

ПРИКЛАДНА ЛІНГВІСТИКА

УДК 811.161.2'32'374

DOI: [https://doi.org/10.17721/um/54\(2024\).204-221](https://doi.org/10.17721/um/54(2024).204-221)

Наталія ДАРЧУК, д-р філол. наук, проф.

ORCID ID: 0000-0001-8932-9301

e-mail: n.darchuk@knu.ua

Київський національний університет імені Тараса Шевченка,

Київ, Україна

Оксана ЗУБАНЬ, канд. філол. наук, доц.

ORCID ID: 0000-0002-2644-3892

e-mail: oxana.zuban@knu.ua

Київський національний університет імені Тараса Шевченка,

Київ, Україна

Валентина РОБЕЙКО, асист.

ORCID ID: 0000-0003-2266-7650

e-mail: valentya.robeiko@knu.ua

Київський національний університет імені Тараса Шевченка,

Київ, Україна

Юлія ЦИГВІНЦЕВА, д-р філософії, мол. наук. співроб.

ORCID ID: 0000-0002-9684-3840

e-mail: tsyhvintseva@nas.gov.ua

Інститут української мови НАН України, Київ, Україна

ІНДЕКСУВАННЯ НЕГАТИВНОГО СЕНТИМЕНТУ УКРАЇНСЬКОМОВНОГО ТЕКСТУ СИСТЕМОЮ “TEXTATTRIBUTOR 1.0”

Вступ. У XXI ст. інформаційне поле є повноцінним полем бою. В українському інформаційному просторі все більшої ваги набуває проблема токсичності тексту, мови ворожнечі (ненависті). Саме тому інтерес дослідників до маркерів негативної тональності тексту, особливо медійного, невпинно зростає. Описано структуру та результати роботи окремого модуля автоматичної системи статистичної параметризації українськомовних текстів “TextAttributor 1.0” – лінгвістичну експертизу токсичності тексту.

Завдання розв'язано двома методами: методом словників і правил (обчислення статистичних параметрів) та методом машинного навчання.

Результати дослідження ґрунтуються на матеріалі корпусу мережесих медіатекстів політичного дискурсу обсягом 10 млн слововживань. Для реалізації мети створено лексикографічну базу даних, до якої ввійшли три словники: емоціогенів, мови ворожнечі й токсичних сполук, а також сформовано навчальну та контрольну вибірки текстів, на яких проводили оцінювання параметрів обраної моделі методом машинного навчання. У межах проєкту обрано ефективну для обчислень архітектуру на основі методу fastText та його інструментальних засобів. Індекс токсичності обчислено за допомогою вербальної ідентифікації негативного сентименту тексту на базі створеного лінгвістичного та програмного забезпечення й унаслідок через згенеровану системою лінгвістичну експертизу тексту, що відображає статистичну картку семантичних класів негативної лексики за класифікаційними маркерами лексикографічних списків та висновок нейронної мережі.

Висновки. Система “TextAttributor 1.0” перебуває на етапі тестування і вдосконалення функцій, тому у статті описано проміжну β -версію системи, проте отримані результати з визначення токсичності свідчать, що розроблена методика квантування вербальних засобів за семантичними параметрами (негативна емотивність) методами словників і правил та машинного навчання є ефективною для досягнення поставлених завдань і дає можливість не лише визначити межу між токсичним і нейтральним текстом, а й на підставі властивих тексту лексичних категорій наблизитися до розв'язання проблеми автоматичного виявлення наративу тексту. Методику розроблення модуля визначення токсичності медійного тексту в системі “TextAttributor 1.0” описано й оприлюднено на сторінці вебзастосунку у квітні 2024 р., але у формі науково-дослідницької статті цю інформацію друкують вперше.

Ключові слова: українськомовний медіатекст, токсичний текст, мова ворожнечі, тональність, негативний сентимент, індекс токсичності тексту.

Вступ

Тональність тексту – це позитивне, негативне або нейтральне ставлення його автора до об'єктів, явищ, ситуацій, осіб, згаданих у тексті. Тональність є важливою ознакою тексту, вона

впливає на його сприйняття та розуміння читачем, дає змогу авторові реалізувати свої комунікативні цілі.

Аналіз тональності тексту нині цікавить не лише мовознавців, а й політологів, управлінців, маркетологів, рекламників, іміджмейкерів та інші категорії фахівців, що працюють із певним брендом (у широкому розумінні цього слова). Особливо важливим видається це завдання для медійних текстів, під якими розуміємо тексти з різних джерел, що “передають, зберігають та відтворюють інформацію, яка впливає на громадську думку” [Словник з медіаграмотності, 2022].

Увага до інформації та її емоційного складника завжди була підвищеною, особливо в епоху гібридних воєн. Як слушно зауважує Г. Почепцов, “поява нових медіа полегшила проведення міжнародних інформаційних інтервенцій, бо вони створюють інформаційно-комунікативний простір, у якому пришвидшено всі процеси обміну. Державний апарат не в змозі реагувати і прогнозувати можливі наслідки таких інтервенцій. І майже на кожному кроці він програє” [Почепцов, 2016, с. 44]. Сучасна повномасштабна російсько-українська війна надала цій проблемі ще більшої значущості. В умовах такого інформаційного простору важливо мати інструменти, які дадуть змогу автоматично аналізувати й класифікувати тексти за негативним або позитивним сентиментом чи за різними емоціями.

