

МЕТОДИ КВАНТИТАТИВНОЇ ЛІНГВІСТИКИ В ДОСЛІДЖЕННІ ІДІОСТИЛЮ ПИСЬМЕННИКА (НА МАТЕРІАЛІ ТВОРІВ ЛЮКО ДАШВАР)

**Л. В. Шитик, доктор філологічних наук, професор,
завідувачка кафедри українського мовознавства і прикладної лінгвістики,
Черкаський національний університет імені Богдана Хмельницького
(Черкаси, Україна)**

e-mail: l_shytyk@ukr.net

ORCID: <https://orcid.org/0000-0001-5941-672>

Ідентифікатор Scopus-Author: 57210982685

**Т. В. Щербина, кандидат філологічних наук,
доцент кафедри українського мовознавства і прикладної лінгвістики,
Черкаський національний університет імені Богдана Хмельницького
(Черкаси, Україна)**

e-mail: Tanya_Bevz@i.ua

ORCID: <http://orcid.org/0000-0003-0008-5983>

Статтю присвячено лінгвостатистичному дослідженню ідіостилю Люко Дашвар. Мета наукового пошуку – вивчити специфіку ідіолекту письменниці шляхом моделювання статистичного профілю її художніх текстів, а також проаналізувати кількісні показники прийомів уведення в художні твори традиційних і модифікованих фразеологічних одиниць як виразної ознаки індивідуального мовлення авторки.

Визначено квантитативні параметри романів письменниці, як-от: обсяг тексту, обсяг словника, індекс різноманітності, середня повторюваність слова в тексті, кількість *h* нарах *legomena*, індекс винятковості тексту, індекс винятковості словника, кількість високочастотної лексики в тексті, кількість високочастотної лексики в словнику, індекс концентрації в тексті та індекс концентрації в словнику, індекс читабельності. Виконано квантитативну параметризацію текстів Люко Дашвар, репрезентовано статистичні характеристики фразеологізмів у романах письменниці. Звернено увагу на квантитативні маркери традиційних і модифікованих фразеологічних одиниць у романах «Молоко з кров'ю», «На запах м'яса» та «Рай.Центр», що відкриває нові можливості для вивчення стилістичних, лексичних та комунікативних особливостей художніх творів. Цей підхід дає змогу оцінити не лише якісний, а й кількісний аспект використання мовних одиниць, що допомагає виявити закономірності їх уживання в літературному дискурсі.

Ключові слова: ідіостиль, Люко Дашвар, лінгвостатистичний аналіз тексту, квантитативні параметри тексту, обсяг тексту, обсяг словника, індекс різноманітності, індекс винятковості, *h* нарах *legomena*, високочастотна лексика.

Актуальність. В епоху розвитку інформаційних технологій лінгвістичні явища потребують не лише якісного, але й кількісного вивчення, тому особливо актуальними стали міждисциплінарні дослідження, наприклад застосування статистичних методів аналізу мовних об'єктів. Кількісні виміри тексту, зокрема його лексичного наповнення, – це один із вагомих компонентів усебічного студіювання художнього мовлення й характеристики індивідуального стилю автора, яскравою ознакою якого є використання фразеологізмів – найвиразніших художньо-образних засобів мови. Кардинальні зміни суспільно-політичних орієнтирів, потреба в номінації нових понять і явищ, посилення експресивних функцій мови сприяли індивідуально-авторському переосмисленню фразеологічного матеріалу й трансформації стійких висловів, вивчення яких допомагає проникнути у сферу значення фразеологізмів, виявити глибину семантичних зв'язків компонентів з однойменними словами вільного вживання. Доповнюючи якісний лінгвістичний аналіз авторських трансформацій стійких мовних одиниць, квантитативні характеристики, з огляду на їхню безперечну об'єктивність, відіграють особливу роль у дослідженні ідіостилу письменника.

Аналіз останніх досліджень і публікацій. Методи квантитативної лінгвістики впродовж останніх десятиліть перебувають у колі зацікавлень як зарубіжних, так і вітчизняних учених. Зокрема, в українському мовознавстві кількісна інтерпретація лінгвістичних одиниць у різних текстах стала об'єктом наукових розвідок В. В. Левицького [13], А. П. Загнітка [9], В. І. Перебийніс [20], Г. Альтмана [2] С. Н. Бук [3] та ін.; лінгвостатистичні характеристики антропонімів у складі українських паремій описала Л. І. Лонська [14]. Квантитативні методи дослідження лінгвістичних явищ дають змогу виявити закономірності створення автором текстів, глибше пізнати внутрішню структуру твору, у чому переконують напрацювання О. О. Павличко [17], О. А. Павлишенко [18], І. М. Кульчицького [11] та ін.

Одним із сучасних напрямів квантитативної лінгвістики є стилеметрія, об'єкт якої – вимірювання статистичних стильових характеристик текстів та їхніх частин для впорядкування та систематизації (дослідження В. І. Перебийніс [19], С. Н. Бук [6], І. М. Кульчицького і З. О. Гладун [8] та ін.). Усе частіше вектор мовознавчих розвідок спрямований на кількісні характеристики індивідуального мовлення на матеріалі корпусів текстів. Зокрема, корпусно-статистичний підхід до дослідження ідіолекту Романа Іваничука застосували І. М. Кульчицький та К. І. Гнатовська [12], Н. Я. Лотоцька [15]; мову прози М. Зерова на тлі корпусу текстів проаналізувала М. О. Шведова [22]; статистичні характеристики ідіолекту Марії Матіос репрезентували Н. Я. Лотоцька та Р. Н. Мох [16]. Результати цих студій переконують, що застосування квантитативних методів для аналізу художніх текстів дає змогу об'єктивно представити індивідуальні особливості творчості письменників.

