

О.О. Окунькова, ст. викладач  
Д.О. Варганова, ст. викладач  
О.Ю. Дячук, ст. викладач  
М.С. Колошук, асистент

Державний університет «Житомирська політехніка»

## Квантитативні дослідження творів Люко Дашвар

(Представлено: к.т.н. Єфіменко А.А.)

*Квантитативна лінгвістика – це потужний інструмент, який дозволяє досліджувати мову за допомогою математичних методів і статистичного аналізу. Ці методи дозволяють перетворити якісні лінгвістичні дані на кількісні, що дає змогу проводити більш об'єктивний та поглиблений аналіз текстових даних.*

*У статті використано методи описової статистики, такі як середнє значення, середнє квадратичне відхилення, міра коливання середньої частоти, стандартна та відносна похибка. Для опису основних характеристик текстів використовуються методи розрахунку частотності слів, довжини та типу речень, насиченість тексту діалогами тощо.*

*Описано можливість використання методів математичної статистики у лінгвістиці, основні напрямки їх застосування та впровадження одержаних результатів для дослідження текстової інформації. Наприклад, визначення авторства тексту та аналіз емоційного забарвлення.*

*Обґрунтовано актуальність проведення нових досліджень над наявними та новими збірками текстів, що дає змогу як підтвердити раніше виявлені статистичні параметри та закономірності будови текстів, так і визначити нові, структурні особливості сучасної української мови. Дослідження, зазначені в статті, є основою статистичної типології текстів та мов.*

*Для автоматизації розрахунку результатів аналізу було використано мову програмування Python. Python є одним з найпопулярніших інструментів для дослідження текстів завдяки своїй гнучкості та великій кількості бібліотек. За допомогою Python можна автоматизувати багато рутинних завдань, пов'язаних з обробкою природної мови.*

*Квантитативні показники отримано в результаті досліджень творів сучасної української письменниці Л.Дашвар.*

**Ключові слова:** статистика; частотність; варіаційний ряд; абсолютне значення; відносне значення; квантитативні дослідження.

**Постановка проблеми.** Квантитативний аналіз тексту дає можливість дослідити його індивідуальні особливості на всіх рівнях від фонетики до синтаксису, визначити принципи їх використання, переконатись у необхідності їх дослідження та вивчення. Це є особливо актуальним для української мови, яка довгий час зазнавала переслідувань та обмежень, утисків та заборон довгий час, і лише після набуття Україною незалежності почала шлях вільного розвитку в усіх напрямках. В останні роки значно зросла доля україномовного контенту в усіх сферах життя. Тому надзвичайно важливо продовжити вивчення законів та принципів сучасного українського мовотворення.

**Метою статті** є дослідження сучасних україномовних художніх текстів засобами математичної статистики та інструментів мови програмування Python та вивчення найбільш ефективних підходів до їх аналізу.

**Аналіз останніх досліджень та публікацій.** Застосування математичних методів дослідження мови відоме з давніх часів. Як правило, раніше їх використовували не постійно, в той час як зараз їх застосовують свідомо і цілеспрямовано. Активне використання математичних методів почалося ще в середині ХХ століття. Одним з перших дослідників мови був Фердинанд де Соссюр, який звернувся до застосування статистичних методів для лінгвістичних досліджень. Науковець також порівнював мову з грою в шахи, у якій важливі правила та значення, а матеріал, з якого зроблені фігури, не має значення [10].

Статистичні дослідження української мови розпочато у другій половині минулого століття, коли в Інституті мовознавства ім. О.О. Потебні АН УРСР було створено групу структурно-математичної лінгвістики. На сьогодні статистичні дослідження виконують в Українському мовно-інформаційному фонді, Інституті лінгвістики Київського національного університету ім. Тараса Шевченка, Львівському національному університеті ім. Івана Франка, Національному університеті «Львівська політехніка» та ін. [2, 7].

Впровадження прикладних комп'ютерних технологій в українське мовознавство розпочалося зі створення теоретичної та методико-процедурної бази (В.Перебийніс «Статистичні параметри стилів» 1967, «Кількісні та якісні характеристики системи фонем української мови» 1970).