Інтерес дослідників до визначення тональності тексту невпинно зростає. За останні кілька років з’явилося багато публікацій закордонних та українських науковців, присвячених різним підходам і наявним методам виявлення тональності [Залуцька та ін., 2023; Романюк, & Романишин, 2013; Стахів, & Скопівський, 2024; Шаховська, & Гірак, 2017; Шингалов та ін., 2017; Ялова, 2021; Reveilhac, & Morselli, 2022; Nandwani, & Verma, 2021; Cui et al., 2023; Cambria et al., 2024], розробленню систем та інструментів автоматичного сентимент-аналізу текстів [Іванов, & Коваленко, 2017; Оленич та ін., 2021], дослідженню негативної тональності різнотипних текстів [Shkvorchenko, 2023; Paul et al., 2024] та ін.

Наша ідея створення автоматичної системи атрибуції українськомовних текстів, яка, з-поміж іншого, буде визначати негативний sentiment, виникла на основі багаторічних досліджень колективу в галузі комп'ютерної лінгвістики загалом та цікавості до sentiment-аналізу зокрема [Дарчук та ін., 2016; Дарчук, 2019; Зубань, & Білоус, 2018; Толочко, 2023].

Дослідницький корпус мережових медіатекстів політичного дискурсу обсягом 10 млн слововживань [Корпус української мови, 2003–2024] засвідчив, що в українському інформаційному просторі з'являється все більше текстів, які містять мову ворожнечі та не нехтують різноманітними прийомами токсичного впливу на читача. Саме тому актуальним вважаємо виявлення насамперед негативного sentimentу медіатекстів з метою нейтралізації їхньої токсичної дії.

Методи

Автоматичне визначення індексу токсичності є окремим модулем системи “TextAttributor 1.0” [TextAttributor 1.0, 2024]. У створенні системи використано методи компонентного аналізу, дистрибутивного аналізу, безпосередніх складників, квантування. Індекс токсичності тексту обчислюється за допомогою вербальної ідентифікації токсичного sentimentу тексту та розв'язується двома методами: методом словників і правил (обчислення статистичних параметрів) і методом машинного навчання.

Результати

Високий індекс токсичності є ознакою токсичного тексту, який ми трактуємо широко – як публікацію, що містить роздратування, домагання, погрози, непристойні коментарі, кіберзалякування, ідкі зауваження та злісні жарти (тролінг), обурення про політику, життя, друзів, мову ворожнечі. Мовою ворожнечі, своєю чергою, вважаємо “будь-яку комунікацію, що принижує чи дискримінує особу або групу осіб” [Словник з медіаграмотності, 2022] на основі того, ким вони є.

Оскільки важливою сферою, яка завжди привертала підвищену увагу суспільства, є політичний дискурс, матеріалом нашого дослідження стали дописи в інтернеті, тексти новинних

стрічок, трансльовані телебаченням, статті в інтернет-виданнях різних регіонів України, коментарі до політичних публікацій тощо.

Завдання дослідження передбачали два послідовні етапи виконання: 1) створення системи, яка генерує лінгвістичну експертизу токсичного тексту із визначенням індексу токсичності за словниками та за правилами; 2) формування навчальної вибірки за наявною системою для машинного навчання і створення нейронної мережі визначення індексу токсичності тексту.

Попередній лінгвістичний аналіз дослідницького корпусу дозволив виокремити підкорпус токсичних текстів, на основі яких виконавці проекту визначали лексеми та сполуки з негативним сентиментом за попередньо розробленими класифікаціями, тональними словниками та відповідно маркували їх. Результатом цієї роботи стала лексикографічна база даних, до якої увійшли три словники: емоціогенів, мови ворожнечі й токсичних сполук.

Лексикографічний список № 1 “Словник емоціогенів” містить 5 000 лексем із негативною тональністю. До списку увійшли повнозначні слова (іменники, прикметники, прислівники та дієслова) з негативною тональністю (за шкалою –2), напр., *аморальний, безкарність, хабарництво, корупція, хабарник, красти, злодіяти, жахливо* (рис. 1).

Слово: Знайти

☒	багнюка	негативна тональність
☒	базікання	негативна тональність
☒	байдуже	негативна тональність
☒	байдужий	негативна тональність
☒	байдужість	негативна тональність
☒	байдушний	негативна тональність
☒	байдушно	негативна тональність
☒	балаган	негативна тональність
☒	баламутний	негативна тональність
☒	баланда	негативна тональність
☒	балачка	негативна тональність
☒	балуваний	негативна тональність
☒	банда	негативна тональність
☒	бандитизм	негативна тональність
☒	бандитський	негативна тональність

Рис. 1. Фрагмент списку емоціогенів

Цей словник було сформовано за тональними словниками Н. Дарчук [Дарчук та ін., 2016] та О. Толочко [Толочко, 2023] обсягом 9 321 одиниця з визначенням тональності за бінарною шкалою (позитив/негатив), укладеним за лексиконом українськомовної публіцистичної лексики обсягом 40 тис. лексем.

