

УДК 811.111'37'42  
DOI <https://doi.org/10.32782/fohium/2024.5.14>

## ТОНАЛЬНОСТІ «КАЗОК БАРДА БІДЛА» ЯК ВИЯВ ОСОБЛИВОСТЕЙ ПЕРЕКЛАДУ

**Роксолана Назарчук**

*кандидат філологічних наук, доцент,  
доцент кафедри прикладної лінгвістики,  
Національного університету «Львівська політехніка»  
ORCID ID 0000-0002-0482-2883  
roksolana.z.nazarchuk@2023@lpnu.ua*

**Софія Лисяк**

*студентка магістратури,  
Інститут комп'ютерних наук та інформаційних технологій  
Національного університету «Львівська політехніка»  
ORCID ID 0009-0004-4384-1755  
sofia.lysiak.mfpl.2023@lpnu.ua*

**Ключові слова:** *сентимент-аналіз, LIWC, позитивна тональність, негативна тональність, текст оригіналу, текст перекладу.*

Запропоновану статтю присвячено аналізу особливостей перекладу «Казок барда Бідла» на українську мову. Зокрема, увагу зосереджено на сентимент-аналізі окремих оповідань зі збірки Дж. К. Роулінг: «Чаклун і стрибучий казанок», «Фонтан фортуни» та «Волохате серце мага» в оригіналі та їхньому українському перекладі В. Морозова за редакцією О. Негребецького та І. Малковича. Визначено прикметні риси перекладу згаданих оповідань відповідно до показників тональності. Закцентовано увагу на залежності тональності від відповідності перекладу, вибору лексики та вжитку трансформацій, а також від самого жанру досліджуваного матеріалу. Узагальнено комплекс теоретичних питань, пов'язаних із постановням та розвитком сентимент-аналізу, тлумаченням поняття тональності тексту, зокрема окреслено завдання, поставлені перед аналізом думок, та методи їх досягнення. Визначено поняття сентимент-аналіз як сукупність методів комп'ютерної лінгвістики, що спрямовано на виявлення присутності емоційних настроїв у текстах. Описано загальні характеристики та можливості програмного забезпечення LIWC-22, а також особливості останньої версії у порівнянні до попередніх. За допомогою застосування проаналізовано три оповідання разом із доданими до них коментарями «Албус Дамблдор коментує...» у їхньому оригіналі та перекладі на українську. Визначено й протлумачено кількісні показники, які відрізняють оригінал і переклад, категорій позитивно та негативно забарвленої лексики, а також лайки, яка не належить до жодної з цих груп. У статті наведено графіки для чіткого окреслення результатів, а також додано середні показники позитивної та негативної тональності в оповіданнях для порівняння з отриманими результатами. Обґрунтовано перспективи дослідження текстів з використанням сентимент-аналізу та функцій програмного забезпечення LIWC-22.

## THE TONES OF ‘THE TALES OF THE BARD OF BEADLE’ AS A MANIFESTATION OF TRANSLATION PECULIARITIES

**Roksolana Nazarchuk**

*PhD in Linguistics, Associate Professor,  
Lviv Polytechnic National University*

**Sofia Lysiak**

*Master's student,  
Institute of Computer Science and Information Technologies,  
Lviv Polytechnic National University*

**Key words:** *sentiment analysis, LIWC, positive tone, negative tone, source text, target text.*

The proposed article is devoted to the analysis of the peculiarities of the translation of ‘The Tales of the Bard of Beadle’ into Ukrainian. In particular, the attention is focused on the sentiment analysis of individual stories from J.K. Rowling's collection ‘The Tales of the Bard of Beadle’: ‘The Wizard and the Hopping Pot’, ‘The Fountain of Fair Fortune’ and ‘The Warlock’s Hairy Heart’ in the original and their Ukrainian translation by V. Morozov, edited by O. Negrebetsky and I. Malkovych. The distinctive features of the translation of the mentioned stories in accordance with the tone indicators are determined. Attention is focused on tone dependence on the translation's relevance, the choice of vocabulary and the use of transformations, and the genre of the material under study. A set of theoretical issues related to the emergence and development of this field, the interpretation of the concept of text tone are summarised, and the tasks set for the analysis of opinions and the methods of their achievement are outlined. The article defines the concept of ‘sentiment analysis’ as a set of computer linguistics methods aimed at detecting the presence of emotional moods in texts (emotionally coloured vocabulary). The general characteristics and capabilities of the LIWC-22 software are described, as well as the improvements of the latest version compared to the previous ones. Using this application, three short stories were analysed together with the attached comments ‘Albus Dumbledore on...’ in their original and Ukrainian translations. The quantitative indicators of the categories of positively and negatively coloured vocabulary, as well as swearing that do not differ significantly between the original and the translation, are identified and analysed. The article provides graphs to outline the results clearly and includes average indicators of positive and negative tones in the stories for comparison with the results obtained. Prospects for the study of texts using sentiment analysis and the functions of the LIWC-22 software are substantiated.