Мета статті – вивчити специфіку ідіолекту Люко Дашвар через моделювання статистичного профілю текстів письменниці, а також проаналізувати кількісні показники прийомів уведення в художні твори традиційних і модифікованих фразеологічних одиниць як однієї з маркаційних ознак індивідуального мовлення авторки.

Матеріалом дослідження послуговували романи Люко Дашвар «Молоко з кров'ю», «На запах м'яса» та «Рай.Центр».

Специфіка теми зумовила комплексний підхід до вибору взаємодоповнювальних **методів** вивчення досліджуваного матеріалу. Традиційний лінгвістичний аналіз ідіостилію письменника не виявляє характерних квантитативних особливостей розподілу фразеологізмів у досліджуваних текстах, тому задля аналізу кількісного співвідношення різновидів досліджуваних фразеологічних одиниць застосовано прийоми статистичного методу; із метою з'ясування прикладних аспектів дослідження прийомів уведення фразеологічних одиниць використано методи квантитативної та комп'ютерної лінгвістики.

Зважаючи на обсяг досліджуваного матеріалу, методи квантитативної лінгвістики вможливають об'єктивне, цілісне й точне вивчення індивідуального мовлення, однак, як слушно зауважують лінгвісти, недоліком їх використання є те, що «статистична методика більше орієнтована на порівняння та аналіз декількох авторів між собою» [12, с. 969–970], одним із критеріїв виокремлення індивідуальної частотної лексики є неповторюваність лексики одного автора в інших, а це «дещо ускладнює використання цього методу у випадку вивчення ідіостилію одного письменника» [там само, с. 970].

Результати дослідження та їх обговорення. Застосування статистичних методів потребує попередньої квантифікації лінгвістичного матеріалу, тобто числового представлення якісних явищ, після чого їх можна вивчати як кількісні об'єкти. Як констатують дослідники, донедавна в лінгвостатистичних працях, метою яких було вивчення частотних текстових закономірностей, аналізували порівняно невеликі обсяги [докл. див.: 18, с. 2], поділяючи твір на фрагменти – вибірки – й на підставі їх аналізу робили висновки про увесь обсяг матеріалу. Однак такий підхід, на нашу думку, не дає повної інформації про досліджуваний об'єкт, зокрема внеможливорює достовірні висновки про стилістичні особливості твору чи авторський ідіолект, для якого характерне використання унікальних одиниць. Такі специфічні мовні явища потребують аналізу великих за обсягом і збалансованих вибірок, якими в нашому дослідженні є романи Люко Дашвар.

Обрані твори відрізняються обсягом («Молоко з кров'ю» – 272 с., «На запах м'яса» – 368 с., «Рай.Центр» – 270 с.), їх можна схарактеризувати за різними лінгвістичними і

статистичними параметрами: кількістю лексичних одиниць; слів, належних до того чи до того лексико-граматичного класу; різноманітних поєднань слів, речень тощо. Задля укладання статистичного профілю досліджуваних творів створено підкорпус на базі Генерального регіонально анотованого корпусу української мови (ГРАК) [23], до складу якого ввійшли романи «Молоко з кров'ю», «На запах м'яса» та «Рай.Центр». Застосування функціоналу ГРАКу з метою квантитативного аналізу створеного підкорпусу дало змогу одержати такі статистичні показники: обсяг тексту (N), обсяг словника (V), кількість нарах legomena (V1), кількість високочастотної лексики в тексті (V10t), кількість високочастотної лексики в словнику (V10) (див. табл. 1). Для встановлення тенденцій та особливостей мовленнєвої практики письменника важливою є інформація про кількість не лише лексичних, але й синтаксичних одиниць у тексті, зокрема речень (S), списки яких у кожному з досліджуваних романів Люко Дашвар проаналізовано вручну, за розділовими знаками кінця речення (див. табл. 1). Так само вручну із загальної кількості крапок, знаків оклику, знаків питань, трикрапок і комбінацій цих знаків вилучено розділові знаки середини речення, наприклад трикрапку, що передає обірваність думки.

Таблиця 1

Загальні кількісні характеристики романів Люко Дашвар

Статистичний показник	«Молоко з кров'ю»	«На запах м'яса»	«Рай.Центр»
Обсяг тексту (N) (загальна кількість слів у тексті)	54788	91932	66324
Обсяг словника (V) (кількість різних слів у тексті; кількість лематизованих слів у тексті)	1600	2826	2019
Нарах legomena (V1) (слова, що трапляються в тексті лише 1 раз)	7419	12301	10111
Кількість високочастотної лексики в тексті (V10t)	746	1220	877
Кількість високочастотної лексики в словнику (V10)	806	1400	1001
Кількість речень у тексті (S)	8199	16512	10883
Кількість знаків у тексті (C)	293082	517668	375005

Використовуючи наявні в науковій літературі методики моделювання статистичного профілю ідіолекту письменників [4; 7; 16], на основі встановлених кількісних даних обчислено такі статистичні характеристики для кожного тексту в підкорпусі: індекс різноманітності (V/N), середню повторюваність слова в тексті (N/V), індекс винятковості тексту (V1/N), індекс винятковості словника (V1/V), індекс концентрації в тексті (V10t/N) та індекс концентрації в словнику (V10/V), індекс синтаксичної різноманітності (N/S), автоматизований індекс читабельності (див. табл. 2).