Лексикографічний список № 2 “Словник мови ворожнечі” охоплює 3 000 лексем. До цього списку увійшли такі групи лексики: негативні назви людей – 1 620 (напр.: *барига, западенець, мент, риг, тарган, тварюка, язичник*), обсценна лексика – 613 (напр.: *вз*бати, ви*бон, відпо*ерити, з*обувати, о*уїти, роз*об*) та лайлива лексика й вульгаризми – 787 (напр.: *всирачка, обдовбуватися, облапати, радіосрач, самодрочіння, серун, сміхуєчок, скурвитися, соціоблядствувати, сраний, схарений*). Одиниці списків були марковані семантичними характеристиками за розробленими класифікаціями. Список № 2 охоплює 18 семантичних ознак, напр.: негативні назви людини за віком, зовнішністю, асоціальною поведінкою, належністю до політичного напрямку, місцем проживання, віросповідання, інтелектуальними здібностями, за порівнянням із рослиною чи твариною тощо (рис. 2). Кожне слово мови ворожнечі отримало мітку за однією із них, напр.: **s** – сексизм, **r** – расизм, **e** – ейджизм та ін.

Класифікатор: Негативні назви людини darчук

☐ за віком	бюба, бабуля
☒ за асоціальною поведінкою	убивця, злодій, крадій
☒ за негативними рисами характеру	егоїст
☒ за зовнішністю	дриш, чмо
☒ за назвами частин тіла або фізіологічними процесами	гіпно
☒ за належністю до політичного напрямку	зобіп
☒ за ознаками захворювання	афганський карлик, алкаш, торчок
☒ за ознакою соціального статусу або за сімейним станом	бомж
☒ за порівнянням із міфічними істотами	відьма, гном, зомбі
☒ за порівнянням із рослинами	зелена пілскава
☒ за порівнянням із тваринами	свиня, криса, лвар
☒ за проживанням на певній території	селючка, западенець
☒ за професією, майновим станом	мент, барига
☒ за релігійними ознаками	сатаніст, упоротий сектант
☒ за розумовими інтелектуальними здібностями	діот
☒ за соціальною активністю/пасивністю	заводила, пофіліст
☒ расизм – за національністю	жид, негритос
☒ сексизм, завуальшоване висловлювання	соблюдателі традицій, Унікальна ви наша

Рис. 2. Семантичні класифікаційні ознаки категорії “негативні назви людини” словника мови ворожнечі

Лексикографічний список № 3 “Токсичні сполуки” складають 1 500 стійких сполук, що відображають мову ворожнечі, різного типу негативно забарвлені експресивні одиниці, токсичні словосполучення, тобто фрази, які виражають негативний сентимент ідіоматично. У цьому списку одиниці класифіковано за 26 семантичними ознаками. Кожній сполуці присвоєна семантична мітка токсичності, мови ворожнечі чи експресивності, напр.: **МВ 2.** Створення негативного образу (*дурнуваті харі крішани, гейропський лесбуха*); **МВ 4.** Твердження про неповноцінність, моральні недоліки (*нахабний хач, роми – брехуни*); **МВ 6.** Звинувачення в негативному впливі на суспільство, державу, у спробах захоплення влади або в територіальній експансії (*влада жидів, майдаунний імбецил, кривава хунта*); **МВ 10.** Погрози про фізичне знищення або будь-яке насильство (*смерть жидам, денацифікація київського режиму*); **Т 1.** Заклики до насильства (*В шию його!, наваляти по безтолковці*); **Т 2.** Створення негативного образу (*алкаш запухлий, бабуля Юля, безголовий кандидат, дешева обгортка*); **Е 1.** Емоційне роздратування автора повідомлення, обурення про політику/життя/друзів тощо (*абсолютне зло, адський фетини*); **Е 2.** Звинувачення в негативному впливі, твердження про злочини (*вандальський набіг, диверсійно-розвідувальна група ворога*) тощо (рис. 3).

Класифікатор: klasifikator.spiluk@ukr.net

МВ 1. Емоційне роздратування автора повідомлення, обурення про політику/життя/друзів тощо	абсолютне зло, адський фетини
МВ 2. Звинувачення в негативному впливі, твердження про злочини	вандальський набіг, диверсійно-розвідувальна група ворога
МВ 3. Експресивні	у якого це хач?
МВ 4. Емоційні фразеологізми	стати кінцем, цілувати дупу
МВ 5. Заклики до насильства, будь-якого: дискримінації.	захопити Україну, росквіля – на нока
МВ 6. Погрози про фізичне знищення або будь-яке насильство (англ. Threat).	смерть жидів, денацифікація київського режиму
МВ 7. Заклики до насильства, будь-якого.	гейроп!, не лесбуй!
МВ 8. Створення негативного образу.	дурнуваті харі крішани, гейропський лесбуха
МВ 9. Випрадання та твердження про історичні випадки насильства, злочини чи кримінальність, перекошування та суніння щодо історичних фактів насильства та дискримінації.	«бандерці» винищували народи разом із фашистами, київські катофашисти
МВ 10. Твердження про неповноцінність, моральні недоліки.	нахабний хач, роми – брехуни
МВ 11. Звинувачення, твердження про злочини чи кримінальність (не історичні).	цигані обчищають будинок пенсіонера
МВ 12. Звинувачення в негативному впливі на суспільство, державу, у спробах захоплення влади або в територіальній експансії.	влада жидів, майдаунний імбецил, кривава хунта
МВ 13. Згадка у привласненню або образливому контексті, непристойні конспірації.	мусоризація слова не разали, ідувати сидопатні і мусорозачани
МВ 14. Звинувачення в промисловості.	небажати та (до) непереносимі пропеліції сексуального характеру
МВ 15. Денацифікація, денацифікація (англ. Denazification).	а шню його!, наваляти по безтолковці
Т 1. Заклики до насильства.	жирний осел, ксьенний напелетка
Т 2. Цікування.	панове т тушководи, черевний иракобіс
Т 3. Ідії зауваження та злочини жарту, націєни на провокацію у читачів емоційної відгуки, або тролінг (англ. Trolling).	алкаш запухлий, бабуля Юля, безголовий кандидат, дешева обгортка
Т 4. Створення негативного образу.	аморальний пошукчик, бездушна тварина
Т 5. Твердження про неповноцінність, моральні недоліки.	агент Креліш, еллітніська шайка
Т 6. Звинувачення, твердження про злочини чи кримінальність (не історичні).	авадаські пси, вапний реваншист
Т 7. Звинувачення в негативному впливі.	йди кифе, валя вже за крати, за парабрі
Т 8. Згадка у привласненню або образливому контексті, непристойні конспірації.	«говориш» довідаєш, амеросіані слабуни, жидівний кур
Т 9. Непристойні жарти та нотки.	я би мави, давати їси парич
Т 10. Погрози про фізичне знищення.	евакуувати в пекло, на плавуку