**Вступ.** Через зростання кількості текстів загалом й літературних творів зокрема, а також постійну потребу в їхньому відтворенні різними мовами виникає нагальна необхідність відбору та відстеження якості перекладу. Одним з ефективних методів оцінки є порівняння тональності текстів перекладу та оригіналу, що дає змогу виявити відмінності в емоційній насиченості, стильових особливостях, підібраній лексиці та точності передачі сенсу. Щоб забезпечити таку оцінку, використовують сучасні лінгвістичні прийоми й методики, які автоматизують аналіз великих текстів. Це дає змогу не лише порівнювати оригінал і переклад, але й здійснювати процес з високим рівнем об'єктивності та точності. Глумачимо

застосування LIWC (Linguistic Inquiry and Word Count) як один із можливих прийомів зіставлення мови оригіналу та перекладу, що спрощує процес аналізу та підведення підсумків стосовно якості перекладу за допомогою автоматично підрахованих кількісних показників.

**Матеріал і методи дослідження.** Сентимент-аналіз – галузь лінгвістики, яка зародилася минулого сторіччя та з початку цього набрала великих обертів як у теоретичному, так і практичному аспектах. Серед сучасних досягнень машинного навчання в аналізі тональностей є ряд програмного забезпечення, що уможливорює швидкі та точні кількісні підрахунки (Pennebaker et al., 2022; Pennebaker, 2007). З-поміж важливих

напрямків розвитку галузі, якими зацікавлені закордонні дослідники, стало дослідження алгоритмів машинного навчання та опрацювання новаторських підходів для аналізу відгуків споживачів та інших текстів (Б. Ліу (Liu, 2010; Liu, 2012), аналіз настроїв у масивах даних (Л. Лей та Д. Ліу (Lei, 2021)), використання комбінованих моделей машинного навчання та покращення інтерпретації текстових даних (К. Іглесіас та А. Морено (Iglesias, 2019)). Об'єктом зацікавлення українських колег є дослідження впливу культурних та мовних особливостей на результати настроєвого аналізу (Дарчук, 2019), проблематика автоматизованого розпізнавання настроїв у різних жанрах текстів та вдосконаленні алгоритмів для обробки неформального та мікротексту (Немеш, 2015; Онищенко, 2004).

Тональність літературного перекладу має вирішальне значення для передачі емоційного забарвлення тексту, яке забезпечує автентичне сприйняття читачем. За перекладу оповідань важливо не лише передати зміст, а й зберегти тональність, що дозволяє відтворити ефект оригінального твору, дублюючи усі його особливості та новаторства автора. Використання інструментів автоматизованого аналізу, зокрема програми LIWC дає можливість вимірювати емоційне наповнення тексту та порівнювати його з оригіналом.

**Мета статті** – на матеріалі оповідань «Чаклун і стрибуний казанок», «Фонтан фортуни» та «Волохате серце мага» зі збірки «Казки барда Бідла» Дж. К. Роулінг у перекладі В. Морозова за редакцією О. Негребецького та І. Малковича зіставити тональності оригінального й перекладного текстів й таким чином простежити особливості перекладу.

