

СТРУКТУРНА, ПРИКЛАДНА ТА МАТЕМАТИЧНА ЛІНГВІСТИКА

УДК 811.111+81'33:81'42

DOI <https://doi.org/10.32782/2710-4656/2022.5.2/09>

Назарчук Р. З.

Національний університет “Львівська політехніка”

Роса І. В.

Національний університет “Львівська політехніка”

ТОНАЛЬНІСТЬ ТВІТІВ ПРО УКРАЇНУ У ЧАСІ РОСІЙСЬКО-УКРАЇНСЬКОЇ ВІЙНИ

Запропоновану статтю присвячено аналізу тональностей дописів про Україну у соціальній мережі «Твіттер» під час російсько-української війни. Зокрема, увагу зосереджено на вивченні твітів лідерів та представників урядів провідних країн світу, а саме США, Великої Британії, Німеччини, Франції та Польщі. Засвідчено особливості твітів про Україну до початку війни, а також у першій та останній на момент дослідження місяці війни. Закцентовано увагу на зростанні ваги соціальних мереж як результату розвитку інформаційно-комунікативних технологій, зокрема схарактеризовано твіттер як платформу, де висвітлюють свої позиції відомі політичні й громадські діячі і яка сприяє трансформації аудиторії з пасивного сприймача інформації до її творця. Узагальнено комплекс теоретичних питань, пов'язаних із тлумаченням поняття тональності тексту на сучасному етапі розвитку лінгвістики. Сентимент-аналіз визначено як клас методів контент-аналізу в комп'ютерній лінгвістиці, призначений для автоматизованого виявлення емоційно забарвленої лексики в текстах і/або емоційної оцінки того, про що йдеться у повідомленні, самим автором тексту. Показано, що основне завдання аналізу тональності полягає у встановленні категорії (позитивної, негативної чи нейтральної) висловленої думки. За допомогою програмного забезпечення LIWC-22 проаналізовано 105 твітів про Україну, 17 з яких було опубліковано безпосередньо до російсько-української війни, 53 – у перший місяць і 35 – в останній на момент дослідження місяці війни (серпень-вересень 2022 року). Виявлено, що позитивна тональність переважає в дописах про необхідність забезпечення захисту для України, підтримки з боку інших країн, надання матеріальної допомоги тощо. У твітах, які стосуються виявів агресії з боку росії, руйнувань міст та сіл України, жертв серед мирного населення та ін., очікувано переважає негативна тональність. У статті подано зразки твітів офіційних представників кожної з п'яти згаданих країн і показано їхню тональність у різні періоди війни, а також простежено динаміку тональностей (наведено кількісні показники змін).

Ключові слова: аналіз тональності, твіттер, позитивна тональність, негативна тональність, нейтральна тональність, війна, Україна.

Постановка проблеми. У сучасній комп'ютерній лінгвістиці, а саме у галузі опрацювання природної мови, аналіз тональності тексту належить до поширених напрямків дослідження. Стрімка еволюція інформаційно-комунікативних технологій призвела до активного використання соціальних мереж, різноманітних форумів та інших платформ, що мають значний вплив на формування суспільних думок, оцінок, настроїв, а також спричинила появу в інтернеті величезної кількості даних. Збір й тлумачення

людським ресурсом згенерованих користувачами текстів – затратне в часі та зусиллях завдання. Отож з початком ХХІ століття починає стрімко розвиватися сентимент-аналіз, поступово стаючи однією з найактуальніших дослідницьких сфер і отримуючи широке застосування не лише в комп'ютерній лінгвістиці, але й соціології, політології, менеджменті, економіці, медицині, сфері обслуговування тощо.