Рис. 3. Семантична класифікація токсичних сполук

Зауважмо, що укладена лексикографічна база даних – це складна поліпараметрична система, у якій одиниці (слова, синтагми) супроводжуються набором семантичних ознак, що у сполученні з кількісними даними частотного словника аналізованого тексту забезпечує аналіз токсичності цього тексту, представлений у системі “TextAttributor 1.0” такими способами: 1) індексом токсичності тексту; 2) статистичною картою токсичності тексту; 3) визначенням у тексті слів із негативним настроєм.

Індекс токсичності тексту (*itox*) вираховується за частотою вживання в тексті різних класів лексики з негативною тональністю: слів-емоціогенів (*e*), лексем мови ворожнечі (*m*), токсичних стійких сполук (*t*). Також у формулі індексу токсичності враховується коефіцієнт (*K*), що посилює значення слів на означення мови ворожнечі і токсичних сполук, який дорівнює –2, та обсяг тексту (*n*):

$$Itox = (e + |K| (m + t)) / n \times 10,$$

Наприклад, індекс токсичності тексту *Націю формують перемоги, а не політкоректність* (119 речень, 1018 слововживання) є високим – 0,81, однак типовим для середніх за довжиною текстів медійного стилю (рис. 4).

Характеристика негативного настрою тексту:

- (itox-0.81) **Типові ознаки медійного стилю:** індекс токсичності тексту формує когнітивно-прагматичні установки на негативні значення 0,2 - 0,9

ПОРІВНЯННЯ АТРИБУЦІЇ ТЕКСТІВ

Підрахувати векторну відстань

ЛІНГВІСТИЧНА ЕКСПЕРТИЗА ТОКСИЧНОСТІ ТЕКСТУ

Категорія	Назва	Кількість
Емоціогени	негативна тональність	57
Класифікація сполук	T3. Твердження про неповноцінність, моральні недоліки.	1
Фразеологізми		10
Класифікація сполук	МВ 5. Звинувачення, твердження про злочини чи кримінальність (не історичні).	1
Негативні назви людини	за професією, майновим станом	1
Негативні назви людини	за соціальною активністю/пасивністю	3
Негативні назви людини	за асоціальною поведінкою	4
Негативні назви людини	за проживанням на певній території	1
Класифікація сполук	T4. Звинувачення, твердження про злочини чи кримінальність (не історичні).	1

Націю формують перемоги, а не політкоректність

Рис. 4. Індекс токсичності та лінгвістична експертиза токсичності тексту

Лінгвістична експертиза токсичності тексту відображає статистичну картку семантичних класів негативної лексики за класифікаційними маркерами лексикографічних списків. Зокрема, в аналізованому тексті трапляються такі категорії лексики з негативним сентиментом: емоціогени з негативною тональністю – 57 (іменники *війна, ворог, беззаконня, біда, діди́вщина, зло, репресії, сутички, тортури, прикметники жорстокий, проросі́йський*, дієприкметник *окупований*, прислівник *хитро*, дієслова *бити, вимагати, загинути, знищувати, плакати* та ін.), токсичні сполуки (*Московський патріархат, тотальна корупція, огидна сатанинська Московія*), негативні назви людини за різними ознаками (*ворог, зрадник, раб, убивця* тощо).

Окремо визначено стійкі сполуки та фразеологізми, а також деякі лексеми, що не впливають на токсичність тексту, вони підсвічені іншими кольорами. Текст із визначеними слововживаннями негативного сентименту, що вербалізують категорії статистичної картки, подано на рис. 5.

Націю формують перемоги, а не політкоректність

Я сам із гідного сходу. Знаю багатьох людей, яких зачепили **Револуція гідності** й **війна** на Донбасі. Але вони все одно залишилися з **патриотичними** поглядами. Нині частина України – **окупована**, багато хто виїхав за кордон. Діялося на їхні думки в **соцмережах** – люди почали шукати Батьківщину **втрачених** й **полювати** за Україною і **Збройні сили**, тріпотіть повернутися додому. **Війна** змінила суспільство, воно стало більш національним. **На жаль**, щоб українці відчули себе українцями, **потом** була **війна**.

Мама мря, не можна сказати, що була глибоко віруючою, але щоввечора промовляла "Отче наш" і намагалася донести мені, що Бог є. Усі мої колискові були українською.