**Результати та обговорення.** Термін настроєвий аналіз використовують для позначення методів опрацювання текстової інформації, що застосовують з метою виявити емоційний тон або настрої, виражений в тексті. Б. Ліу дає цьому терміну більш вичерпне пояснення, стверджуючи, що «аналіз тональності – це низка методів, прийомів й інструментів для виявлення та вилучення суб'єктивної інформації з мови, наприклад, думок, настроїв, оцінок, ставлень та емоцій людей щодо організацій, проблем, послуг» (Liu, 2010: 1; Liu, 2012: 7), а також ототожнює його із поняттям дослідження думок (opinion mining). Т. Насукава був першим, хто у 2003 році вжив термін настроєвий аналіз та позначив його як «завдання з'ясувати почуття авторів, виражені в позитивних чи негативних коментарях, запитаннях і проханнях, шляхом аналізу великої кількості документів» (Nasukawa, 2003: 2; Kanayama, 2004: 494). Аналіз тональності сягає своїм корінням ще ХХ століття й був застосований на початку 1940-х років для дослідження

громадської думки (Lei, 2021), проте його розквіт припадає на ХХ століття й, за підрахунками М. Ментуле, здійсненими в 2018 році, 99% із понад 5000 опублікованих результатів досліджень датовані після 2005 року (Mäntylä, 2018).

Настроєвий аналіз виражають зазвичай у повністю полярних термінах: позитивних (сприятливих) та негативних (несприятливих) (Lei, 2021: 1). Н. Дарчук згадує про третій – проміжний показник, що визначає нейтральне ставлення автора (Дарчук, 2019: 10). Отож настроєвий аналіз передбачає ідентифікацію вираження почуттів, полярності та сили висловлювань, а також ставлення до суб'єкта (Nasukawa, 2003: 2).

Аналіз тональності – конструкція з багатьма прибудовами та надбудовами, адже вимагає вирішення різноманітних задач опрацювання природної мови, а саме аналізу мікротексту, вияв сарказму, вилучення концепцій, виокремлення аспектів розпізнавання анафор, тлумачення суб'єктивності, розпізнавання іменованих сутностей та особистості, а також профілювання користувачів (Cambria, 2017). Настроєвий аналіз застосовує два основні типи методів: «на основі лексики, які часто називають неконтрольованим машинним навчанням (unsupervised machine-learning), і контрольоване машинне навчання (supervised machine-learning), а також існує можливість поєднання цих двох шляхів, тобто гібридний метод» (Lei, 2021: 14).

Суттєвим у тлумаченні тональності є жанр текстів. По-перше, не у всіх текстах зустрічається феномен тональності. По-друге, «у текстах можуть бути множинними не лише автори та їхні оцінки, а й об'єкти, а також ситуації, що значно ускладнює процес автоматичного аналізу тональності, як і багатозначність оцінної лексики, яка значною мірою залежить від контексту» (Дарчук, 2019: 10). Н. Дарчук пропонує звертати увагу на чотири моменти за аналізу настроїв: «1) суб'єкт тональності – автор повідомлення (може надавати характеристику об'єктам висловлювання цілісно в той чи інший спосіб і зводиться до позитивного/негативного представлення тональності); 2) об'єкт тональності – про що йде мова (зміст написаного, що визначається, за ключовими словами); 3) аспект тональності – характеристика об'єкта; 4) оцінка тональності – настроєвий аналіз (ставлення автора до описуваного предмета, аспекту, властивостей)» (Дарчук, 2019: 10).

Основне завдання цього дослідження – оцінити роль настроєвого аналізу у визначенні якості перекладу та відповідності використаної лексики, адже автор Дж. К. Роулінг та перекладач В. Морозов пишуть для різних мовних аудиторій, пристосовуючись до їхніх реалій. Для цієї оцінки використано прийоми, які може запропонувати лексикон LIWC-22.

LIWC (Linguistic Inquiry and Word Count) – програма для аналізу тексту, що вперше з'явилася як дослідницький проєкт у 1992 році й стала рішенням для невдач ранніх комп'ютерних методів (Pennebaker, 2022). Основне завдання застосування – вивчати зразки усного та письмового мовлення. Згодом після першої версії вийшли друга (LIWC2001), третя (LIWC2007), четверта (2015), а тепер і п'ята (LIWC-22) версії (Pennebaker et al., 2022). Кожного разу оригінальну програму оновлювали, що впливало на складніший програмний дизайн та кількість одиниць у словниках, розширювало діапазон можливостей (з 2015 р. додано й український словник, сформований С. Засекіним). До прикладу, найновіший внутрішній словник містить понад 12 000 слів, словосполук, фраз й вибраних емотиконів, коли третя версія містила лише 4500 одиниць (Pennebaker et al., 2022; Pennebaker, 2007). Остання версія LIWC-22 суттєво змінила не лише словник, а й програмні опції, щоб відобразити нові напрямки в аналізі текстів. Як і попередні версії, програму призначено для швидкого та ефективного аналізу окремих або декількох мовних файлів. Водночас застосунок надає можливість користувачеві досліджувати вживання слів різними способами, адже вихідні дані LIWC містять ім'я файлу, загальну кількість і відсоток слів, що були захоплені в тексті для кожного мовного виміру (вибір яких здійснюється самостійно) (Pennebaker et al., 2022).