Таблиця 2

Індекси статистичного профілю романів Люко Дашвар

Показник	«Молоко з кров'ю»	«На запах м'яса»	«Рай.Центр»
Індекс різноманітності (багатство словника) (V/N)	0,03	0,03	0,03
Середня повторюваність слова в тексті (N/V)	34,24	32,53	32,84

Індекс винятковості тексту (V1/N)	0,135	0,133	0,152
Індекс винятковості словника (V1/V)	4,63	4,35	5
Індекс концентрації в тексті (частка тексту, яку охоплюють слова з високою частотою) (V10t/N)	0,013	0,013	0,013
Індекс концентрації в словнику (частка словника, яку становлять слова з високою частотою) (V10/V)	0,50	0,49	0,49
Індекс синтаксичної різноманітності (N/S)	6,68	5,57	6,09
Індекс читабельності (I _{ARI})	7,10	7,87	8,24

Під час вивчення показників лінгвістичних характеристик, визначених за кількома різними вибірками, постає питання про істотність розходження між ними. Проаналізувавши кількісні дані, які характеризують лексичний рівень досліджуваних романів (див. табл. 1), стверджуємо, що найвищі показники за обсягом слововживань, словником лексем, кількістю слів, що вжиті в тексті не більше ніж один раз, кількістю високочастотної лексики в тексті і в словнику демонструє твір «На запах м'яса». Аналіз кількості синтаксичних одиниць, зокрема речень, теж вирізняють цей роман з-поміж двох інших. Найнижчі показники за вищезгаданими параметрами має роман «Молоко з кров'ю».

Однак вивчення індексів статистичного профілю творів Люко Дашвар (див. табл. 2) засвідчує рівномірність репрезентації кількісних даних. Зокрема, однаковим для всіх аналізованих романів є один із найпотужніших у діагностиці тексту коефіцієнт словникової різноманітності, цінність якого полягає не лише у виявленні лексичних характеристик, а також в окресленні рис ідіостилу автора художнього твору. Зовсім не розрізняються досліджувані тексти й за індексами концентрації високочастотних слів у тексті, незначні відмінності між показниками концентрації високочастотних слів у словнику.

Засвідчено відмінності між показниками середньої повторюваності слова в тексті: трохи вищий, ніж в інших творах, індекс у романі «Молоко з кров'ю» (34,24), несуттєво відрізняються індекси в романах «На запах м'яса» (32,53) і «Рай.Центр» (32,84).

Привертають увагу загалом високі для художніх творів [пор.: 16, с. 68] індекси винятковості тексту і винятковості словника, що є ознаками варіативності лексики: найвищі ці показники в романі «Рай.Центр» – 0,152 і 5 відповідно; у творі «Молоко з кров'ю» – 0,135 і 4,63, у тексті «На запах м'яса» – 0,133 і 4,35. Крім того, роман «Молоко з кров'ю» відрізняється від інших досліджуваних текстів вищим показником синтаксичної різноманітності – 6,68 (пор. «Рай.Центр» – 6,09; «На запах м'яса» – 5,57).

Однією з важливих кількісних характеристик будь-якого тексту є показник його складності для сприймання та розуміння, який визначають за індексами читабельності [докладніше див.: 10, с. 82]. Автоматичний індекс читабельності (I_{ARI}) для досліджуваних творів Люко Дашвар визначено за формулою, що враховує кількість друкованих знаків в аналізованому тексті (C), слів (W) і речень (S):

$$I_{ARI} = 4,71 \times (C/W) + 0,5 \times (W/S) - 21,43$$

Найвищий індекс читабельності – у романі «Рай. Центр» (8,24), простіші для сприймання – «На запах м'яса» (7,87), «Молоко з кров'ю» (7,10).

Визначення показників лінгвістичних характеристик трьох різних вибірок передбачає необхідність з'ясування істотності статистичного розходження між значеннями їхніх коефіцієнтів за допомогою розрахунку χ^2 , який у лінгвостатистиці має назву «критерій однорідності». Для цього побудуємо таблицю, у якій кількість рядків визначена відповідно до кількості вибірок, а кількість колонок – до кількості показників (індексів), які зіставляємо (див. табл. 3).

Установлення істотності / неістотності статистичного розходження

	I ₁	I ₂	I ₃	I ₄	I ₅	I ₆	I ₇	I ₈	ΣT
«Молоко з кров'ю» (T ₁)	0,03	34,24	0,135	4,63	0,013	0,50	6,68	7,10	53,328
«На запах м'яса» (T ₂)	0,03	32,53	0,133	4,35	0,013	0,49	5,57	7,87	50,986
«Рай.Центр» (T ₃)	0,03	32,84	0,152	5	0,013	0,49	6,09	8,24	52,855
ΣI	0,09	99,61	0,42	13,98	0,039	1,48	18,34	23,21	157,169

Використовуємо схему обчислення χ^2 , запропоновану В. І. Перебийніс [19, с. 72–73], застосувавши формулу:

$$\chi^2 = S \times \left(\sum \frac{(\ln T_n)^2}{\sum \ln \Sigma T_n} - 1 \right)$$

Після необхідних обчислень отримано $\chi^2 = 1,84$. Для оцінювання істотності / неістотності розходження використано таблицю критичних значень χ^2 [там само, с. 155]. Визначено кількісний показник ступенів свободи (f), що обчислений як добуток кількості показників мінус один і кількості вибірок мінус один. Для нашої вибірки це $f = (8 - 1) \times (3 - 1) = 14$.