Аналіз останніх досліджень і публікацій. За короткий період свого існування сентимент-аналіз

пройшов інтенсивний шлях розвитку й на сьогоднішній день може похвалитися не лише потужними практичними, але й теоретичними здобутками. Вивченню машинних методів аналізу тональності та класифікації текстової інформації, опису програмних застосунків, призначених для здійснення моделювання, присвячено ряд наукових праць українських дослідників, серед яких, наприклад: С. Альбота [6], Н. Дарчук [1], О. Немеш [2], І. Онищенко [3], А. Романюк [2], Н. Шаховська [5] та ін. Закордонні колеги, напрацьовуючи різноманітні підходи до сентимент-аналізу, досліджують психологічні та соціальні аспекти мережевого мовлення (Б. Ліу [11], Ф. Поцці й Е. Ферсіні [14], К. Іглесіас і А. Морено [7] та ін.), роль семантики мовлення та шляхи її тлумачення й моделювання автоматизованими системами (наприклад, Р. Стайн [15]), впроваджують здобутки машинного навчання у сентимент-аналіз (Л. Лей і Д. Ліу [9], Б. Ліу [10] та ін.), описують особливості вияву стилістичних прийомів (гумору, іронії, сарказму) й «спау думок» (наприклад, фальшивих відгуків для несправедливої реклами або антиреклами певних продуктів) – див. про це детальніше [14, с. 142–155] тощо.

Постановка завдання. Стрімкий розвиток сентимент-аналізу подиктований наявністю майже невичерпних джерел для вивчення – новинних ресурсів, блогів, текстів соціальних мереж, торговельних платформ з можливістю коментування дописів чи продуктів. Багато компаній (Google, Twitter, IBM, SAP, Microsoft та ін.) стимулюють розвиток технологій текстових досліджень, надаючи різноманітні рішення для отримання даних та їхнього опрацювання, що призводить до появи все точніших та ефективніших методів [9, с. 7]. Результати сентимент-аналізу застосовують для прогнозування політичних рейтингів, настроїв, намірів, уподобань людей, емоційних відгуків стосовно товарів (див. про це, наприклад, [7; 9]). Як стверджують дослідники, тональність текстів активно досліджують у таких сферах: бізнес і фінанси, політика, медицина й охорона здоров'я, розваги (особливо кіноіндустрія), наука (академічне письмо) [9, с. 6]. Технології автоматичного комп'ютерного оцінювання виконують багато завдань, даючи змогу швидко опрацювати великі обсяги текстових даних, проте їм кинуть і серйозні виклики. Сентимент-аналіз має справу з особливим виявом людської свідомості й підсвідомості – емоціями, які можуть бути виражені по-різному, і тому на даний момент немає однозначного підходу, скажімо, до тлумачення слен-

гізмів, діалектизмів, метафори й метонімії, стилістичних прийомів (іронії, сарказму тощо). Отож наша розвідка – ще один крок на шляху до розуміння ідентифікації тональностей, в чому і полягає її актуальність.

Мета статті – з використанням текстового аналізатора LIWC-22 [8] з'ясувати тональність твітів світових лідерів США, Великої Британії, Німеччини, Франції та Польщі про Україну, створених безпосередньо до російсько-української війни, у першій і в останні (серпень-вересень 2022 року) на момент дослідження місяці війни.

Виклад основного матеріалу. Терміном «сентимент-аналіз» позначають процес використання алгоритмів і комп'ютерних технологій для планомірного вияву, вилучення і класифікації суб'єктивної інформації й емоційних станів, виражених у тексті. Як зауважує Б. Ліу, «аналіз тональності тексту – це клас методів тлумачення змісту в комп'ютерній лінгвістиці, призначений для автоматизованого виявлення в текстах емоційно забарвленої лексики й емоційної оцінки автора певних об'єктів, мова про які йде в тексті [10, с. 1]. Поняття аналізу тональності має різні синоніми та близькі за змістом терміни: сентиментометрія (sentiment metrics), бренд-моніторинг (brand monitoring), соціомедіааналіз (social media analysis), розвідка думок (opinion mining), «підслуховування» думок (opinion listening), сентимент-аналіз (sentiment analysis) [10, с. 1]. Офіційним постановом методу вважають 2003 рік, коли його назву вперше вжив Т. Насукава, проте початками сентимент-аналізу необхідно трактувати перші дослідження думок під час Другої світової війни та розвідки про суб'єктивність у текстах на початку 1990-х [12, с. 16].