Мисиста КОХАНОВСЬКА, **Будівельні** – 63 років, лідер добровільного руху Організації українських націоналістів. Справжнє ім'я **Микола Дієба**. Народився 25 травня 1971 року в селі Придольське Закарпатської області. Біл безбатька. Мати працювала на копальні. До 32 років проживав російською мовою, **надавав розгромаження** на чужині та училищі в Мелітополі. 2003-го переїхав до Києва. Працював будівельником, охоронцем. Тоді ж долучився до антивоєнної руху, 2008 року вступив до Кодексу українських націоналістів. Був організатором та учасником багатьох акцій протесту. Вивчав совіту за спеціальністю: "Економіка підприємств", здобув у Перелісві Хмельницького державного педагогічного університету ім. Григорія Сковороди 30 вересня 2007-го вперше на Софійській вулиці запалив Пастушкову вату на честь дня народження Маршала Жуковського. 2011 року вступив до лав ОУН, склавши присягу на вірність Україні. 2013-го **був бранцем** у повстанні пам'ятника Леніну в центрі Києва. Під час Революції гідності був сотником Параски Квилюкської ОУН імені **Батька-Козака** і **Козацька**. З початком російської агресії вів на фронт, де очолював добровільний батальйон ОУН, **чор бранця**, зокрема, у **густоті** Савур-Могили, захистив Донецького аеропорту. У травні 2015 року створив громадянську організацію "Добровільний рух ОУН", де його обрали головою. Проводив з початком великої війни очолює Добровільне формування територіальної громади Києва ОУН. Згодом пішов на фронт. Служив як радийол у відділці Сил Оборони України. У шлюбі одружено. Дружина Катерина, 47 років. Дітей немає. **Загинув** 10 червня в бою на передовій біля Вовчанська на харківському напрямку.

Жив у Києві.

Жив у Києві

У нашій сім'ї було небагато книжок, але "Кобзар" був. Мама читала мені його. Не любив, бо там було багато зажурних речей – і я плакав. Згодом читав його замість Букваря.

Українцем усвідомив себе в **совєцькй армії**. Служив у **будбаті в Москві**. Дідівщина й сучічки були саме через те, що я – **українець**.

Рис. 5. Фрагмент тексту з вербалізаторами негативного сентименту

Як бачимо, ця методика дає можливість з'ясувати межу між токсичним і нейтральним текстом, визначити статистику найчастотніших моделей вираження негативного настрою та на підставі властивих тексту лексичних категорій наблизитися, напр., до розв'язання проблеми автоматичного виявлення наративу тексту.

Другий режим роботи системи “TextAttributor 1.0” – машинне навчання. На вході системи ми мали корпуси текстів, розмічених за принципами, що передбачають виявлення наявності або відсутності в тексті ознак токсичності. Кожен корпус розділявся на навчальну та контрольну вибірки, а також, для достатньо великих корпусів, – валідаційну вибірку. На навчальному корпусі ми проводили оцінювання параметрів обраної моделі розв’язання поставленого завдання. На валідаційній вибірці налаштовували гіперпараметри моделі. Остаточні показники ефективності моделі вимірювали на контрольній вибірці з урахуванням доступності обчислювальних ресурсів, які необхідні для експлуатації моделі. У створенні системи “TextAttributor 1.0” за методом машинного навчання було сформовано і використано не достатньо великі корпуси відносно коротких українськомовних інтернет-текстів (блоги, коментарі, статті тощо). Унаслідок серії експериментальних досліджень у межах описаного проєкту було обрано ефективну для обчислень архітектуру на основі методу fastText і його інструментальних засобів [Joulin et al., 2016]. Цей метод базується на векторному представленні слів та оцінці розподілу ймовірності документів за наперед визначеними класами. Слова подаються у векторному вигляді на підставі автоматичного розбиття слів на частини (subwords), що моделює відкритість словника. Останнє є важливим з огляду як на високу флективність української мови, так і на необхідність враховувати наявні в текстах нові, ще не внесені у словник, слова, а також слова з орфографічними помилками. Ефективність використаного підходу також визначається технічними умовами експлуатації системи та текстовими ресурсами для навчання моделей. Для бінарної класифікації, яка виявляє документи з токсичністю, було використано підготовлений корпус обсягом приблизно 12 тисяч документів (текстів). Найкращий результат за узагальненою метрикою ($F1 = 79,4\%$) отримано для таких значень гіперпараметрів: розмірність простору векторного представлення слів – 56, початкова швидкість навчання – 0,15, кількість епох – 500, лексичний контекст – біграми, довжина

субслів – від 2 до 5, поріг прийняття рішення – 0,4. Водночас за чутливості 85 % досягається точність близько 70 %.

Аналіз токсичності тексту *Націю формують перемоги, а не політкоректність* методом машинного навчання показав нульовий ступінь токсичності, що загалом підтверджує результати лінгвістичної експертизи за методом словників і правил про допустимий обсяг слів з негативним сентиментом для такого обсягу тексту медійного стилю.

Дискусія і висновки

Автоматична система “TextAttributor 1.0” перебуває на етапі тестування визначення токсичності українськомовних текстів, проте отримані результати свідчать, що розроблена методика квантування вербальних засобів за семантичними параметрами (негативна емотивність) за методами словників та правил і методами машинного навчання є ефективною для досягнення поставлених завдань. Режим роботи модуля на основі словників і правил дає можливість не просто отримати висновок про негативну тональність окремого тексту, а й уявляє мовні засоби, якими вона досягається. Окрім того, методика словників і правил забезпечила автоматичне формування датасету для машинного навчання в завданні визначення токсичності тексту. Реалізація цього завдання на основі глибокого машинного навчання дала змогу отримати результати, зіставні зі системами для інших мов, параметри яких оцінено на вибірках із подібними характеристиками. Система “TextAttributor 1.0” є зручною й ефективною для використання в різноманітних текстологічних дослідженнях.