Розходження вважають істотним, якщо значення χ^2 більше за табличне для заданого рівня значущості. У нашій вибірці 1,84 – значно менше число, ніж критичні значення, подані в таблиці [23,7 (5 %) і 29,1 (1 %)]. Отже, розходження в статистичних показниках між трьома аналізованими романами Люко Дашвар мінімальне.

Залучення корпусних технологій та лінгвостатистичних методів дослідження до вивчення ідіостилю розширює діапазон аналізу проблеми, однак не вичерпує її, тому кількісні дані варто поєднувати з характеристикою власне лінгвістичних явищ. Зокрема, мовна палітра Люко Дашвар багата на фразеологізми, не лише традиційні, але й okazionalni, які письменниця використовує для досягнення найвищого емоційно-експресивного навантаження. Спробуємо за допомогою χ^2 -критерію зробити статистичну перевірку розподілів фразеологічних одиниць в одному з аналізованих творів, тобто відповідність спостережуваних даних розподілу деякому теоретичному закономірному розподілу. Критерій χ^2 та його ступінь вільності з'ясуємо на матеріалі роману «На запах м'яса». Основне завдання – дослідити статистичну однорідність вибірок стосовно функціонування трансформованих фразеологічних одиниць.

Таблиця 4

Установлення статистичної однорідності функціонування фразеологічних одиниць

m=2 \ n=3	i=1	i=2	i=3	Σjki
j=1	16	15	10	41
j=2	5	8	3	16
Σjki	21	23	13	N=57

У таблиці **m** – це кількість вибірок, **n** – кількість підвибірок. У процесі дослідження поділено текст на дві вибірки за кількістю сторінок; перша вибірка **j = 1**: 3–179, друга **j = 2**: 180–359 сторінки; підвибірки (**i**) першої вибірки (**j = 1**): **i = 1** – 3–60 с.; **i = 2** – 61–120 с.; **i = 3** – 121–179 с.; підвибірки (**i**) другої вибірки (**j = 2**): **i = 1** – 180–240 с.; **i = 2** – 241–300 с.; **i = 3** – 301–359 с.; **Σjki** – це загальна кількість частот у вибірках і підвибірках.

У першій вибірці модифіковані фразеологічні одиниці зафіксовано 41 раз, у трьох її підвибірках отримано частоти 16, 15, 10. У другій вибірці фразеологізми представлені 14 разів, у трьох її підвибірках мали частоти 5, 8, 3.

Використавши вже згадану формулу [19, с. 72–73], обраховуємо $\chi^2 = 0,795$. Довірчу ймовірність визначаємо після обчислення ступеня вільності ($f = 2$), що йому відповідають у

таблиці критичних значень два числа: 9,2 та 6,0. Перше з них – значення при довірчій імовірності 99 %, друге – при 95 %. Зазвичай беруть довірчу ймовірність 95 %. Показник $\chi^2 = 0,79 < 9,2$ (менший), тож розходження між порівнюваними вибірками несуттєве, вибірки статистично однорідні стосовно досліджуваного явища.

Отже, за допомогою критерію χ^2 з'ясовано, що дві вибірки одного роману є статистично однорідними. Дослідження ступеня вільності довело, що розходження в кількісному складі між порівнюваними вибірками не суттєве.

Ступінь щільності зв'язку не лише між кількісними, але і якісними характеристиками мовних явищ можна визначити за допомогою коефіцієнта рангової кореляції Спірмена. Відповідно до алгоритму розрахунку коефіцієнта кореляції рангів, спочатку замінюють кількісні (або напівкількісні, або якісні) ознаки ряду x та y на ранги, ранжуючи від меншої величини до більшої (або від більшої до меншої на розсуд дослідника); визначають умовні відхилення, тобто різницю рангів за кожним рядком; зводять умовні відхилення у квадрат; визначають суму квадратів умовних відхилень; підставляють отримані дані у відому формулу й обчислюють коефіцієнт кореляції [5, с. 55].

Оскільки обчислення коефіцієнта кореляції рангів зазвичай проводять на малій кількості чисел спостережень, особливого значення набуває оцінка статистичної значущості (достовірності) цього коефіцієнта. Ранговий коефіцієнт завжди менш точний, ніж парний коефіцієнт кореляції.

Його обчислюємо за формулою:

$$p = 1 - \frac{6 \sum_{j=1}^n (R_j - r_j)^2}{n(n^2 - 1)}$$

P – це коефіцієнт кореляції рангів, n – кількість текстів, Σ – сума рангів, R_j – ранг за першою ознакою, r – ранг за другою ознакою.

Коефіцієнт рангової кореляції Спірмена визначаємо на підставі аналізу трьох романів Люко Дашвар: «Молоко з кров'ю», «На запах м'яса», «Рай. Центр», досліджуючи вживання в них трансформованих фразеологічних одиниць.

$J = 1$ («Молоко з кров'ю» 272 сторінки); $j = 2$ («На запах м'яса» 368 сторінок); $j = 3$ («Рай. Центр» 270 сторінок). Ранг за обсягом та за частотою з'ясовуємо в порядку від більшого до меншого; ранг за обсягом (R) присвоюємо залежно від кількості сторінок, а ранг за частотою (r) – залежно від частоти вживання фразеологічної одиниці в тому чи в тому тексті. Різницю рангів ($R-r$) з'ясовуємо за допомогою віднімання суми одного рангу від суми іншого та підносимо знайдену різницю до квадрата, щоб знайти квадратичну суму $(R-r)^2$. Ранжуємо в послідовності від меншого до більшого (див. табл. 5).