Проаналізовані оповідання Дж. К. Роулінг «Чаклун і стрибучий казанок», «Фонтан фортуни» та «Волохате серце мага» досить короткі за обсягом, тому їх можна досліджувати повністю

або сегментувати для вияву місць перекладацьких трансформацій. Цікавим є те, що українські тексти коротші від англійських приблизно на 700 слів кожен. Запропоновані оповідання представляють жанр казки й рясніють забарвленою лексикою, що робить їх хорошим матеріалом для отримання показників різних категорій та вивчення тональності загалом. Варто згадати й те, що кожна казка має своєрідний додаток – рубрику «Албус Дамблдор коментує казку...», що також було проаналізовано з метою визначити, як відрізняються сентименти самої казки та її трактування та як це впливає на загальні показники.

Для оцінки тональності завантажено файли з казками та їхніми обговореннями. Це зроблено окремо для англійських та українських варіантів, адже здійснення аналізу вимагає обирати словники, що збігаються з мовою матеріалу. Як вже було згадано раніше, український словник утворено у 2015 році, а тому він не є таким обширним і оновленим, як англійська версія LIWC-22. Отож показники для відстеження відібрано **tone\_pos**, **tone\_neg** для англійської мови та **posemo**, **negemo** для української, які відображають не стільки емоції, як настрої твору. Вони обидва містять слова з позитивними (*well, love, новий, хороший*) та негативними (*too much, hate, поганий, помилковий*) емоціями. Відповідно до цих двох категорій наведено графіки (рис. 1, 2), де відображено частки позитивних чи негативних лексем твору, що в оповіданнях англійською мовою не перевищують 4.00% та 5.00% українською мовою (застосунок використовує відсотковий спосіб відображення показників). Номери 1, 2 та 3 відповідають

