

ФІЛОЛОГІЧНІ НАУКИ

DOI: <https://doi.org/10.32839/2304-5809/2019-11-75-198>

УДК 811.161.2'272'233'322.5:82-31Іваничук7.08

Бахурська Л.З., Лотоцька Н.Я.

Національний університет «Львівська політехніка»

АКТУАЛЬНІСТЬ ТА ВАЖЛИВІСТЬ ЛІНГВОСТАТИСТИЧНОГО ДОСЛІДЖЕННЯ ПИСЬМЕННИКА РОМАНА ІВАНИЧУКА У РОМАНІ «ОРДА»

Анотація. Стаття присвячена актуальності дослідження ідіостилу автора як способу володіння засобами мови, статистичного підрахунку вживаності слів у творі письменника Романа Іваничука «Орда». Дослідження індивідуального стилю автора дає змогу проаналізувати та визначити закономірності функціонування та розвитку мови в часі та просторі, а також оцінити внесок автора в систему художніх засобів. У статті досліджено використання статистичних даних у лінгвістичному профілі автора. Дані, які зібрані процедурами статистичної лінгвістики, використовуються для виявлення стилістичних особливостей письменника Романа Іваничука. У дослідженні використано структурне роззначення, яке допомагає виявити індекси винятковості та різноманітності автора, багатогранності його словника та готує текст до процесу ідентифікації текстової форми слова як представника тої лексеми, що зберігається у словнику.

Ключові слова: ідіостиль, лінгвостатистичні дослідження, вживаність слів, статистика, структурне роззначення.

Bakhurska Liliia, Lototska Natalia

Lviv Polytechnic National University

THE RELEVANCE AND IMPORTANCE OF LINGUISTIC AND STATISTICAL STUDY OF THE WRITER ROMAN IVANYCHUK IN THE NOVEL "ORDA"

Summary. The article is devoted to the topicality of the study of the author's idiosyle as a way of mastering the means of speech, the statistical calculation of the use of words in the work of the writer Roman Ivanychuk «Orda». The study of the author's individual style makes it possible to analyze and determine the patterns of functioning and development of language in time and space, as well as to evaluate the author's contribution to the system of artistic means. The study of the individual style of the author includes research and analysis of the choice of language and stylistic means of the writer, the main idea and its manifestation. It is determined that the individual style of the author is a complex but structurally integrated system of aesthetically creative selection and understanding of different linguistic elements inherent in the text activity of author Roman Ivanychuk. The style of the author depends mainly on the selection of specific language means in accordance with the basic idea behind his creativity and personal linguistic and stylistic preferences. The article explores the usage of statistics in the author's linguistic profile. The data collected through statistical linguistics are used to identify the stylistic features of writer Roman Ivanychuk. The study uses structural marking to help identify indexes of the author's exclusivity and diversity, the versatility of his vocabulary, and prepares the text for the process of identifying the textual form of a word as representative of that token stored in the dictionary. Structural text marking provides information about the structure of text, allowing us to separate words, borders of sentences and paragraphs. The statistical survey of large sets of texts enables one to delve deeper into the laws of the functioning of linguistic units in speech. Statistical methods are one of the methods that are best devoid of subjectivity; they provide a basis on which to draw linguistic conclusions. Statistics allowed us to generalize complex numerical linguistic data to draw accurate conclusions, since intuitive comparisons cannot always be used to tell us anything significant about linguistic variability.

Keywords: idiosyle, linguistic statistics, word usage, style, statistics, structural marking.

Постановка проблеми. Індивідуальний стиль автора є невід'ємною частиною вивчення лінгвостилістики та когнітивної лінгвістики та є багаторівневим проявом особистості [8, с. 2]. Зацікавлення до індивідуальності автора та його впливу на мистецьку спадщину завжди було актуальним. Власне лінгвістичні дослідження, які почали вивчати це питання, набули своєї популярності в середині 20-х років ХХ ст. і досі поняття «ідіостилу автора» є одним з найбільш суперечливих.