Таблиця 5

Ранги за обсягом та частотою

№	Обсяг тексту	Частота ФО	Ранг за обсягом R	Ранг за частотою r	Різниця рангів (R-r)	Різниця рангів (R-r) ²
j = 1	272	34	3	2	1	1
j = 2	368	52	1	1	0	0
j = 3	270	43	2	3	-1	-1
						Сума: 0

Оцінку достовірності коефіцієнта кореляції рангів визначають за допомогою різних методів залежно від числа спостережень. Якщо число парних спостережень менше за число 9 ($n = 9$), оцінку значущості проводять за спеціальною таблицею [див. 5, с. 78]. Коли число спостережень від 10 і більше, оцінку значущості обчислюють за допомогою критерію t за формулою:

$$t = p \sqrt{\frac{n-2}{j-p^2}}$$

Число спостережуваних нами явищ дорівнює трьом, тому результати звіряємо з таблицею, у якій n – це кількість досліджуваних явищ, а достовірність представлена при 95 % та 99 %.

$$P = 1 - \frac{6 \cdot 0}{3(9 - 1)} = 1$$

У таблиці достовірність для числа явищ репрезентована для 99 %, вона повністю збігається, тобто перебуває в межах норми.

Щоб визначити відношення поширення трансформованих фразеологізмів у романі «Рай.Центр», треба врахувати, що лінгвометрія вимірює й аналізує кількісні характеристики одиниць різних рівнів мови. Для визначення таких характеристик використовують коефіцієнти мовної різноманітності.

Коефіцієнт лексичної різноманітності тексту – це відношення кількості слів, у нашому разі фразеологізмів, до загальної кількості слів тексту, тобто:

$$K_l \equiv \frac{L}{S}$$

K_l – це коефіцієнт лексичної різноманітності; L – це кількість модифікованих ФО, які вжиті в тексті; S – це загальна кількість слів у тексті.

Щоб визначити коефіцієнт поширення трансформованих ФО, потрібно поділити суму модифікованих фразеологізмів на загальну кількість слів у тексті:

$$K_l = \frac{43}{66324} = 0,000648$$

Значення коефіцієнта лежить у межах 0–1. Що більшим є отриманий десятковий дріб, тим вищою є лексична різноманітність, а отже, цей показник для роману суттєвий.

Як слушно зауважують лінгвісти, статистичні показники не можуть визначати естетичних можливостей художніх творів, а, відображаючи лише деякі структурні особливості текстів, «можуть лише вказувати на приховані від простого спостереження особливості індивідуального чи функціонального стилю» [20, с. 152; 21, с. 362], тому вивчення оригінального лінгвістичного портрета Люко Дашвар потребує поєднання квантитативних методів із власне лінгвістичними.

Висновки і перспективи. Залучення корпусних технологій і методів квантитативної лінгвістики дало змогу вивчити статистичний профіль творів Люко Дашвар, засвідчити рівномірність репрезентації кількісних даних у кожному з аналізованих романів. Істотність статистичного розходження між значеннями коефіцієнтів встановлено за допомогою розрахунку χ^2 ; визначено ступінь вільності трансформованих фразеологічних одиниць на матеріалі роману «На запах м'яса». З'ясовано, що аналізовані вибірки статистично однорідні, а дослідження ступеня вільності довело, що розходження в кількісному складі між порівнюваними вибірками не суттєве. Визначено коефіцієнт різноманітності для трансформованих фразеологізмів у романі «Рай.Центр» та коефіцієнт рангової кореляції Спірмена на матеріалі трьох романів «Молоко з кров'ю», «На запах м'яса» та «Рай.Центр», на підставі чого доведено лексичну різноманітність текстів.

Отже, квантитативні методи надають точні, структуровані дані про особливості авторського стилю, мовну насиченість тексту та функційне навантаження мовних одиниць. Вони є ефективним інструментом для дослідження творчості Люко Дашвар, оскільки дають змогу глибше проаналізувати її мовні стратегії, стилістичні уподобання та унікальність у сучасному літературному контексті.

У перспективі плануємо комплексне статистичне зіставлення всіх художніх творів письменниці для з'ясування специфіки використання синтаксичних конструкцій.

Список використаної літератури

1. Алексієнко Л. О., Дарчук Н. П. Принципи створення параметризованої бази даних за поетичними текстами Лесі Українки. *Леся Українка і сучасність* (до 130-річчя від дня народження Лесі Українки): зб. наук. праць. Луцьк, 2004. С. 344–352.
2. Альтман Г. Мода та істина в лінгвістиці. *Проблеми квантитативної лінгвістики*. Чернівці: Рута, 2005. С. 3–10