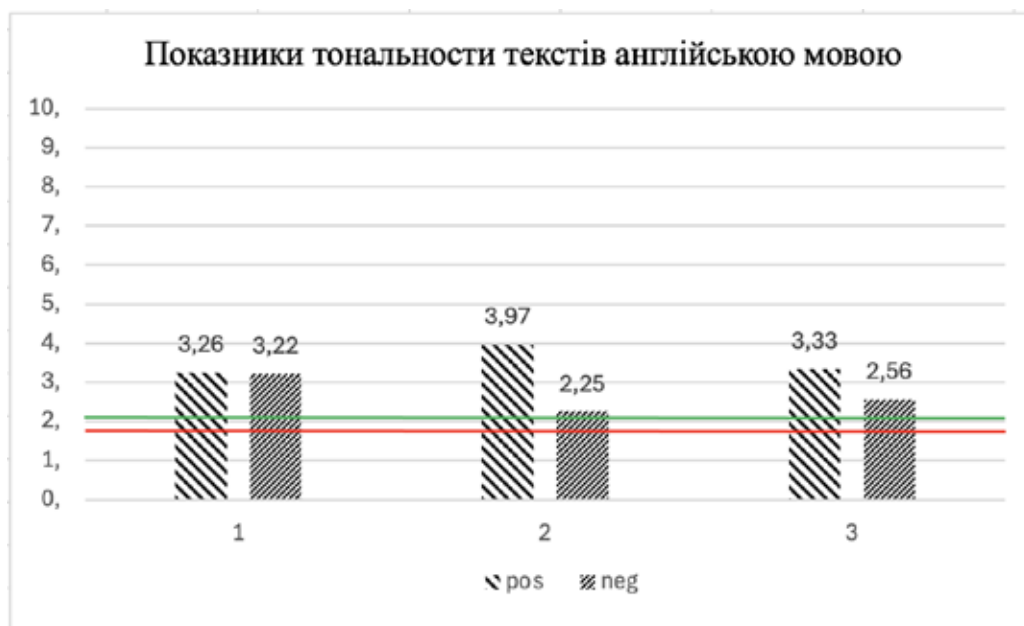
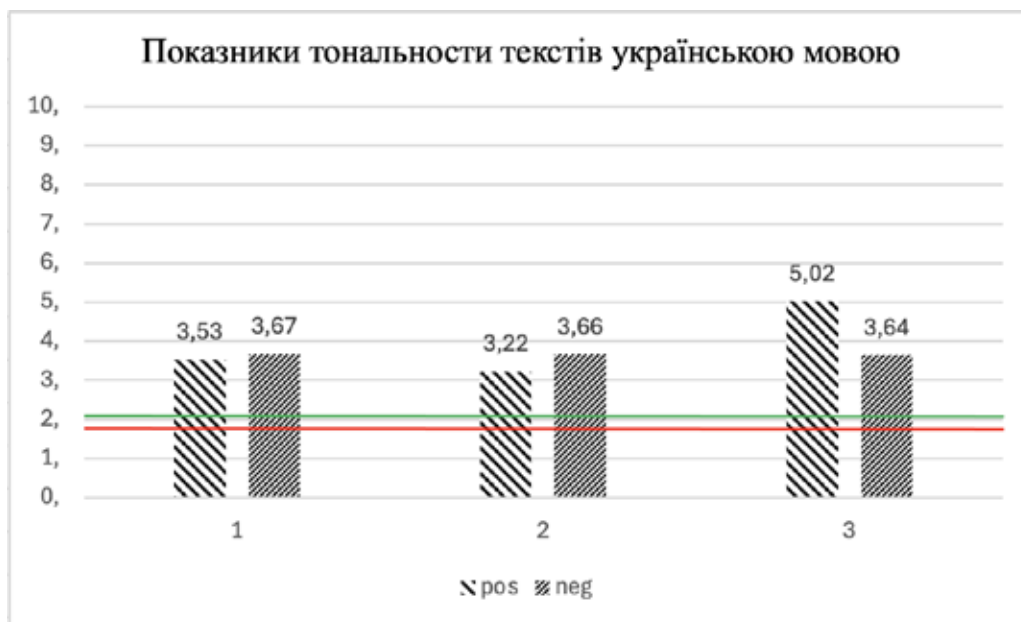


Рис. 1. Показники тональності текстів англійською мовою



**Рис. 2. Показники тональності текстів українською мовою**

послідовності розташування казок у збірці: 1 – «Чаклун і стрибучий казанок», 2 – «Фонтан фортуни» та 3 – «Волохате серце мага». Варто зауважити, що тональність текстів різниться: переклад першої казки (posemo – 3.53%, negemo – 3.67%) практично відповідає оригіналові (tone\_pos – 3.26%, tone\_neg – 3.22%) зі зсувом у кілька десятих відсотка; щодо двох інших, то у них є певні розбіжності. У казці «Фонтан фортуни» мовою оригіналу позитивна тональність переважає (tone\_pos – 3.97%, tone\_neg – 2.25%), коли у мові перекладу навпаки (posemo – 3.22%, negemo – 3.66%). В оповіданні «Волохате серце мага» кількісні показники перекладу (posemo – 5.02%, negemo – 3.64%) перевищують показники оригіналу (tone\_pos – 3.33%, tone\_neg – 2.56%), але позитивна тональність переважає у двох варіантах.

Для порівняння у документації LIWC розташовано таблиці зі середніми значеннями для всіх категорій текстів, які можна надавати обчисленням у програмі (романів, твітів, блогів, імейлів, оповідань та ін.), що дає змогу встановити, наскільки висока чи низька тональність досліджуваного матеріалу. Зелена лінія на графіках – середнє число позитивних показників, яка дорівнює 2.18% та червона – негативних – 1.75%. Порівнюючи показники досліджуваних текстів оригіналу та їхнього перекладу із середніми значеннями, представленими LIWC-22 на базі Test Kitchen Corpus для жанру казки, спостерігаємо, що позитивна, а також негативна тональність «Казок барда Бідла» дещо вища.