Попри високу увагу до цього поняття, в сучасному науковому просторі досі не визначено спільних підходів до тлумачення поняття «ідіостиль» як складової лінгвістичного дослідження. Лінгвіст В. Виноградов стверджував, що загальні риси

та процеси розвитку мови найкраще виражені через ідіостиль. Індивідуальний стиль автора дослідник трактує, як складну, але структурно поєднану систему естетично творчого добору та розуміння різних мовних елементів притаманних текстовій діяльності певного автора [2, с. 135].

Ідіостиль асоціюється із системою мовних та стилістичних засобів, властивих художній ідентичності автора та із специфічними рисами мови автора, до яких можуть належати лексичні, семантичні, граматичні або стилістичні особливості текстів. Для того, що повністю зрозуміти ідіостиль письменника завдання вчених полягає не лише в тому, щоб проаналізувати мовні засоби, але й у вивченні біографії, життєвого досвіду та психологічної особистості автора.

Статистика може бути використана не лише в математичних дослідженнях, а й власне бути пов'язаною з вивченням ідіостилю автора. На жаль, серед лінгвістичної спільноти статистичні методи часто уникають. Статистика дозволяє нам узагальнити складні числові лінгвістичні дані, щоб зробити з них висновки, адже інтуїтивні порівняння не завжди можуть бути використані для того, щоб сказати щось істотне про мовну варіативність [11, с. 6]. Статистичні методи є важливим пунктом при укладанні частотних словників, адже допомагають підібрати найуживанішу лексику та дослідити лексичний склад мови. Не менш корисними ці методи є при укладанні корпусу, адже один із статистичних методів – процес зняття омонімії, полегшує автоматичну обробку тексту. Кількісні статистичні методи застосовують у перекладознавстві в процесі зіставлення та порівнянні двох мов при перекладі.

Статистичне обстеження великих сукупностей текстів зумовлюється необхідністю глибше проникнути в закони функціонування мовних одиниць у мовленні. У своїй роботі дослідники-лінгвісти застосовують статистичні методи для виміру зв'язків між мовними елементами, встановлення подібності чи відмінності між кількісними сукупностями, для розподілу мовних одиниць у мові, та навіть для встановлення тенденцій розвитку мови [6, с. 1–2]. Застосування статистичних у лінгвістиці забезпечує високу надійність результатів, адже вона базована на точних кількісних даних, які можна перевірити [3, с. 27–28]. Дані, зібрані процедурами статистичної лінгвістики, використовують для виявлення стилістичних особливостей авторів, з'ясування джерел текстів, розшифрування історичних текстів та вирішення проблем стенографії, теорії комунікацій та інформатики.

Аналіз досліджень та публікацій. Проблеми «ідіостилю» присвячені праці В.А. Кухаренко, І.А. Тарасової, Н.С. Болотнової, Г.Я. Солганіка, Н.І. Головченко, В.В. Леденової, В.Г. Щукіна та інші [8, с. 2]. Незважаючи на високу зацікавленість цим поняттям, досі не визначено спільних підходів до тлумачення дефініції «ідіостиль» як складової лінгвістичного дослідження. До прикладу, український учений О. Пономарів вказує, що загальні риси та процеси розвитку мови краще виражаються через ідіостиль [10, с. 332] та те, що в літературному стилі показано багатство національної мови [5, с. 15].

Суперечливим серед дослідників є не лише питання щодо дефініції поняття «ідіостилю», але й ототожнення та розмежування понять «ідіостиль» та «ідіолект». Р.Х. Робінсон інтерпретує ідіолект як мовленнєві навички особи [Цит. за: 1, с. 40], а М. Юсслер говорить про те, що ідіостиль це так звана своєрідна форма мовлення індивіда [9, с. 186].

Якщо говорити про лінгвостатистику у вузькому сенсі, то це вивчення деяких математичних проблем, пов'язаних з лінгвістичним матеріалом, а головним чином з видами статистичних розподілів мовних одиниць в тексті. Багато дослідників (В.І. Перебийніс, An. Woods та ін.) присвятили свої праці використанню статистичних методів у лінгвістиці. Дослідниця В. Перебийніс визначила безпосереднє співвідношення між статистикою та мовними дослідженнями, наголосивши на тому, що мова має кількісні характе-

ристики та є складною системою, яка підпорядковується статистичним законам [4, с. 20].

