатором такий: ви надсилаєте JSON, простий текст або HTML, що містить письмові дані для аналізу, до програми. Сервіс приймає до 128 КБ тексту, що становить близько 1000 речень. Служба повертає результати JSON, які повідомляють про тон вашого введення (рис. 1).

Отже, у результаті аналізу перших 29 заголовків ми отримали інформацію про те, що в них наявні тони страху, радості та смутку. Наведення курсору на ці показники дозволяє відкрити пояснення, що кожен із них означає і в якій мірі присутній у тексті. Файл JSON містить обрахований показник кожного з тонів у проаналізованому тексті (рис. 2).

На рівні речення можна переглянути розподіл отриманих результатів відповідно до тону: інтенсивністю кольору позначено рівень тону для кожного речення відповідно до наведеної шкали (рис. 3).

Заголовки-речення, для яких програма визначила емоційні тони, було погруповано відповідно до тональності. Наводимо результати, у дужках вказуємо індекс тональності, який був присвоєний програмою кожному реченню згідно зі шкалою.

1) Аналітичний тон, що відображає міркування та аналітичне ставлення людини до речей, виявлено в таких заголовках:

Russia says it will retake annexed areas despite retreat (0.72);

What does Russia's troop call-up mean for Ukraine? (0.76);

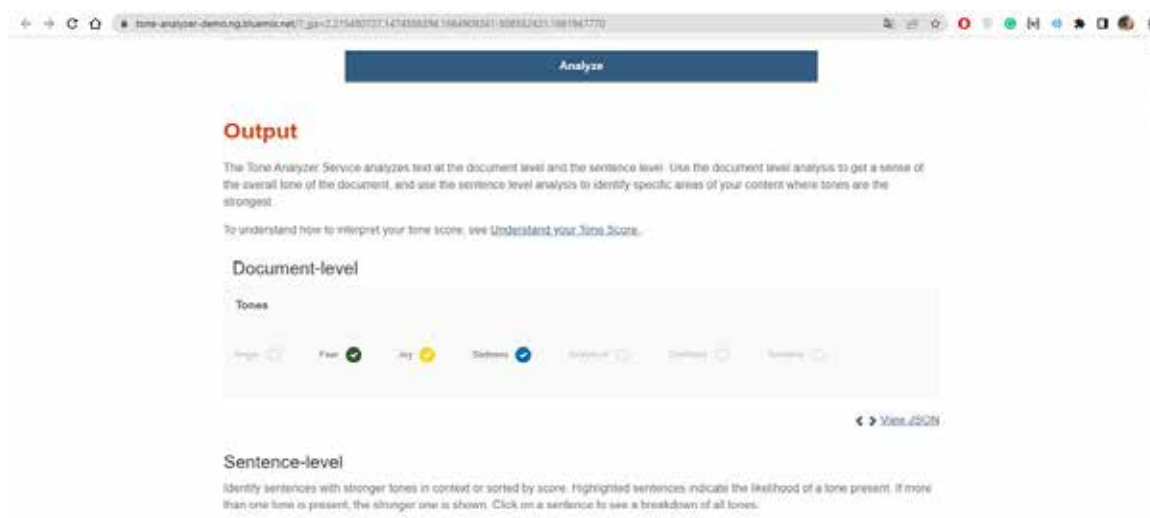


Рис. 1. Результати аналізу на рівні документа

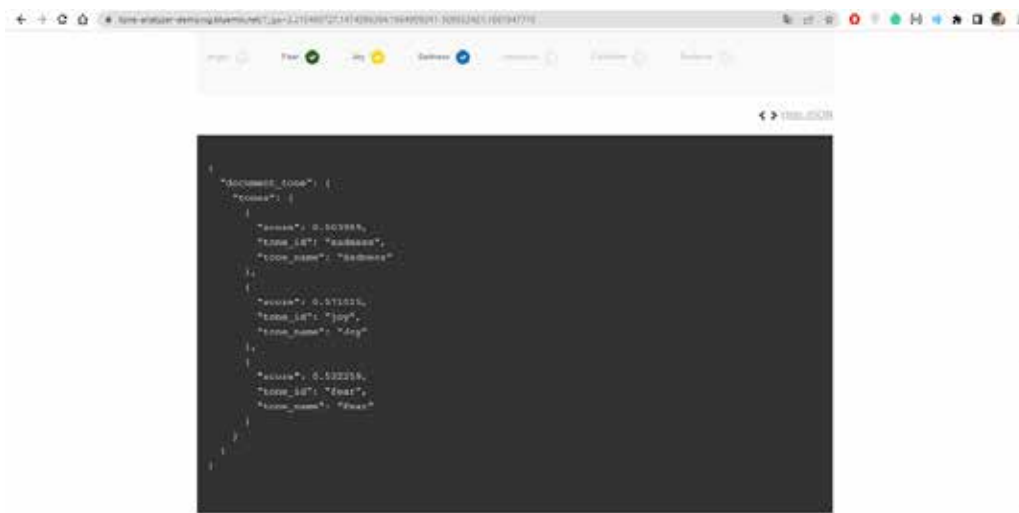


Рис. 2. Файл JSON



Рис. 3. Результати аналізу на рівні речення

Why EU leaders struggle to agree on gas price cap (0.76);

Lyman retreat sees rare criticism of Russian leaders (0.76).

У наведених прикладах заголовки сигналізують про те, що в статті будуть представлені певні міркування чи оцінки експертів.

2) Тон, що показує ступінь впевненості людини, наприклад, у такому заголовку: The enemy is near, but we are standing firm (0.87). Тут чітко видно, що та людина, якій належить ця цитата, не має жодних сумнівів щодо віри в Збройні Сили України.