3. Бук С. Багатство словника тексту: досвід полоністики. *Současná slovanská lexikologie a lexikografie*. Прага, 2014. С. 276–279.
4. Бук С. Квантитативна параметризація текстів Івана Франка: проект та його реалізація. *Вісник Львівського університету. Серія філологічна*. 2013. Вип. 58. С. 290–307.
5. Бук С. Основи статистичної лінгвістики : [навчально-методичний посібник]. Львів : Видавничий центр ЛНУ імені Івана Франка, 2008. 124 с.
6. Бук С. Статистичні характеристики лексики основних функціональних стилів української мови: спроба порівняння. *Лексикографічний бюлетень*. 2006. № 13. С. 166–170.
7. Бук С. Сучасні методи дослідження мови письменника у слов'янознавстві. *Проблеми слов'янознавства*. 2012. № 61. С. 86–95.
8. Гладун З. О., Кульчицький І. М. Основні підходи дослідження індивідуального стилю письменника. *Молодий вчений*. 2019. № 11 (75). С. 944–946.
9. Загнітко А. Категорійний і парадигмальний простір сучасної лінгвоперсонології. *Studia linguistica*. 2016. № 9. С. 150–159.
10. Кричківська А. М., Паращин Ж. П., Швед О. В., Губицька І. І., Болібрех Л. Д., Новіков В. П. Застосування інформаційних технологій для стандартизації методології створення навчальної літератури. *Вісник Національного університету «Львівська політехніка»*. 2014. № 803 : Інформатизація вищого навчального закладу. С. 81–85.
11. Кульчицький І. М., Данчевська Ю. О. Деякі аспекти створення корпусу художніх творів В. С. Стефаника. *Прикладна лінгвістика та лінгвістичні технології. Megaling 2012* : зб. наук. праць. Київ, 2013. С. 143–148.
12. Кульчицький І. М., Гнатівська К. І. Дослідження ідіолекту письменника: корпусно-статистичний підхід. *Молодий вчений*. 2019. № 11 (75). С. 968–970.
13. Левицький В. В. Квантитативні методи в лінгвістичесі. Вінниця : Нова книга, 2007. 264 с.
14. Лонська Л. І. Антропоніми в складі українських паремій: лінгвостатистичне дослідження. *Мовознавчий вісник* : зб. наук. праць. Черкаси, 2018. Вип. 24–25. С. 69–77.
15. Лотоцька Н. Я. Ідіолект Романа Іванчука: корпуснобазований та лінгвокогнітивний підходи : дис. ...докт. філософії : 035 / Нац. ун-т «Львівська політехніка». Львів, 2021. 324 с.
16. Лотоцька Н. Я., Мох Р. Н. Специфіка ідіолекту Марії Матіос: статистичний профіль. *Закарпатські філологічні студії*. Ужгород, 2023. Вип. 29. Т. 1. С. 64–70.
17. Павличко О. О. Щодо статистичних параметрів авторського стилю (на матеріалі творів Е. М. Ремарка). *Мовні і концептуальні картини світу* : зб. наук. праць. Київ : ВПЦ «Київський університет», 2010. Вип. 29. С. 186–191.
18. Павлишенко О. А. Квантитативні характеристики лексико-семантичних полів дієслова в авторських текстах англословної художньої літератури : автореф. дис. ...канд. філол. наук : 10.02.04 / Львів. нац. ун-т ім. Івана Франка. Львів, 2017. 24 с.
19. Перебийніс В. І. Статистичні методи для лінгвістів. Вінниця : Нова книга, 2002. 168 с.
20. Перебийніс В. С., Муравицька М. П., Дарчук Н. П. Частотні словники та їх використання. Київ : Наукова думка, 1985. 204 с.
21. Цьох Л. Й., Іваночко В. Я. Паралельні корпуси у дослідженнях відтворення ідіостилу у перекладі (на основі паралельного корпусу роману Дж. Лондона «White Fang»). *Молодий вчений*. 2021. № 10 (98). С. 358–363.
22. Шведова М. О. Мова прози Миколи Зерова на тлі корпусу текстів 1910–1935 років. *Мовна особистість: лінгвістика і лінгводидактика*. Київ ; Черкаси : вид-во ФОП Гордієнко Є. І., 2020. 402 с.
23. Шведова М., фон Вальденфельс Р., Яригін С., Рисін А., Старко В., Ніколаєнко Т. та ін. Генеральний регіонально анотований корпус української мови (ГРАК). Київ ; Львів ; Єна, 2017–2024. URL : <http://uacorpus.org>.

Список використаних джерел

1. Дашвар Люко. Молоко з кров'ю : роман. Харків : Книжковий Клуб «Клуб Сімейного Дозвілля», 2010. 272 с.
2. Дашвар Люко. На запах м'яса : роман. Харків : Книжковий Клуб «Клуб Сімейного Дозвілля», 2013. 368 с.
3. Дашвар Люко. Рай. Центр : роман. Харків : Книжковий Клуб «Клуб Сімейного Дозвілля», 2010. 272 с.

References

1. Aleksiienko, L. O. & Darchuk, N. P. (2004). Pryntsypy stvorennia parametryzovanoi bazy danykh za poetychnymy tekstamy Lesi Ukrainky [Principles of creating a parameterized database of Lesya Ukrainka's poetic texts]. In: *Lesia Ukrainka i suchasnist (do 130-richechia vid dnia narodzhennia Lesi Ukrainky) [Lesya Ukrainka and modernity (to the 130th anniversary of Lesya Ukrainka's birth)]*: zb. nauk. prats. Lutsk, 344–352 (in Ukr.).
2. Altman, G. (2005). Moda ta istyna v lnhvistytsi [Fashion and truth in linguistics]. In: *Problemy kvantytatyvnoi lnhvistyky [Problems of Quantitative Linguistics]*. Chernivtsi: Ruta, 3–10 (in Ukr.).
3. Buk, S. (2014). Bahatstvo slovyka tekstu: dosvid polonistyky [The richness of the text's vocabulary: the experience of Polish studies]. In: *Současná slovanská lexikologie a lexikografie [Contemporary Slavic lexicology and lexicography]*. Praha, 276–279 (in Ukr.).
4. Buk, S. (2013). Kvantytatyvna parametryzatsiia tekstiv Ivana Franka: proekt ta yoho realizatsiia [Quantitative parameterization of Ivan Franko's texts: the project and its implementation]. In: *Visnyk Lvivskoho universytetu. Seriiia filolohichna [Bulletin of Lviv University. Philological Series]*, 58, 290–307 (in Ukr.).

