

*Барановська Л. М.,**студентка IV курсу кафедри прикладної лінгвістики
Національного університету «Львівська політехніка»**Альбота С. М.,**кандидат філологічних наук, доцент,
доцент кафедри прикладної лінгвістики
Національного університету «Львівська політехніка»*

ТРАГЕДІЯ ЛЮДСЬКОЇ ДОЛІ: КОРПУСОБАЗОВАНИЙ ПІДХІД У ДОСЛІДЖЕННІ ІДІОЛЕКТУ БОРИСА ГРІНЧЕНКА НА МАТЕРІАЛІ ОПОВІДАННЯ «КАТОРЖНА»

Анотація. Проблема дослідження авторської мовної картини світу, властивостей лексичних та стилістичних засобів та способів вербалізації концептуальної сфери письменника вже протягом декількох століть посідає вагомe місце в мовознавчих дослідженнях науковців усього світу. Подібна проблематика привертає увагу лінгвістів не лише в процесі дослідження мови як загального явища, а також як індивідуального мовлення, яке ґрунтується на його своєрідних ознаках. Через художній стиль письменника передається його індивідуальність та власна мовна картина світу. З метою узагальнення всіх мовних особливостей письменника дослідники використовують таке поняття як «ідіолект», його визначення та підходи до вивчення є предметом дискусій протягом багатьох століть як у зарубіжному, так і українському мовознавстві. У запропонованому дослідженні наша увага зосереджується на лінгвостатистичному підході до визначення ідіолекту письменника. Для виконання дослідження було обрано соціально-психологічне оповідання українського письменника Бориса Грінченка, написане у 1888 році. Тематика твору торкається проблеми взаємостосунків між дітьми та батьками, важливості моральних цінностей у суспільстві, жорстокої поведінки людей, а також проблеми нещасного дитинства з подальшим впливом на все життя особистості. Для дослідження твору тексту було попередньо вичитано, для видалення зайвих графічних знаків та апострофів, опрацьовано за допомогою програмного забезпечення «AntConc». З метою опрацювання статистичних показників було використано функціонал програми «Microsoft Excel». З опрацьованого тексту було укладено список вжитих у ньому слівформ. Розрахунки статистичних показників було проведено з використанням власних математичних формул. Під час дослідження ідіолекту письменника найбільш ефективним методом є не лише аналіз лексичних та граматичних одиниць, а також безпосередній аналіз постійно відтворюваних одиниць – колокацій. Корпус текстів дозволяє провести статистичний аналіз сполучуваності лексичних одиниць, які дають змогу об'єктивно визначити якісні особливості ідіолекту письменника. У програмі «AntConc» було створено «швидкий корпус» для підрахунку найбільш вживаних лексичних одиниць, а також пошуку колокацій.

Ключові слова: частина мови, ідіолект, семантика, корпусна лінгвістика, статистичний показник.

Постановка проблеми. Мовознавці розглядають явище авторського ідіолекту як унікальний та своєрідний феномен,

що виникає через вплив соціальних, психологічних, культурних та духовних особливостей певного народу, його традицій, національної специфіки, представником яких є митець. За допомогою певних експресивних засобів письменник зображує власну картину світу, яка відтворює в собі усі національні особливості. Щоб ідентифікувати особливості ідіолекту письменника необхідно з'ясувати, які мовні засоби у його творчості є переважаючими, яким чином вони формують мовну картину світу цього автора, висвітлюють його світогляд та спосіб мислення. Незважаючи на те, що поняття «ідіолект» давно відоме в лінгвістиці, воно все ще залишається предметом дискусій. Дослідження ідіолекту письменника допомагає краще оцінити та зрозуміти стиль автора, манеру його оповіді, оцінити вживання певних лексичних та граматичних засобів методами квантитативного аналізу.

Аналіз останніх досліджень та публікацій. Сучасна наукова розвідка надає великого значення проблемі дослідження мовлення конкретної особи або навіть групи осіб. Авторський стиль мови найбільш виразно та чітко відображає ідіолект письменника, тому щоб його дослідити науковці аналізують власне текст.

Збільшення зацікавленості до питання семантики різномовних мовних одиниць, їх функціонування та вплив на читача розпочалося ще з середини 20-х років минулого століття. Багато наукових праць у цьому напрямку опираються на дослідження Гумбольдта [1, 2] та О. О. Потебні [3].

Проблему визначення індивідуальної природи особистості досліджували наступні науковці: І. Бодуен де Куртене [4], Е. Сепір [5], Ф. де Сосюр [6], К. Фосслер [7] та ін. Серед українських дослідників, які вивчали мову особистості великого значення набули праці С. Єрмоленко [8], Ф. Бацевича [9], І. Фаріон [10] та К. Голобородько [11] та ін.