У перспективі розроблений інструмент автоматичного визначення токсичності тексту плануємо поєднати зі семантичним та синтаксичним автоматичним аналізом, а також із психолінгвістичним і соціолінгвістичним експериментами з метою встановлення фактів впливу на читача. Описана методика у проєкції на семантичний аналіз лексичної таксономії, розроблений у лабораторії комп’ютерної лінгвістики [Дарчук та ін., 2016], дасть можливість на підставі властивих тексту семантичних категорій наблизитися до розв’язання проблеми автоматичного виявлення

наративу тексту, поглибити текстологічний опис впровадженої лінгвістичної експертизи, відкрис шлях до відстежування поширювачів фейків тощо.

Внесок авторів: Наталія Дарчук – концептуалізація, методологія, написання (перегляд і редагування); Оксана Зубань – методологія, валідація даних, написання (перегляд і редагування); Валентина Робейко – методологія, валідація даних; Юлія Цигвинцева – методологія, валідація даних, написання (оригінальна чернетка).

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

Дарчук, Н., Зубань, О., Лангенбах, М., & Ходаківська, Я. (2016). АГАТ-семантика: семантичне розмічування корпусу української мови. *Українське мовознавство*, 1(46), 92–102. [https://doi.org/10.17721/um/46\(2016\)](https://doi.org/10.17721/um/46(2016))

Дарчук, Н. (2019). Лінгвістичні засади автоматичного сентимент-аналізу українськомовного тексту. *Science and education a new dimension*, 189, 10–13. <https://doi.org/10.31174/SEND-Ph2019-189VII55-02>

Залуцька, О., Молчанова, М., Мазурець, О., Мельник, О., & Скрипник, Т. (2023). Метод інтелектуального аналізу емоційної тональності текстової інформації для визначення поведінкових намірів нейромережевими засобами. *Вісник Хмельницького національного університету. Технічні науки*, 1, 5, 67–73. <https://elar.khmnmu.edu.ua/server/api/core/bitstreams/7d4a9f0d-53d5-4633-bf69-e360e5206a70/content>

Зубань, О. М., & Білоус, В. В. (2018). Автоматичне визначення лексичної тональності українськомовного тексту (на матеріалі новин ТСН). *Інформаційні технології та взаємодії (IT&I'2018)*, (с. 182–183). <http://iti.fit.univ.kiev.ua/wp-content/uploads/ITI-2018.pdf>

Іванов, Є. М., & Коваленко, С. В. (2017). Розробка web-додатка для аналізу тональності текстової інформації. *Інформаційні технології: наука, техніка, технологія, освіта, здоров'я*, 1, 18–19. <https://core.ac.uk/download/pdf/161789959.pdf>

Корпус української мови (2003–2024). *Mova.info: лінгвістичний портал*. <http://www.mova.info/corpus.aspx>

Оленич, І., Притула, М., Сінкевич, О., & Хамар, О. (2021). Система автоматичного визначення тональності тексту. *Електроніка та інформаційні технології*, 15, 16–23. <https://doi.org/10.30970/eli.15.2>

Почепцов, Г. (2016). *Сучасні інформаційні війни*. Видавничий дім “Києво-Могилянська академія”.

Романюк, А., & Романишин, М. (2013). Тональний словник української мови на основі сентимент-анотованого корпусу. *Українське мовознавство*, 43, 63–74.

Словник з медіаграмотності (2022). <https://filter.mkip.gov.ua/wp-content/uploads/2022/10/slovnuk.pdf>

Стахів, М., & Скопівський, М. (2024). Від аналізу тексту до моделювання природної мови: комплексне дослідження. *Вісник Хмельницького національного університету. Серія: Технічні науки*, 333(2), 242–250. <https://doi.org/10.31891/2307-5732-2024-333-2-38>

Толочко, О. (2023). *Тональний словник української мови*. <https://github.com/Oksana504/sentimentdictionary-uk>

Шаховська, Н. Б., & Гірак, Х. Ю. (2017). Шкалювання емоційно забарвлених слів для використання у методах класифікації тональності. *Вісник Національного університету “Львівська політехніка”*, 872(1), 195–203. <https://science.lpnu.ua/sites/default/files/journal-paper/2018/jun/13016/ilovepdfcom-195-203.pdf>

Шингалов, Д. В., Мелешко, Є. В., Минайленко, Р. М., & Резніченко, В. А. (2017). Методи автоматичного аналізу тональності контенту у соціальних мережах для виявлення інформаційно-психологічних впливів. *Техніка в сільськогосподарському виробництві, галузеве машинобудування, автоматизація*, 30, 196–202. http://nbuv.gov.ua/UJRN/znpkntu_2017_30_29

Ялова, К., Яшина, К., Говорущенко, Т., & Тарасюк, О. (2021). Сентимент-аналіз засобами нейронної мережі. *Математичне моделювання*, 1(44), 30–37. [https://doi.org/10.31319/2519-8106.1\(44\)2021.235906](https://doi.org/10.31319/2519-8106.1(44)2021.235906)

Cambria, E., Zhang, X., Mao, R., Chen, M., & Kwok, K. (2024). SenticNet 8: Fusing emotion AI and commonsense AI for interpretable, trustworthy, and explainable affective computing. In *International Conference on Human-Computer Interaction (HCI)*. <https://sentic.net/publications/>