Тональність усього тексту дослідники визначають як сукупність лексичних тональностей його складових одиниць із урахуванням правил їхнього поєднання [4; 9; 10; 13]. Завдяки тональності окремих слів можна визначити тональність повідомлення загалом. Емоційне забарвлення у сентимент-аналізі найчастіше позначають триадою позитивного, негативного й нейтрального. Наприклад, слова «чудовий, щастя, доброта можуть бути позначені як позитивні, а такі слова, як поганий, потворний, сум – негативні» [10, с. 5]. Нейтральна тональність зазвичай означає, що жодної суб'єктивної думки не було висловлено. Аналіз тональності тісно пов'язаний з класифікацією текстів на суб'єктивність, що категоризує висловлення з точки зору об'єктивності та суб'єктивності. Однак варто зауважити, що не

лише суб'єктивні висловлення містять сентимент. Об'єктивні фрази також можуть містити приховану думку (див. про це [16]).

Аналізуючи тональність, слід звертати увагу на чотири моменти: 1) суб'єкт тональності (автора повідомлення); 2) об'єкт тональності (про що йде мова); 3) аспект тональності (характеристику об'єкта); 4) оцінку тональності – сентимент [3]. Як уважає Н. Дарчук, суб'єкт тональності може надавати характеристику об'єктам висловлення цілісно в той чи інший спосіб. Об'єкт тональності – це зміст написаного, який можна визначати, наприклад, за ключовими словами. Аспект тональності – це характеристика, яку надають об'єктові тональності. А сентимент – це власне тональна оцінка – ставлення автора до описуваного предмета, аспекту, властивостей [1, с. 10].

Основне завдання цього дослідження – визначити емоційне забарвлення твітів лідерів провідних країн світу під час російсько-української війни. Довжина повідомлень в соціальній мережі «Твітер» (вони є досить короткими, не більше 280 символів, й змістовно насиченими), як і їхній неформальний стиль, роблять згадані зразки тексту досконалим і плідним матеріалом для вивчення тональності, а також дають змогу отримати точні результати сентимент-аналізу. Варто згадати й інші особливості твіттера, важливі для дослідження: відправник й отримувач повідомлень «працюють» майже одночасно; кількість підписників у певних користувачів, зокрема у лідерів провідних країн світу, сягають мільйонів (інколи державні служби використовують твітер для оповіщень про надзвичайні ситуації); у повідомленнях присутня суб'єктивність (сугестивний компонент), і вони часто є реакцією на якусь подію та містять аргументацію (аргументативний компонент), адже зазвичай мета допису – переконати у правильності позиції відправника інформації (мовця).

Тональність у дослідженні було визначено за допомогою лексикона LIWC-22 (Linguistic Inquiry and Word Count) [8], що є одним із найточніших серед програмних забезпечень, призначених для аналізу тексту (вивчення різних зразків емоційних, когнітивних, структурних та технологічних текстових компонентів [13]). LIWC добре зарекомендував себе і пройшов внутрішню та зовнішню перевірку у процесі, що триває понад десять років із залученням фахівців у галузі психології, соціології та лінгвістики [13]. Програма аналізує формальні та неформальні тексти, дописи з соціальних медіа, книги та короткі істо-

рії. LIWC використовує власний словник з майже 4500 одиниць, організованих в одну (або більше) зі 76 категорій [13]. До прикладу, за допомогою LIWC виявляли політичні настрої на базі твітів, передбачали початок депресії у людей через тексти соціальних медіа, розрізняли щасливі пари від нещасливих на основі їхніх миттєвих повідомлень [13]. На жаль, LIWC не містить тлумачення таких лексичних елементів, як акроніми, сленгізми, емотикони, що є доволі важливими для сентимент-аналізу. Необхідно також взяти до уваги, що середнє значення позитивної тональності текстів, відібраних із соціальних мереж становить 5.93, а негативної – 2.34.

Світові лідери таких країн, як США, Велика Британія, Німеччина, Франція та Польща почали активно обговорювати ситуацію в Україні приблизно за три тижні до початку війни. У твітах зазвичай йшлося про підтримку України у випадку, якщо війна розпочнеться, можливість запобігти конфліктові дипломатичним шляхом, подальші шляхи захисту та надання допомоги. Для прикладу аналізу тональності було обрано твіти урядів згаданих країн, опубліковані упродовж першого місяця війни (рис. 1).