Мета статті полягає у визначенні ідіолекту автора та його особливостей на основі корпусобазованого підходу у романі українського письменника Бориса Грінченка «Каторжна».

Рішення до запропонованої мети можуть бути досягнені шляхом виконання наступних завдань:

- визначити наукові підходи до вивчення поняття «ідіолект»;
- виокремити основні лексичні та граматичні категорії, які розкриватимуть особливості авторського ідіолекту;

– встановити типи для стилю автора граматичні та лексичні ознаки;

– висвітлити труднощі, які виникають у процесі дослідження ідіолекту певного письменника, щоб удосконалити методику і полегшити майбутню наукову розвідку у цьому напрямку.

Об'єктом поданого дослідження є ідіолект Бориса Грінченка в романі «Каторжна».

Предметом праці виступають виступають лінгвістичні особливості ідіолекту Бориса Грінченка, визначені за допомогою корпусобазованого підходу.

Виклад основного матеріалу. Поняття «ідіолект» вже давно відоме мовознавчій науці, проте досі залишається предметом дискусій. Вивчення ідіолекту письменника дозволяє вивчити та оцінити стиль автора, вживання певних лексичних та граматичних категорій, манеру його оповіді за допомогою методів квантитативного аналізу. Американський лінгвіст Б. Блох у 1948 запропонував термін «ідіолект». Під цим терміном мовознавець розумів збірку можливих висловлювань використання мови у певний час з метою взаємодії з іншою людиною [12]. Сучасне мовознавство виділяє соціолінгвістичний, стилістичний, лінгвокультурологічний та лінгвокогнітивний підходи до розуміння поняття «ідіолект».

Розвиток комп'ютерних технологій та доступного програмного забезпечення надає статистичним дослідженням у рамках корпусної лінгвістики можливості для стрімкого розвитку.

Досліджуване оповідання містить 3598 слововживань. Кількість різних словоформ становить 1330 разів. Кількість унікальних слів, тобто таких, які вживаються у тексті лише один раз, становить 918 слів, у відсотковому відношенні це становить 25,5% від загальної кількості вжитих у тексті слів. Важливим кроком для дослідження ідіолекту автора є визначення кількості високочастотної лексики (лексичні одиниці, які використовуються у тексті більше десяти разів); індексу винятковості тексту, яких характеризує варіативність лексики у тексті; індекс концентрації лексики у тексті, який визначає частку тексту, яку займають слова з високою частотою.

Для обчислення показників було використано середовище «Microsoft Excel», і отримано наступні результати:



Рис. 1. Діаграма розподілу частки вживання частин мови у тексті

Таблиця 1

Статистичні показники тексту

Кількість високочастотної лексики (V10t)	40
Індекс винятковості тексту (I _{ext})	0,37
Індекс концентрації у тексті (I _{ext})	0,011

З отриманих показників можна зробити висновок про середнє різноманіття лексики в романі Бориса Грінченка «Каторжна», оскільки у тексті частка слів, що вживаються лише один раз не є великою на відповідний обсяг тексту.

Наступним кроком був підрахунок слів за частинами мови, отримані результати становлять наступне:

Таблиця 2

Статистичні показники розподілу вживання за частинами мови

Частина мови	Кількість вживань у тексті
Іменники	646
Прикметники	153
Дієслова	750
Займенники	662
Прислівники	284
Сполучники	499
Прийменники	276
Числівники	42
Частки	275
Вигуки	11
Загальна кількість слів	3598

У процесі опрацювання тексту програмою AntConc було визначено, що найчастіше вживаються наступні слова: сполучник «і» – 136 разів, частка «не» – 126 разів, займенник «вона» – 117 разів. Відсоток їх вживання до загальної кількості слів у тексті 3,77%, 3,5% та 3,25% відповідно.

Найбільш вживаною частиною мови було дієслово, яке налічувало 750 слововживань, кількість займенників становила 662 слововживання, кількість вживання іменників – 646.

Необхідно зазначити, що серед дієслів високу частотність мають словоформи слова «плакати», серед іменників – «мачуха», «каторжна» та «журба».

На основі розрахунку частотності частин мови можна визначити наступні показники:

- індекс епітетизації (I_{nat}) – вказує, скільки іменників у тексті припадає на один прикметник;
- індекс дієслівних означень (I_{vat}) – вказує, скільки прислівників припадає на одне дієслово;
- ступінь номінальності – визначає кількість іменників, що припадають на одне дієслово.