5. Buk, S. (2008). Osnovy statystychnoi linhvistyky [Fundamentals of statistical linguistics]: navchalno-metodychnyi posibnyk. Lviv: Vydavnychiy tsentr LNU imeni Ivana Franka, 124 (in Ukr.).
6. Buk, S. (2006). Statystychni kharakterystyky leksyky osnovnykh funktsionalnykh styliv ukrainskoi movy: sprobа porivniannia [Statistical characteristics of the vocabulary of the main functional styles of the Ukrainian language: an attempt at comparison.]. In: *Leksykohrafichnyi biuleten* [Lexicographic bulletin], 13, 166–170 (in Ukr.).
7. Buk, S. (2012). Suchasni metody doslidzhennia movy pysmennyka u slovianoznavstvi [Modern techniques for studying writer's language in Slavic linguistics]. In: *Problemy slovianoznavstva* [Problems of Slavic studies], 61, 86–95 (in Ukr.).
8. Hladun, Z. O. & Kulchytskyi, I. M. (2019). Osnovni pidkhody doslidzhennia indyvidualnoho stylu pysmennyka [Basic approaches to the study of the individual style of the writer]. In: *Molodyi vchenyi* [Young scientist], 11(75), 944–946 (in Ukr.).
9. Zahnitko, A. (2016). Katehoriinyi i paradyhmalnyi prostir suchasnoi linhvopersonolohii [Categorical and Paradigmatic Aspect of Modern Linguopersonology]. In: *Studia linguistica*, 9, 150–159 (in Ukr.).
10. Krychkovska, A. M., Parashchyn, Zh. P., Shved, O. V., Hubytska, I. I., Bolibrukh, L. D. & Novikov, V. P. (2014). Zastosuvannia informatsiinykh tekhnolohii dlia standartyzatsii metodolohii stvorennia navchalnoi literatury [Application of information technologies for standardization of the methodology for creating educational literature.]. In: *Visnyk Natsionalnoho universytetu «Lvivska politekhnika»* [Bulletin of Lviv Polytechnic National University], 803: Informatyzatsiia vyshchoho navchalnoho zakladu, 81–85 (in Ukr.).
11. Kulchytskyi, I. M. & Danchevska, Yu. O. (2013). Deiaiki aspekty stvorennia korpusu khudozhnykh tvoriv V. S. Stefanyka [Some aspects of creating a corpus of V. S. Stefanyk's literary works]. In: *Prykladna linhvistyka ta linhvistychni tekhnolohii. Megaling 2012* [Applied linguistics and linguistic technologies. Megaling 2012]: zb. nauk. prats. Kyiv, 143–148 (in Ukr.).
12. Kulchytskyi, I. M. & Hnatovska, K. I. (2019). Doslidzhennia idiolektu pysmennyka: korpusno-statystychnyi pidkhid [Investigation of the writer's idiolect: a corpus-statistical approach]. In: *Molodyi vchenyi* [Young scientist], 11(75), 968–970 (in Ukr.).
13. Levitskii, V. V. (2007). Kvantitativnie metody v lingvistike [Quantitative methods in linguistics]. Vinnitsa: Nova kniga, 264 (in Ukr.).
14. Lons'ka, L. I. (2018). Antroponimy v skladi ukrainskykh paremii: linhvostatystychnie doslidzhennia [Anthroponemes-woman's names in the structure of Ukrainian proverbs: quantitative investigation]. In: *Movoznavchyi visnyk* [Linguistic Bulletin]: zb. nauk. prats. Cherkasy, 24–25, 69–77 (in Ukr.).
15. Lototska, N. Ya. (2021). Idiolekt Romana Ivanychuka: korpusnobazovanyi ta linhvokohnityvnyi pidkhody [The idiolect of Roman Ivanychuk: corpus-based and linguo-cognitive approach]. PhD dissertation (Philology). Lviv, 324 (in Ukr.).
16. Lototska, N. Ya. & Mokh, R. N. (2023). Spetsyfika idiolektu Marii Matios: statystychnyi profil [The specificity of Maria Matios' idiolect: a statistical profile]. In: *Zakarpatski filolohichni studii* [Transcarpathian Philological Studies]. Uzhhorod, 29, 1, 64–70 (in Ukr.).
17. Pavlychko, O. O. (2010). Shchodo statystychnykh parametriv avtorskoho stylu (na materiali tvoriv E. M. Remarka) [On the statistical parameters of the author's style (based on the works of E. M. Remarque)]. In: *Movni i kontseptualni kartyny svitu* [Linguistic and conceptual worldviews]: zb. nauk. prats. Kyiv: VPTs «Kyivskiy universytet», 29, 186–191 (in Ukr.).
18. Pavlyshenko, O. A. (2017). Kvantytatyvni kharakterystyky leksyko-semantychnykh poliv diieslova v avtorskykh tekstakh anhlomovnoi khudozhnoi literatury [Quantitative characteristics of lexical-semantic fields of verbs in the author's texts of English literature]. Extended abstract of PhD dissertation (Germanic languages). Lviv, 24 (in Ukr.).
19. Perebyinis, V. I. (2002). Statystychni metody dlia linhvistiv [Statistical methods for linguists]. Vinnytsia: Nova knyha, 168 (in Ukr.).
20. Perebyinis, V. S., Muravytska, M. P. & Darchuk, N. P. (1985). Chastotni slovnyky ta yikh vykorystannia [Frequency dictionaries and their use]. Kyiv: Naukova dumka, 204 (in Ukr.).
21. Tsokh, L. Y. & Ivanochko, V. Ya. (2021). Paralelni korpusy u doslidzhenniakh vidtvorennia idiostyliu u perekladi (na osnovi paralelnoho korpusu romanu Dzh. Londona «White Fang») [Parallel corpora in the study of idiostyle reproduction in translation (based on the parallel corpus of J. London's «White Fang»)]. In: *Molodyi vchenyi* [Young Scientist], 10 (98), 358–363 (in Ukr.).
22. Shvedova, M. O. (2020). Mova prozy Mykoly Zerova na tli korpusu tekstiv 1910–1935 rokiv [The language of Mykola Zerov's prose against the background of the corpus of texts from 1910–1935.]. In: *Movna osobystist: linhvistyka i linhvodydaktyka* [Linguistic personality: linguistics and linguodidactics]. Kyiv; Cherkasy: vyd-vo FOP Hordiienko Ye. I., 402 (in Ukr.).
23. Shvedova, M., fon Valdenfels, R., Yaryhin, S., Rysin, A., Starko, V., Nikolaenko, T. ta in. (2017–2024). Heneralnyi rehionalno anotovanyi korpus ukrainskoi movy (HRAK) [General Regionally Annotated Corpus of Ukrainian (GRAK)]. Kyiv; Lviv; Yena. Available at: <http://uacorporus.org>.