З офіційних сторінок уряду Великої Британії було проаналізовано 12 твітів. Якщо допис стосувався забезпечення захисту та підтримки суверенітету, впевненості у перемозі України над ворогом, позитивної тональності було більше, ніж негативної. Наприклад, у твіті про санкції проти агресора позитивна тональність становить 5.71, а негативна – 2.89: «I thank President @vonderleyen for her close cooperation on sanctions so far. We will continue working with our allies to prepare a package of economic measures that would punish Russia if Ukraine's sovereignty is breached» (@BorisJohnson, 15.02). У інших твітах, де йшлося про небезпеку для України з боку росії, негативна тональність завжди була понад середніх показників і сягала приблизно 5, а позитивна становила зазвичай 0. За декілька тижнів до війни активно почали писати про Україну представники влади Німеччини. Тональність була переважно позитивною і відповідала середньому значенню, адже у дописах йшлося про необхідність поважати суверенітет кожної держави, а також про намагання утриматися від війни взагалі або вирішити її дипломатичним шляхом.

У перший місяць російсько-української війни твіти почали ще активніше з'являтися у соціальній мережі. З дописів представників влади США можна побачити, що позитивна тональність пере-



Рис. 1. Тональність твітів у перший місяць війни

важає і сягає високих показників, адже повідомлення присвячені сміливості українців, санкціям проти агресора, наданню фінансової допомоги українському народу. Наприклад, такий допис: «President Putin thought he could roll into Ukraine – and the world would roll over. Instead, he met a wall of strength he never imagined. He met the Ukrainian people» (@POTUS, 2.03) не містить негативної тональності, натомість позитивна становить 10.

У дописах президента Франції негативна тональність домінує, адже велику увагу приділено тому, що війна несе небезпеку для всього людства. Отож серед 11 проаналізованих твітів лише у трьох переважає позитивна тональність, показник якої понад середнє, а також виявлено 6 дописів із негативною і 3 – з нейтральною тональністю. Для прикладу подаємо повідомлення, позитивна тональність якого становить 0, а негативна – 10: «I spoke to President Putin this morning. He refuses to stop his attacks on Ukraine at this point. It is vital to maintain dialogue to avoid human tragedy. I will continue my efforts and contacts. We must avoid the worst» (@EmmanuelMacron, 3.03).

Протягом першого місяця війни президент Польщі створив 14 дописів, 5 з яких мають позитивну тональність із понад середніми показниками; виявлено ще 6 повідомлень, забарвлених негативно, інші були нейтральними. У твітах, де показник позитивної тональності був високий, зазначалося про незламність українського народу, впевненість у перемозі України та про підтримку з боку Польщі. Подаємо приклад твіту, який має однаково позитивну й негативну тональність: «Afternoon, I visited the POL-UKR bordercrossing in

Korczowa and the reception point for refugees. Agata was in the accommodation places of women and children from Ukraine. I sad to @ZelenskyUa: «Tell your boys that they can fight calm, their relatives are safe in Poland» (@AndrzejDuda, 5.03).

На офіційних сторінках представників Великої Британії у твіттері від кінця лютого і протягом березня було опубліковано багато дописів із позитивною тональністю, як-от: «Tonight I met personnel at @RAFBrizeNorton to thank them for their efforts in coordinating the UK response to the invasion of Ukraine. The UK has sent defensive military aid to Ukraine and offered a full package of support to @NATO to strengthen its eastern flank» (@BorisJohnson, 27.02). Набагато менше твітів було з негативною тональністю: «In another economic blow to the Putin regime following the illegal invasion of Ukraine, the UK will move away from dependence on Russian oil throughout this year, building on our severe package of international economic sanctions» (@BorisJohnson, 8.03). У першому прикладі показник позитивної тональності становить 11.11, негативної – 0, а у другому прикладі навпаки: позитивна тональність 0, а негативна – 5.56.

За перший місяць російсько-української війни твітер німецької влади засвідчив приблизно однакову кількість дописів як з позитивною, так і негативною тональністю. «The people of #Ukraine are fighting for freedom and democracy – values that we share with them. We are on their side. We will therefore be supplying them with defensive weapons. And we have therefore approved tough sanctions such as cutting Russian banks off from SWIFT» (@Bundeskanzler, 27.02) – чудовий

приклад твіту, в якому позитивна й негативна тональність має показник 2.22.