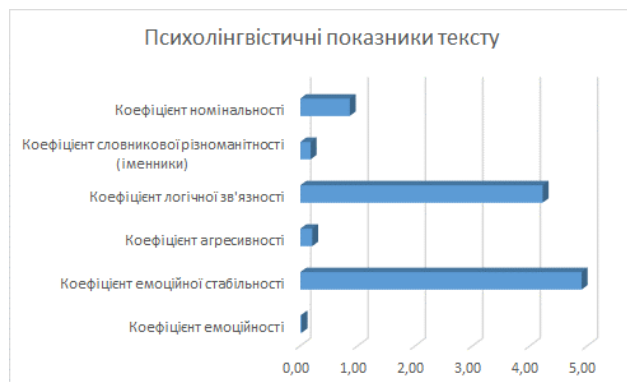


Рис. 2. Психолінгвістичні показники тексту

Таблиця 3

Показники реляційності між частинами мови

	Значення показника
Індекс епітетизації (I_{ep})	4,2
Індекс дієслівних означень (I_{od})	0,37
Ступінь номінальності	0,86

У поданому оповіданні на один прикметник припадає близько 4 іменників, індекс вживання дієслівних означень малий, тобто вірогідність вживання дієслів з прислівниками мала, дещо більший показник, що позначає ступінь номінальності – на одне дієслово може припасти один іменник, проте такі сполуки не завжди присутні.

Для поданого дослідження важливою складовою було визначення психолінгвістичних показників тексту [13]. Найвищі коефіцієнти становлять показники емоційної стабільності та логічної зв'язності, що є результатом вживання великої кількості іменників та дієслів у тексті, найнижчим є коефіцієнт емоційності (0, 02%).

Додатково було визначено одиниці із негативним семантичним забарвленням [14].



Рис. 3. Розподіл вживань слів із негативною семантикою

Кількість таких слів становить 222 одиниці, це становить 6,17% від усієї кількості тексту та свідчить про відносно вагомому частку слів на позначення негативної семантики.

Також за допомогою програмного забезпечення «AntConc» також було проведено пошук N-грам, тобто пошук частотних списків послідовностей для слова з негативним забарвленням «каторжна» та отримано наступні результати:

	Типе	Rank	Freq	Range
1	каторжна а вона тільки глянула	1	1	1
2	каторжна а до батька яка	1	1	1
3	каторжна а з мачушиними дітьми	1	1	1
4	каторжна але каторжна обхопила руками	1	1	1
5	каторжна визвірювався на неї батько	1	1	1
6	каторжна вона кинулась до мачухи	1	1	1
7	каторжна від усіх ляна від	1	1	1
8	каторжна докія та й було	1	1	1
9	каторжна забігла тоді вставала вона	1	1	1
10	каторжна обхопила руками куці і	1	1	1
11	каторжна стала усім вона каторжна	1	1	1
12	каторжна та й годі мабуть	1	1	1
13	каторжна так до мене в	1	1	1
14	каторжна так її всі звали	1	1	1
15	каторжна ти кричала мачуха геть	1	1	1
16	каторжна усім на світі усі	1	1	1
17	каторжна що її нічим не	1	1	1
18	каторжна я тобі дам як	1	1	1

Можна простежити, що за значенням виокремлені грами мають таке ж негативне забарвлення, як і власне опрацьоване слово.

Висновки та перспективи подальших пошуків. На основі проведеного дослідження було з'ясовано, що для творчості українського письменника Бориса Грінченка характерне середнє різноманіття лексики, оскільки кількість слів, використаних у тексті один раз, становить 918 слововживань та кількість слів, що використовуються автором більше 10 разів становить лише 40 слововживань.

Для оповідання «Каторжна» притаманне використання великої кількості таких самостійних частин мови, як дієслова, іменники та займенники. За результатами психолінгвістичного дослідження було визначено, що найвищими показниками у творі є показники емоційної стабільності та логічної зв'язності.

Також було цікаво визначити, що зазвичай, щоб вказати на головну героїню оповідання, інші герої вживали не її ім'я, а займенник «вона». Оповідання «Каторжна» є яскравим прикладом зображення нелегкої людської долі, відповідно у цьому творі було використано велику кількість лексичних засобів із негативною семантикою.

У процесі дослідження ідіолекту письменника дослідник може стикнутись із низкою труднощів, які можуть уповільнити цей процес. Наприклад, недостатнє функціональне забезпечення, щоб полегшити статистичні розрахунки (особливо в випадках, коли текст має великий об'єм).