Одним з найпоширеніших методів лінгвостатистичного аналізу – це аналіз, заснований на так званому законі Зіпфа, що зводиться до рівняння виду: $F \times i = \text{const}$, де F – частота слова в частотному словнику, а i – ранг цього слова. Таким чином, закон Зіпфа виступає як модель опису розподілу слів по частоті, однак на кожен випадок вводяться коригувальні коефіцієнти певним чином позбавляють його так званої «пророчої сили». Проблема зв'язку частоти слова і його рангу зникається з проблемою оцінки лексичного багатства тексту або сукупності текстів.

Виділення не вирішених раніше частин загальної проблеми. У будь-якому дослідженні творчості письменника важливим етапом є розуміння власне його «індивідуального стилю». Дослідження «ідіостилю» дають змогу оцінити внесок автора в систему художніх засобів, проаналізувати та визначити закономірності функціонування та розвитку мови в часі та просторі. Використання статистики у лінгвістичних дослідженнях є сучасним підходом для опису, наприклад, лексичного рівня у художніх творах. Поняття індивідуального стилю автора не завжди вважається важливим етапом у дослідження його творчості та певних лінгвістичних даних пов'язаних із специфікою вживання мовних одиниць у творах автора. Статистичні методи допомагають визначити найуживанішу лексику, що є важливим етапом при укладанні частотних словників у лексикографії. На превеликий жаль, серед лінгвістичної спільноти статистичні методи в основному ігноруються або уникаються. Причинами цього може бути те, що статистику відносять лише до царини математичних наук, а також існує відчуття, що ці способи можуть порушити «магію» літературного тексту.

Мета статті. Головною метою цієї роботи полягає дослідження творчості письменника Романа Іваничука у романі «Орда» з лінгвістичної точки зору. Актуалізувати дослідження літературних творів на статистичному рівні та показати важливість використання статистики на прикладі результатів отриманих з структурного роззначення роману «Орда».

Виклад основного матеріалу. Для проведення лінгвостатистичного дослідження роману «Орда» Р. Іваничука було виконано кілька етапів підготовки для отримання необхідних результатів.

Першим етапом є структурне роззначення речень твору Романа Іваничука «Орда». Варто звернути увагу на те, що структурне роззначення тексту представляє інформацію про структуру тексту, дає змогу відокремити слова, межі речень та абзаців. Попередньо проведено структурне роззначення речень тексту Романа Іваничука, що уможливить визначення частотності, довжину речень, розподіл кількості букв у словах та реченнях твору, числові характеристики вибірки, визначення абсолютних та відносних частоти. Отримані результати структурного роззначення допоможуть визначити особливості ідіолекту Романа Іваничука у творі «Орда» та виявити індекси винятковості та різноманітності автора, багатогранності його словника.

За допомогою структурного роззначення, ми отримали загальні дані про роман, які допоможуть у подальших порівняннях та висновках.

Таблиця 1
Довжина тексту у різних одиницях

Одиниця	Кількість
Буква	317576
Слово	61136
Речення	3614
Абзац	1331
Розділ	19
Частина	1

Обробка статистичного роззначення твору дала нам результат того, що у одній частині роману «Орда» Р. Іванчука є 19 розділів, які загалом містять 1331 абзаців, які, у свою чергу, складають 3614 речень. Усі ці речення загалом нараховують 61136 слів, які в свою чергу містять 317576 букв.

Структурне роззначення готує текст для подальшого проведення процесу лематизації, а якщо бути більш точними, то до процесу ідентифікації текстової форми слова як представника тої лексеми, що зберігається у словнику [7, с. 30]. Оскільки, однією із актуальних проблем, які зустрічаються при обробці художнього тексту є наявність морфологічної та лексичної омонімії, яка в подальшому ускладнює автоматичну обробку тексту, то при укладанні корпусу, наступним важливим кроком є процес зняття омонімії.