Пропри те, що війна триває досі, кількість тематичних твітів значно знизилася, інформації про розвиток подій в Україні стає все менше. Аналіз твітера президента США за серпень-вересень 2022 року виявив, що серед дев'яти проаналізованих дописів п'ять містять позитивну тональність, два – негативну і два – нейтральну, а показники негативної й позитивної тональності майже однакові (зауважмо, що згадані показники понад середнє). Наприклад, у привітанні з Днем Незалежності України позитивна тональність сягає 12.20, а негативна, як можна очікувати, дорівнює нулю.

Протягом серпня-вересня 2022 року на офіційній сторінці президента Франції Еммануеля Макрона тільки один твіт стосується України. У ньому переважає негативна тональність із показником 2.70: «Russian armed forces must know that the war crimes committed in Ukraine will not go unpunished. A team of French experts have already provided assistance on the ground and we have donated a mobile DNA testing laboratory» (@EmmanuelMacron, 1.08).

На сторінках офіційних представників Великої Британії й надалі присутня велика кількість дописів про Україну, в яких переважає позитивна тональність і показники якої є понад середнє. Проте збільшилася кількість твітів, де позитивна і негативна тональність однакові або дорівнюють нулю. Також виявлено декілька твітів із високим показником негативної тональності (скажімо, 11.90, наприклад: «Vladimir Putin is once again violating international law with his threats to annex more of Ukraine. We will not hesitate to take further action, including imposing more sanctions to cripple Putin's war machine. We will ensure he loses this illegal war» (@trussliz, 30.08).

Висновки та перспективи. Отже, щоб виявити, вилучити та вивчити емоційні стани та

суб'єктивну інформацію в різних за тематикою й жанрами текстах, сентимент-аналіз як дитя комп'ютерної лінгвістики використовує кількісні методи, а також алгоритми машинного навчання. Програмний продукт LIWC отримав широке застосування для оцінки тональностей у текстах соціальних мереж. Дослідження твітів про Україну, створених представниками влади п'яти провідних країн світу (США, Великої Британії, Німеччини, Франції та Польщі) під час російсько-української війни, виявило домінування позитивної і негативної тональностей. До початку війни середня позитивна тональність в офіційних представників США була 2.31, негативна – 2.89, Великої Британії – 2.49 і 2.11 відповідно, у лідерів Німеччини переважав нейтральний тон, твітів урядів Польщі та Франції про Україну не виявлено. Упродовж першого місяця російсько-української війни позитивна тональність дописів представників Великої Британії та США збільшилася і становила 4.9 і 3.11 відповідно, а негативна зменшилася (1.87 і 1.17). В твітах німецької влади негативна тональність становила 1.97 на відміну від Польщі і Франції, в яких засвідчено показники 4.71 і 4.49 відповідно. Дослідження повідомлень останніх місяців війни дало змогу зафіксувати такі показники: позитивна тональність – 4.73 (Велика Британія), 4.69 (США), 3.9 (Франція), 3.3 (Німеччина), 2.13 (Польща); негативна тональність – 6.38 (Польща), 3.63 (Німеччина), 2.47 (США), 2.46 (Велика Британія), 1.7 (Франція). Точність результатів зумовлена стилем дописів офіційних представників урядів, який передбачає мінімальне використання мовних засобів, що можуть ввести в оману аналізатор тональності.

Сентимент-аналіз, поєднуючи схоластичні теорії емоційних станів з практичними завданнями опрацювання природної мови, наближає мислення комп'ютера до людського, є важливим внеском до розвитку штучного інтелекту, а тому забезпечує значні дослідницькі перспективи.