Перспективою цього дослідження виступає подальший, глибший аналіз ідіолекту та ідіостилу Бориса Грінченка, структурування та поглиблення лінгвостатистичного підходу до їх вивчення. Для глибшого вивчення ідіолекту автора також було б доцільно також виділити та дослідити концептуальну систему письменника та дослідити, які стилістичні засоби автор використовує у своїх творах.

Література:

1. Гумбольдт В. фон. Избранные труды по языкознанию. Москва: Прогресс, 1984. 397 с.
2. Гумбольдт В. Язык и философия культуры. Москва: Прогресс, 1985. 456 с.
3. Потебня О. Мова, національність, денаціоналізація. Видавництво: Українська Вільна Академія Наук у США, 1992, 155 с.
4. Бодуэн де Куртене И.А. Избранные труды по общему языкознанию. Ч. 1–2. Москва: Изд-во АН СССР, 1963. 388 с.
5. Сепир Э. Избранные труды по языкознанию и культурологии. Москва: Прогресс, 1993. с. 223–247.
6. Saussure F. de (1972). Cours de linguistique générale. Paris : Payot, 520 p. URL: <https://plato.stanford.edu/archives/sum2021/entries/idioclects/> (date of access: 26.05.2022).
7. Фосслер К. Эстетический идеализм. Избранные работы по языкознанию. Москва: Издательство ЛКИ, 2007. 144 с.
8. Єрмоленко С.Я. Нариси з української словесності (стилістика та культура мови). Київ: Довіра, 1999. 431 с.
9. Бацевич Ф.С. Когнітивне і лінгвальне в процесах вербалізації (на матеріалі аномальних висловлень) // Мовознавство. 1997. № 6. С. 30–37.
10. Фаріон І.Д. Ідіостиль Івана Пулюя крізь призму динаміки мовної норми (за листами мислителя). Лінгвостилістичні студії. Східноєвропейський університет ім. Лесі Українки. Луцьк, 2017. Вип. 7. с. 176–183.
11. Голобородько К. Ю. Ідіостиль Олександра Олеса: лінгвокогнітивна інтерпретація: [монографія]. Харків: Харківське історико-філологічне товариство, 2010. 527 с.

12. Bloch B. A. (1948). Set of postulates for phonemic analysis. *Language* XXIV, pp. 3–46.
13. S. Albota, «Linguistic and psychological features of the Reddit news post», *International Scientific and Technical Conference on Computer Sciences and Information Technologies*, № 1, pp. 295-299, 2020.
14. S. Albota, «Linguistically manipulative, disputable, semantic nature of the community Reddit feed post», *CEUR Workshop Proceedings. Proceedings of the 5th International conference on computational linguistics and intelligent systems (COLINS 2021)*, № 1, pp. 769-783, April 22-23, 2021.

Baranovska L., Albota S. The tragedy of human destiny: a corpus-based approach in the study of Borys Hrinchenko's idiolect based on material of the narrative «Katorzhna»

Summary. The problem of research of the author's linguistic picture of the world, properties of lexical and stylistic means and ways of verbalization of the writer's conceptual sphere for several centuries occupy an important place in linguistic research of scholars around the world. Such issues attract the attention of linguists not only in the process of studying language as a general phenomenon, but also as an individual speech, which is based on its unique features. The writer's artistic style conveys his individuality and his own linguistic picture of the world. In order to generalize all the linguistic features of the writer, researchers use such a concept as «idiolect», its definition and approaches to study have been the sub-

ject of debate for many centuries in both foreign and Ukrainian linguistics. In the proposed study, our attention is focused on the linguistic-statistical approach to determining the idiolect of the writer. The socio-psychological story of the Ukrainian writer Borys Hrinchenko was written in 1888, and chosen for this study. The theme of the work concerns the problem of the relationship between children and parents, the importance of moral values in society, the cruel behavior of people, as well as the problem of unhappy childhood with subsequent impact on the entire life of the individual. To study the work, the text was pre-read to remove unnecessary graphic characters and apostrophes, and processed using «AntConc» software. In order to process statistical indicators, the functionality of «Microsoft Excel» was used. A list of word forms used in it was compiled from the processed text. Calculations of statistical indicators were performed using own mathematical formulas. In the study of the writer's idiolect, the most effective method is not only the analysis of lexical and grammatical units, but also the direct analysis of constantly reproducing units – collocations. The body of texts allows for statistical analysis of the compatibility of lexical units, which allow to objectively determine the qualitative features of the writer's idiolect. «AntConc» has created a «quick corpus» for counting the most commonly used lexical items, as well as finding collocations.

Key words: part of speech, idiolect, semantics, corpus linguistics, statistical indicator.