Варто зауважити, що лайливі слова як окрему підпорядковану групу не вміщено до складу

одиниць, що формують категорії tone\_positive/tone\_negative чи posemo/negemo. Їх виділено в окрему категорію для зручності підрахунку елементів специфічної лексики. Очевидним є те, що у казках вони не траплятимуться або ж будуть зафіксовані зрідка, про що й свідчать показники: у цьому випадку вони присутні в оригіналі 0.00, 0.08 та 0.05 % у першому, другому та третьому оповіданні відповідно та 0.14, 0.11 та 0.00 % у перекладі. Їхня абсолютна частота близька до нуля та становить приблизно 1–2 слова на текст.

**Висновки.** Отож важливу роль у зіставних дослідженнях тональності оригінальних та перекладних текстів відіграють кількісні методи та алгоритми машинного навчання, що є напрацюванням комп'ютерної лінгвістики і постали як відповідь на потребу виявлення й окреслення категорій сентиментів. Програмне забезпечення LIWC, зокрема його остання версія LIWC-22, набуло своєї популярності у визначенні тональності різножанрових текстів. Його вважають одним з найточніших лексиконів, призначених для студій над зразками емоційних, когнітивних, структурних та технологічних текстових компонентів. Дослідження оригінальних оповідань Дж. К. Роулінг та їхніх перекладів В. Морозова виявило особливості передачі емоційного забарвлення згаданих текстів. Показники негативної та позитивної тональності «Казок Барда Бідла» дещо вищі, ніж середні показники тональностей жанру, представлені програмою LIWC-22. Кількісні показники позитивної тональності досліджених оповідань

мовою оригіналу переважають показники негативної; у мові перекладу домінування позитивного настрою виявлено лише у третьому оповіданні. Крім того, показники тональностей перекладів українською мовою є загалом дещо вищі, ніж такі у мові оригіналу, проте виявлена незначна різниця не свідчить про порушення тонального балансу між оригіналом і перекладом, а, навпаки, засвідчує успішність його збереження. На нашу думку, деякі кількісні коливання у показниках тональностей зумовлені впливом мовних і культурних факторів на процес перекладу, а також намаганням перекладача відтворити не лише змістову частину твору, але й спроектувати відповідні емоційно-експресивні домінанти в цільовій мові.

Запропонована у статті методика вивчення перекладів художньої прози відкриває перспективу для дослідження в такому руслі інших текстів.

#### ЛІТЕРАТУРА

1. Дарчук Н. П. Лінгвістичні засади автоматичного настрою-аналізу українськомовного тексту. *Science and Education a New Dimension Philology*. Budapest, 2019. № 55. С. 10–13.
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. A Practical Guide to Sentiment Analysis / ed. by E. Cambria, S. Bandyopadhyay, A. Feraco. Stirling: Springer International Publishing, 2017. Vol. 5. 196 p.
5. Iglesias C., Moreno A. Sentiment analysis for social media. *Applied Sciences*. Basel, 2019. № 9. P. 1–4.
6. Kanayama H., Nasukawa T., Watanabe H. Deeper Sentiment Analysis Using Machine Translation Technology. *COLING: in proc. of the 20th International Conference, Geneva, Aug 23–Aug 27 2004 Geneva, 2004*. P. 494–500.
7. Lei L., Liu D. Conducting Sentiment Analysis. Cambridge: Cambridge University Press, 2021. 106 p.
8. Liu B. Sentiment Analysis and Opinion Mining. Williston-London: Morgan&Claypool Publishers, 2012. 168 p.
9. Liu B. Sentiment Analysis and Subjectivity. *Handbook of Natural Language Processing*. 2nd ed. / ed. by N. Indurkha., F. J. Damerau. Boca Raton, 2010. P. 627–667.
10. Mäntylä M. V., Graziotin D., Kuutila M. The Evolution of Sentiment Analysis – A Review of Research Topics, Venues, and Top Cited Papers. *Computer Science Review*. 2018. Vol. 27. P. 16–32.
11. Nasukawa T., Yi J. Sentiment analysis: Capturing favorability using natural language processing. *K-CAP03: in proc. of the Second International Conferences, 23 October 2003*. Sanibel Island FL, 2003. P. 70–77.
12. Pennebaker J. W., Francis M. E., Booth R. J. Linguistic Inquiry and Word Count (LIWC). Operator's Manual. Austin, TX: University of Texas at Austin, 2007. 11 p.
13. The Development and Psychometric Properties of LIWC-22 / J. W. Pennebaker et al. Austin, TX: University of Texas at Austin, 2022. 49 p.