Cui, J., Wang, Z., Ho, S.-B., & Cambria, E. (2023). Survey on sentiment analysis: evolution of research methods and topics. *Artificial Intelligence Review*, 56, 8469–8510. <https://doi.org/10.1007/s10462-022-10386-z>

Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., & Mikolov, T. (2016). FastText.zip: Compressing text classification models. *Cornell University. Computer Science. Computation and Language*. <https://doi.org/10.48550/arXiv.1612.03651>

Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11, 81. <https://doi.org/10.1007/s13278-021-00776-6>

Paul, J., Das Chatterjee, A., Misra, D., Majumder, S., Rana, S., Gain, M., & Sil, J. (2024). A survey and comparative study on negative sentiment

analysis in social media data. *Multimedia Tools and Applications*, 1–50. <https://doi.org/10.1007/s11042-024-18452-0>

Reveillac, M., & Morselli, D. (2022). Dictionary-based and machine learning classification approaches: a comparison for tonality and frame detection on Twitter data. *Political Research Exchange*, 4(1), 182–184. <https://doi.org/10.1080/2474736X.2022.2029217>

Shkvorchenko, N. (2023). Political toxicity in the contrastive perspective (Based on American, British and Ukrainian media discourse). *Cognition, communication, discourse*, 26, 152–173. <https://doi.org/10.26565/2218-2926-2023-26-09>

TextAttributor 1.0 (2024). <http://ta.mova.info>

TextAttributor 1.0. Методика (2024). <http://ta.mova.info/methods>

REFERENCES

Cambria, E., Zhang, X., Mao, R., Chen, M., & Kwok, K. (2024). SenticNet 8: Fusing emotion AI and commonsense AI for interpretable, trustworthy, and explainable affective computing. In *International Conference on Human-Computer Interaction (HCII)*. <https://sentic.net/publications/>

Corpus of the Ukrainian language (2003–2024). *Mova.info: linguistic portal* [in Ukrainian]. <http://www.mova.info/corpus.aspx>

Cui, J., Wang, Z., Ho, S.-B., & Cambria, E. (2023). Survey on sentiment analysis: evolution of research methods and topics. *Artificial Intelligence Review*, 56, 8469–8510. <https://doi.org/10.1007/s10462-022-10386-z>

Darchuk, N., Zuban, O., Lanhenbakh, M., & Khodakivska, Ya. (2016). AGAT-semantics: semantic markup of the Ukrainian language corpus. *Ukrainian linguistics*, 1(46), 92–102 [in Ukrainian]. [https://doi.org/10.17721/um/46\(2016\)](https://doi.org/10.17721/um/46(2016))

Darchuk, N. (2019). Linguistic approach for development of computer-based sentiment analysis in the Ukrainian language. *Science and education a new dimension*, 189, 10–13 [in Ukrainian].

Dictionary of media literacy (2022) [in Ukrainian]. <https://filter.mkip.gov.ua/wp-content/uploads/2022/10/slovnyk.pdf>

Ivanov, Ye. M., & Kovalenko, S. V. (2017). Development of a web application for analysing the tone of textual information. *Information technology: science, engineering, technology, education, health*, 1, 18–19 [in Ukrainian]. <https://core.ac.uk/download/pdf/161789959.pdf>

Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., & Mikolov, T. (2016). FastText.zip: Compressing text classification models. *Cornell University. Computer Science. Computation and Language*. <https://doi.org/10.48550/arXiv.1612.03651>

Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11, 81. <https://doi.org/10.1007/s13278-021-00776-6>

Olenych, I., Prytula, M., Sinkevych, O., & Khamar, O. (2021). System of automatic determination of text tone. *Electronics and information technologies*, 15, 16–23 [in Ukrainian]. <https://doi.org/10.30970/eli.15.2>

Paul, J., Das Chatterjee, A., Misra, D., Majumder, S., Rana, S., Gain, M., & Sil, J. (2024). A survey and comparative study on negative sentiment analysis in social media data. *Multimedia Tools and Applications*, 1–50. <https://doi.org/10.1007/s11042-024-18452-0>

Pocheptsov, H. (2016). *Modern information wars*. Vydavnychiy dim “Kyievo-Mohylianska akademiia” [in Ukrainian].

Romaniuk, A., Romanyshyn, M. (2013). Tonal dictionary of the Ukrainian language based on sentiment-annotated corpus. *Ukrainian linguistics*, 43, 63–74 [in Ukrainian].

Reveillac, M., & Morselli, D. (2022). Dictionary-based and machine learning classification approaches: a comparison for tonality and frame detection on Twitter data. *Political Research Exchange*, 4(1), 182–184. <https://doi.org/10.1080/2474736X.2022.2029217>

Shakhovska, N., & Hirak, K. (2017). Ranging of emotive words for the use in the methods of tone classification. *Bulletin of Lviv Polytechnic National University*, 872(1), 195–203 [in Ukrainian]. <https://science.lpnu.ua/sites/default/files/journal-paper/2018/jun/13016/ilovepdfcom-195-203.pdf>

Shkvorchenko, N. (2023). Political toxicity in the contrastive perspective (Based on American, British and Ukrainian media discourse). *Cognition, communication, discourse*, 26, 152–173. <https://doi.org/10.26565/2218-2926-2023-26-09>