Внесені структурні позначки у тексті були використані для проведення подальшого лінгвостатистичного дослідження роману «Орда».

Результатами першого підрахунку є розподіл кількості букв у словах твору. Результати статистичного обстеження при *одиниці виміру – буква*. Розглядаємо розподіл абсолютних та відносних частот.

Таблиця 2
Розподіл кількості букв у словах твору

Кількість букв у словах	Розподіл абсолютних частот (у всьому тексті)	Розподіл відносних частот (у всьому тексті)
1	6597	10,79%
2	8081	13,22%
3	5071	8,29%
4	7430	12,15%
5	7484	12,24%
6	6663	10,90%
7	5913	9,67%
8	5139	8,41%
9	3456	5,65%
10	2348	3,84%
11	1460	2,39%
12	754	1,23%
13	386	0,63%
14	194	0,32%
15	80	0,13%
16	39	0,06%
17	17	0,03%
18	15	0,02%
19	6	0,01%
20	3	0,00%

Найчастотнішими у романі «Орда» є слова, кількість букв яких дорівнює 2 – 8081 (13,22%); 5 – 7484 (12,24%); 4 – 7430 (12,15%); 6 – 6663 (10,90%); 1 – 6597 (10,79%); 7 – 5913 (9,67%); 8 – 5139 (8,41%); 3 – 5071 (8,29%); 9 – 3456 (5,65%); 10 – 2348 (3,84%); 11 – 1460 (2,39%). Слова, кількість літер у яких становить 12 є найменш частотними у творі та їх довжина становить 754 (1,23%); 13 – 386 (0,63%); 14 – 194 (0,32%); 15 – 80 (0,13%); 16 – 39 (0,06%); 17 – 17 (0,03%). Найменш вживаними були слова довжиною 18, 19 та 20 фонем з відповідною кількістю літер – 15 (0,02%), 6 (0,01%) та 3 (0,00%).

Посилаючись на попередні дані, розраховуємо числові характеристики вибірки розподілу кількості букв у словах твору. Результатами дослідження є кількість значень, максимальне та мінімальне значення, розмах, середнє та середнє квадратичне значення, коефіцієнт варіації, стандартна та відносна похибка.

Таблиця 3
Числові характеристики вибірки

Значення	Весь текст
N	61136
max	20
min	1
R	19
\bar{X}	5,19
σ	2,98
v	0,5741
$S\bar{x}$	0,0121
ε	0,0046

Як бачимо ще з таблиці 1 кількість значень, тобто слів у романі (N) є 61136. За результатами дослідження ми бачимо, що максимальне значення кількості букв у словах (max) – 20, тоді як мінімальне значення (min) – 1; розмах (R) – 19; середнє значення кількості букв у словах (\bar{X}) – 5,19; середнє квадратичне відхилення (σ) – 2,98; коефіцієнт варіації (v) – 0,5741; стандартна похибка ($S\bar{x}$) – 0,0121; відносна похибка (ε) – 0,0046.

Наступним етапом дослідження є статистичне обстеження при *одиниці виміру – слова*. Нижче наведено кілька даних у таблиці 4 з розподілом кількості слів у реченнях твору.

Таблиця 4
Розподіл кількості слів у реченнях твору

Кількість слів у реченні	Розподіл абсолютних частот	Розподіл відносних частот
7	200	5,53%
4	199	5,51%
8	197	5,45%
5	195	5,40%
6	193	5,34%
11	138	3,82%
19 та 22	68	1,88%
33 та 35	19	0,53%
55, 57 та 70	5	0,14%
63, 68, 78, 80, 81, 83, 87, 89, 94, 97, 99, 100, 107, 108, 119, 120, 127, 136, 137, 138, 144, 156 та 164	1	0,03%

Висновки та пропозиції. Виконані дослідження демонструють, що за допомогою статистичних підрахунків було визначено особливості ідіолекту Романа Іваничука у творі «Орда». Проведене структурне роззначення твору дало змогу дослідити розподіл кількості букв у сло-

вах твору, довжину тексту у різних одиницях, розподіл кількості слів у реченнях твору. У наступних студіях необхідно розширювати кількість показників для подальшого дослідження індивідуальності автора Р. Іваничука у його текстах.