Список літератури:

1. Дарчук Н. Лінгвістичні засади автоматичного сентимент-аналізу українськомовного тексту. URL: <https://seanewdim.com/wp-content/uploads/2021/04/Linguistic-approach-for-development-of-computer-based-sentiment-analysis-in-the-Ukrainian-language-N.-Darchuk.pdf>
2. 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.
3. 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.
4. Романишин М., Романюк А. Тональний словник української мови на основі сентимент-анотованого корпусу // Українське мовознавство. 2013. Вип. 43. С. 63–74.
5. Шаховська Н., Шаховська Х. Метод аналізу відгуків клієнтів з природномовних текстів // Штучний інтелект. 2018. № 3 (81). С. 18–26.

6. Albota S. Linguistically manipulative, disputable, semantic nature of the community reddit feed post. URL: <http://ceur-ws.org/Vol-2870/paper57.pdf>
7. Iglesias C., Moreno A. Sentiment analysis for social media // Applied Sciences. 2019. № 9 (23):5037. URL: <https://doi.org/10.3390/app9235037>
8. Linguistic Inquiry and Word Count. URL: <https://www.liwc.app/>
9. Lei L., Liu D. Conducting Sentiment Analysis. Cambridge: Cambridge University Press, 2021. 106 p.
10. Liu B. Sentiment Analysis and Opinion Mining. Williston-London: Morgan&Claypool Publishers, 2012. 168 p.
11. Liu B. Sentiment Analysis: Mining Opinions, Sentiments and Emotions and Subjectivity. Cambridge: Cambridge University Press, 2015. 384 p.
12. Mäntylä M., Graziotin D., & Kuutila M. The evolution of sentiment analysis: A review of research topics, venues, and top cited papers // Computer Science Review. 2018. № 27. P. 16–32.
13. Pennebaker J. W., Francis M. E., Booth R. J. Linguistic Inquiry and Word Count (LIWC). Operator's Manual. URL: http://www.gruberpeplab.com/teaching/psych231_fall2013/documents/231_Pennebaker2007.pdf
14. Sentiment Analysis in Social Networks/ ed.by F. A. Pozzi, E. Fersini, E. Messina, B. Liu. Cambridge, MA: Elsevier, 2017. 284 p.
15. Stine R. Sentiment analysis // Annual Review of Statistics and its Application. 2019. № 6. P. 17.1–17.2.
16. Wiebe J., Bruce R., O'Hara Th. Development and use of a gold-standard data set for subjectivity classifications. URL: <https://aclanthology.org/P99-1032.pdf>

Nazarchuk R. Z., Rosa I. V. TONALITY OF TWEETS ABOUT UKRAINE DURING THE RUSSIAN-UKRAINIAN WAR

The proposed paper reveals the tonality of posts about Ukraine on the Twitter social network during the Russian-Ukrainian war. In particular, special attention is paid to studying the tweets of the leaders and government representatives of the world's leading countries, namely the USA, Great Britain, Germany, France and Poland. The peculiarities of tweets about Ukraine before the start of the war and in the first and last months of the war at the time of the research are documented. Attention is focused on the growing importance of social networks due to the development of information and communication technologies. In particular, Twitter is characterised as a platform where well-known political and public figures highlight their positions and contribute to transforming the audience from a passive receiver of information to its creator. A set of theoretical questions related to the interpretation of the concept of text tonality at the current stage of the development of linguistics is summarised. Sentiment analysis is treated as a class of content analysis methods in computer linguistics intended for automated detection of emotionally coloured vocabulary in texts and/ or emotional assessment of what is being said in the message by the author of the text. It has been shown that the main task of tonality analysis is to establish the category (positive, negative or neutral) of the expressed opinion. With the help of LIWC-22 software, 105 tweets about Ukraine have been analysed, 17 of which were published immediately before the Russian-Ukrainian war, 53 in the first month, and 35 in the last months of the war at the time of the study (August-September 2022). It has been found that a positive tonality prevails in posts about the need to protect Ukraine, support from other countries, provision of material assistance, etc. Predictably, a negative tonality prevails in the tweets that refer to manifestations of aggression by Russia, the destruction of cities and villages in Ukraine, victims among the civilian population, etc. The paper also presents samples of tweets by official representatives of each of the five mentioned countries. It shows their tonality in different war periods and traces tonality dynamics (quantitative indicators of changes are provided).

Key words: sentiment analysis, Twitter, positive tonality, negative tonality, neutral tonality, war, Ukraine.