3) Наступні речення наповнені тональністю страху як реакцією на небезпеку, що насувається, це механізм виживання, який є реакцією на якийсь негативний подразник, може бути легка обережність або надзвичайна фобія:

Ukraine recaptures southern villages from Russians (0.59);

What Russian annexation means for Ukraine's regions (0.57);

Ukrainian refugees fear they will become homeless (0.91).

Несподіванкою стало визначення тону страху в першому реченні: можливо, через слово "recaptures", що має негативну конотацію.

4) Тон радості чи щастя має відтінки насолоди, задоволення, він міститься в реченнях, які викликають відчуття благополуччя, внутрішнього спокою, любові, безпеки та задоволення. Серед інших цим тоном програма позначила заголовки, які не можна співвіднести з тоном радості, як-от:

A rare glimpse of daily life in occupied Ukraine (0.65);

Crimean beauty queen fined over Ukrainian song (0.88);

Zelensky and Musk in row over 'peace plan poll' (0.64).

На прикладі цих речень помітно, що машинний аналіз все-таки не може повністю врахувати весь контекст і зчитати підтекст. На формальне визначення тону вплинули такі слова: “glimpse”, “beauty”, “peace”, що поза контекстом є позитивно конотованими.

5) Тон смутку вказує на відчуття втрати та неблагополуччя, він також є маркером пасивної, неенергійної, відчуженої від соціуму й замкненої в собі людини. Аналізатор позначив цим тоном такі заголовки:

Ukraine war in maps: Russia losing more territory (0.59);

Ukrainian medical trainees continue studies in UK (0.58);

UK's winter gas supplies are at 'significant risk' (0.58).

Як бачимо, у першому реченні склалася подібна до вже описаних ситуація: слова “war” та “losing” поза контекстом будуть негативно конотовані, а тому цей заголовок асоціюється зі смутком, але тут не врахований ширший історичний контекст і ролі країн-учасниць війни.

6) Тон невпевненості — це ступінь стримування людини. Програма визначила його, наприклад, у такому заголовку: Ukrainians try to cross

their new Iron Curtain (0.62). Очевидно, що ключовим словом для визначення тональності стало “try”, яке означає не кінцевий результат, а лише описує якісь намагання.

Висновки. Отже, здійснений комп’ютерний аналіз дозволив визначити прогалини в алгоритмі роботи аналізаторів такого типу, як програма IBM Watson™ Tone Analyzer, що залежить від того, в прямому чи переносному значеннях ужито слово, від ясності й повноти контексту. Заголовки є мінімальним текстом, а тому програма не може ідентифікувати їх у ширшому контексті, не може врахувати соціокультурний, історичний чи політичний фон. У перспективі — здійснити аналіз тональності заголовків, але в контексті новинних текстів. Загалом, як показав комп’ютерний аналіз, серед всього масиву англійськомовних заголовків BBC news про війну в Україні переважають марковані тонами смутку й страху.

Наразі, на жаль, немає текстового аналізатора тональності для української мови, тому однозначно перспективним є створення сервісу, подібного до того, що пропонує компанія IBM Watson. Такий програмний продукт був би цікавим як для психолінгвістів, так і журналістів-практиків.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

- Allport F., Lepkin M. Building War Morale with News-Headlines. *Public Opinion Quarterly*. 1943. Vol. 7, № 2. P. 211–242.
- Blood D., Phillip P. Recession headline news, consumer sentiment, the state of the economy and presidential popularity: A time series analysis 1989–1993. *International Journal of Public Opinion Research*. 1995. Vol. 7., № 2. P. 2–22.
- Cambria E., Schuller B., Xia Y., Havasi C. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*. 2013. Vol. 2., № 2. P. 15–21.
- DiMicco J., Millen D. Identity Management: Multiple Presentations of Self in Facebook. *Proceedings of the 2007 International ACM Conference on Supporting Group Work*, 2007. P. 383–386.
- IBM Watson Tone Analyzer. URL: <https://cloud.ibm.com/docs/tone-analyzer> (дата звернення: 28.10.2022)
- Jamieson K. H. *Dirty politics: Deception, distraction, and democracy*. New York: Oxford University Press, 1992.
- Lighthart A., Catal C., Tekinerdogan B. Systematic reviews in sentiment analysis: a tertiary study. *Artificial Intelligence Review*. 2021. Vol. 54, № 7. P. 4997–5053.
- Mudinas A., Zhang D., Levene M. Combining lexicon and learning based approaches for concept-level sentiment analysis. *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining*. 2012. P. 1–8.
- Mutz D. C. *In-Your-Face Politics: The Consequences of Uncivil Media*. Princeton. NJ: Princeton University Press, 2015.
- Taiwo R. Language, Ideology and Power Relations in Nigerian Newspaper Headlines. *Nebula*. 2007. Vol. 3, № 1. P. 218–245.
- Tepavcevic M. Linguistic Specificities of Media Headline Discourse - in the Case of Montenegrin Newspapers. *Media Dialogues / Medijski dijalozi*. 2019. Vol. 14., № 2. P. 75–87.
- Turney P. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Proceedings of the Association for Computational Linguistics*. 2002. P. 417–424.
- Volcani Y., Fogel D. System and method for determining and controlling the impact of text: U.S. Patent Application № 10/376.680.
- Wang H., Lu Y., Zhai C. Latent aspect rating analysis on review text data: a rating regression approach. *Proceedings of the 16th ACM SIGKDD International conference on Knowledge discovery and data mining*. 2010. P. 783–792.