Список літератури:

1. Белл Р.Т. Социолінгвістика: цели, методы и проблемы. Москва : Международные отношения, 1980. 320 с.
2. Виноградов В.В. Стилистика. Теория поэтической речи. Поэтика. Москва : Наука, 1963. 255 с.
3. Перебийніс В.І. Що дає статистика мовознавцям? *Вісник Київського лінгвістичного університету : зб. наук. праць. Серія : Філологія*. Київ : Вид. центр 133 КНЛУ, 2003. Т. 6, № 2. С. 27–32.
4. Перебийніс В.С. Кількісні та якісні характеристики системи фонем сучасної української літературної мови. Київ : Наукова думка, 1970. 270 с.
5. Пономарів О.Д. Стилистика сучасної української мови. Київ : Либідь, 1993. 15 с.
6. Романюк С.К. Застосування статистичних методів у лінгвістичних дослідженнях. Рівне, 2015. 4 с.
7. Сас В.З. Маркери когерентності в ідіодискурсі Франца Кафки. 2017. 75 с.
8. Сидоренко І.А. Підходи до вивчення ідіостилу у лінгвістичних дослідженнях художнього тексту. Київ, 2105. 11 с.
9. Юсслер М. Социолінгвістика. Київ : Вища школа, 1987. 273 с.
10. Vinogradov, V. (1990). *Jazyk ruskoj literatury v osveshhenii* V.V. Vinogradova [Russian literature language by V. Vinogradov]. Moscow : Nauka.
11. Pascual Cantos Gómez Do we need statistics when we have linguistics? DELTA vol.18 no. 2. São Paulo, 2002.

References:

1. Bell, R.T. (1980). *Sotsyolynhvystyka: tsely, metody i problemy* [Sociolinguistics: goals, methods and problems]. Moskva: International Relations, 320 p.
2. Vinogradov, V.V. (1963). *Teoriya poetycheskoi rechy. Poetyka* [Stylistics. The theory of poetic speech. Poetics]. Moskva: Science, 255 p.
3. Perebyinis, V.I. (2003). *Shcho daie statystyka movoznavtsiam?* [What gives statistics to linguists?]. *Bulletin of the University of Kiev Linguistics: Coll. Sciences. wash. Series: Philology*. Kyiv: Kind of. Center 133 KNLU, Vol. 6, No. 2, pp. 27–32.
4. Perebyinis, V.S. (1970). *Kilkisni ta yakisni kharakterystyky systemy fonem suchasnoi ukrainkoi literaturnoi movy* [Quantitative and qualitative characteristics of the phoneme system of modern Ukrainian literary language]. Kyiv: Scientific Thought, 270 p.
5. Ponomariv, O.D. (1993). *Stylistyka suchasnoi ukrainkoi movy* [Stylistics of modern Ukrainian language]. Kyiv: Libid, 15 p.
6. Romaniuk, S.K. (2015). *Zastosuvannia statystychnykh metodiv u lnhvistychnykh doslidzhenniakh* [Application of statistical methods in linguistic research]. Rivne, 4 p.
7. Sas, V.Z. (2017). *Markery koherentnosti v idiodyskursi* [Markers of coherence in Franz Kafka's idiocourse], 75 p.
8. Sydorenko, I.A. (2015). *Pidkhody do vyvchennia idiostylu u lnhvistychnykh doslidzhenniakh khudozhnoho tekstu* [Approaches to the study of idiostyle in linguistic studies of artistic text]. Kiev, 11 p.
9. Jussler, M. (1987). *Sotsyolynhvystyka* [Sociolinguistics]. Kyiv: High School, 273 p.
10. Vinogradov, V. (1990). *Russian literature language by V. Vinogradov* [Jazyk ruskoj literatury v osveshhenii V.V. Vinogradova]. Moscow: Science.
11. Pascual Cantos Gómez (2002). Do we need statistics when we have linguistics? DELTA vol. 18 no. 2. São Paulo.