List of used sources

1. Dashvar, Liuko (2010). Moloko z kroviu [Milk with blood]: roman. Kharkiv: Knyzhkovyi Klub «Klub Simeinoho dozvillia», 272 (in Ukr.).
2. Dashvar, Liuko (2013). Na zapakh miasa [The smell of meat]: roman. Kharkiv: Knyzhkovyi Klub «Klub Simeinoho Dozvillia», 368 (in Ukr.).

3. Dashvar, Liuko (2010). Rai.Tsentr: roman. Kharkiv: Knyzhkovyi Klub «Klub Simeinoho Dozvillia», 272 (in Ukr.).

METHODS OF QUANTITATIVE LINGUISTICS IN THE STUDY OF THE WRITER'S IDIOSTYLE (BASED ON LYUKO DASHVAR'S WORKS)

**Liudmyla Shytyk, Doctor of Philological Sciences, Professor,
Head of the Department of Ukrainian Linguistics and Applied Linguistics,
Bohdan Khmelnytsky National University of Cherkasy (Cherkasy, Ukraine)**

e-mail: l_shytyk@ukr.net

ORCID: <https://orcid.org/0000-0001-5941-672>

Identifier Scopus-Author: 57210982685

**Tetyana Shcherbyna, Candidate of Philological Sciences, Associate Professor
of the Department of Ukrainian Linguistics and Applied Linguistics,
Bohdan Khmelnytsky National University of Cherkasy (Cherkasy, Ukraine)**

e-mail: Tanya_Bevz@i.ua

ORCID: <http://orcid.org/0000-0003-0008-5983>

Abstract. Introduction. In the era of information technology development, linguistic phenomena require not only qualitative but also quantitative study. Considering this, interdisciplinary studies focused on applying statistical methods to the analysis of linguistic objects have become particularly relevant. Quantitative characteristics of the lexical content of a text are one of the important components of a comprehensive analysis of literary speech and the author's individual style characteristics.

Purpose. The purpose of the article is to study the specifics of Luko Dashvar's idiolect by modelling the statistical profile of the writer's texts, as well as to analyse the quantitative indicators of the methods of introducing traditional and transformed phraseological units into works of fiction as a distinctive sign of the author's individual speech.

Methods. Statistical methods were used to analyse the quantitative correlation of the varieties of the studied phraseological units; methods of quantitative and computational linguistics were used to clarify the applied aspects of the study of the methods of introducing phraseological units.

Results. The use of corpus technologies and quantitative linguistics methods made it possible to study the statistical profile of Luko Dashvar's works and to demonstrate the uniformity of the representation of quantitative data in each of the analysed novels. The significance of the statistical difference between the coefficient values was determined using the χ^2 calculation; the degree of freedom of the transformed phraseological units was defined on the material of the novel «The Smell of Meat». The analysed samples were found to be statistically homogeneous, and the study of the degree of freedom proved that the differences in the quantitative composition between the compared samples are not significant. The diversity coefficient for transformed phrasemes in the novel «Rai.Tsentr» and the Spearman's rank correlation coefficient are determined on the basis of three novels «Milk with Blood», «The Smell of Meat» and «Rai.Tsentr». The lexical diversity of the texts is proved on the basis of the determined coefficients.

Originality. For the first time, quantitative parameterization of Luko Dashvar's texts was performed, and the statistical characteristics of the transformed phrasemes of the writer's novels were represented.

Conclusion. Quantitative methods provide accurate, structured data on the features of the author's style, linguistic richness of the text and the functional load of language units. They serve as an effective tool for researching Luko Dashvar's works, as they allow for a deeper analysis of her linguistic strategies, stylistic preferences, and uniqueness in the context of contemporary fiction writing.

Key words: idiolect, Luko Dashvar, linguostatistical text analysis, quantitative text parameters, text volume, vocabulary volume, diversity index, exclusivity index, hapax legomena, general vocabulary.

Надійшла до редакції: 31.08.2024

Прийнято до друку: 02.10.2024