Shyngalov, D., Meleshko, Ye., Mynaylenko, R., & Reznichenko, V. (2017). Methods of automated sentiment analysis on social networks. *Machinery in agricultural production, industry machine building, automation*, 30, 196–202 [in Ukrainian]. http://nbuv.gov.ua/UJRN/znpkntu_2017_30_29

Stakhiv, M., & Skopivskyi, S. (2024). From text analysis to natural language modeling: a comprehensive exploration. *Herald of Khmelnytskyi national university. Series: Technical sciences*, 333 (2), 242–250 [in Ukrainian]. <https://doi.org/10.31891/2307-5732-2024-333-2-38>

TextAttributor 1.0 (2024) [in Ukrainian]. <http://ta.mova.info>

TextAttributor 1.0. Methodology (2024) [in Ukrainian]. <http://ta.mova.info/methods>

Tolochko, O. (2023). *Tonal dictionary of the Ukrainian language* [in Ukrainian] <https://github.com/Oksana504/sentimentdictionary-uk>

Yalova, K., Yashyna, K., Hovorushchenko, T., Tarasiuk, O. (2021). Sentiment analysis with neural network. *Mathematical Modeling*, 1(44), 30–37 [in Ukrainian]. [https://doi.org/10.31319/2519-8106.1\(44\)2021.235906](https://doi.org/10.31319/2519-8106.1(44)2021.235906)

Zalutska, O., Molchanova, M., Mazurets, O., Melnyk, O., & Skrypnyk, T. (2023). Method for intellectual analysis of textual information emotional tonality for determine the behavioral intentions by neural networks means. *Bulletin of Khmelnytsky National University. Technical Sciences*, 1, 5, 67–73 [in Ukrainian]. <https://elar.khmnu.edu.ua/server/api/core/bitstreams/7d4a9f0d-53d5-4633-bf69-e360e5206a70/content>

Zuban, O. M., & Bilous, V. V. (2018). Automatic determination of lexical tone of Ukrainian-language text (based on TSN news). *Information Technology and Interaction (IT&I'2018)*, (pp. 182–183) [in Ukrainian]. <http://iti.fit.univ.kiev.ua/wp-content/uploads/ITI-2018.pdf>

Отримано редакцією збірника / Received: 10.04.24

Прорецензовано / Revised: 03.05.24

Схвалено до друку / Accepted: 10.05.24

Nataliia DARCHUK, DSc (Philol.), Prof.

ORCID ID: 0000-0001-8932-9301

e-mail: n.darchuk@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Oksana ZUBAN, PhD (Philol.), Assoc. Prof.

ORCID ID: 0000-0002-2644-3892

e-mail: oxana.zuban@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Valentyna ROBEIKO, Assist.

ORCID ID: 0000-0003-2266-7650

e-mail: valentyna.robeiko@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Yuliia TSYHVINTSEVA, PhD, Junior Researcher

ORCID ID: 0000-0002-9684-3840

e-mail: tsyhvintseva@nas.gov.ua

Institute of the Ukrainian Language of the National Academy
of Sciences of Ukraine, Kyiv, Ukraine

INDEXING OF THE NEGATIVE SENTIMENT OF THE UKRAINIAN TEXT BY “TEXTATTRIBUTOR 1.0”

Background. In the twenty-first century, the information space is a full-fledged battlefield. In the Ukrainian information space, the problem of text toxicity and hate speech is becoming increasingly important. Therefore, the interest of researchers in markers of negative textual tone,

especially in media texts, is constantly growing. The article describes the structure and results of a separate module of the automatic system of statistical parameterization of Ukrainian-language texts “TextAttributor 1.0” – determination of the text toxicity index.

The tasks are solved by two methods: the method of dictionaries and rules (calculation of statistical parameters) and the method of machine learning.

The results of the study are based on the material of the corpus of online media texts of political discourse with a volume of 10 million word occurrences. To achieve this goal, a lexicographic database was created, including three dictionaries: Emotiogens, Hate Speech, and Toxic Compounds, and training and control samples of texts were formed to estimate the parameters of the selected model using machine learning. The project chose a computationally efficient architecture based on the fastText methodology and tools. The toxicity index is calculated by verbally identifying the negative sentiment of the text based on the linguistic and software-generated and is detected by the system-generated linguistic examination of the text, which displays a statistical map of semantic classes of negative vocabulary by classification markers of lexicographic lists, and the output of the neural network.

Conclusions. The “TextAttributor 1.0” system is at the stage of testing and improving its functionality, so the article describes an intermediate β -version of the system, but the results obtained in determining toxicity show that the developed methodology for quantifying verbal means by semantic parameters (negative emotionality) using dictionaries and rules and machine learning is effective in achieving the tasks set and makes it possible not only to determine the boundary between toxic and neutral text but also to approach the solution of the problem based on the lexical categories inherent in the text. The methodology for developing a module for determining the toxicity of media text in the “TextAttributor 1.0” system was described and published on the web application page in April 2024, but this information is published for the first time in the form of a research article.

Key words: *Ukrainian-language media text, toxic text, hate speech, sentiment, negative sentiment, text toxicity index.*

Проект реалізовано за підтримки Посольства Великої Британії в Україні спільно з громадською організацією “Український науковий

центр лінгвістичних студій” (директор – Костянтин Гончаренко). Дякуємо всім, хто долучився до організації, супроводу та виконання окремих завдань і тестування системи на всіх етапах роботи.

Погляди, висловлені в цій публікації, належать розробникам проєкту і можуть не збігатися з офіційною позицією уряду Великої Британії.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; у рішенні про публікацію результатів.