Заведено розрізняти кількісні й статичні методи. До кількісних методів належать підрахунки частоти вживання мовних одиниць, а статистичні методи передбачають використання різних формул для виявлення правил розподілу мовних одиниць у мовленні, виміру зв'язків між мовними елементами, встановлення тенденцій у розвитку та функціонуванні мови [8].

В наш час статистику все частіше використовують для лінгвістичних досліджень, наприклад, для вивчення особливостей авторської мови та тональності текстів. Статистичні методи істотно видозмінюють та збільшують наші знання як про саму систему мови, так і про особливості її функціонування [4, 5].

Кульчицький І.М. стверджує, що для досліджень у галузі лінгвістики часто використовують імовірно-статистичні методи математичної лінгвістики, а також методи та прийоми математичної теорії множин, а саме моделювання множин, визначення множин, взаємне накладання множин з метою їх порівняння. Порівняльні методи широко застосовують для вивчення тематичних та формальних класів мовних одиниць і, хоч кількісні методи зазвичай зараховують до практичної сфери, їх доволі широко використовують для розв'язання теоретичних проблем мовознавства [6, 7].

Науковиця В.Перебийніс вважає, що всі складні системи підпорядковуються статистичним законам. Такими системами є живі організми, економіка, а також наука та людська мова. Вона вказує на те, що необхідно використовувати статистичні методи у мовознавстві, адже мова є складною системою, а відтак підпорядковується статистичним законам. Вона має кількісні характеристики, притаманні усім рівням системи. Таким чином, і будова мови, і її функціонування в мовленні, і співвідношення мови й мислення, мови й суспільства підкоряються дії статистичних законів, тому вивчати їх необхідно завдяки статистичними методами [8–10].

Робота багатьох мовознавців базується на визнанні факту підпорядкованості мови й мовлення статистичним законам. Статистичне обстеження великого обсягу матеріалу має на меті встановлення деяких характеристик мови (визначення валентності, тобто кола сполучуваності кожної одиниці) і мовлення (ланцюжка зв'язків, що реалізовані в мовленні, питома вага кожного зв'язку і ланцюжки зв'язків у текстах різних функціональних стилів). Робота над лінгвістичними дослідженнями потребує опрацювання матеріалу, що передбачає володіння методами математичної статистики [3, 8, 9].

На сьогодні в Україні квантитативними та статистичними дослідженнями мовних явищ займаються М.Кригін [4], С.Бук [2, 3], М.Крупа [5], І.Кульчицький [6, 7] та ін.

**Викладення основного матеріалу.** Для досягнення поставленої мети було проаналізовано три твори сучасної української письменниці, журналістки та сценаристки Любо Дашвар (Ірини Іванівни Чернової) «Мати все», «Село не люди. Вбити свідка», «Галя без голови». Творчість Л.Дашвар неодноразово нагороджена вітчизняними літературними преміями, зокрема її перший роман «Село не люди» став лауреатом премії «Коронація слова 2007», книга «Молоко з кров'ю» перемогла в номінації «Книга року Бі-Бі-Сі». Також письменниця отримала відзнаку «Золотий письменник України». Твори письменниці неодноразово отримували схвальні відгуки літературних критиків та пересічних читачів.

У процесі аналізу текстів використовувалися методи квантитативної лінгвістики, що є розділом математичної лінгвістики, яка досліджує мову за допомогою статистичних методів. Дослідження текстів базуються на одному з найосновніших понять квантитативної лінгвістики – частоти досліджуваної одиниці, яка обчислюється в абсолютному та відносному еквіваленті. Абсолютна частота – це кількість вживань певної одиниці (літери, слова, словоформи, словосполучення, речення тощо) в обстеженому матеріалі. Відносна частота – відношення абсолютної частоти певної одиниці у вибірці до обсягу вибірки. Вимірюється у відсотках (%) або в частках [6, 11, 12].

Проте в різних вибірках частота одиниці зазвичай неоднакова. У таких випадках потрібно оперувати середньою частотою. Це відношення суми абсолютних частот певної одиниці у підвибірках до кількості підвбірок. Якщо середня частота більша за одиницю, то доцільно її рахувати з точністю до сотих (два знаки після коми), якщо менша за одиницю, то її треба рахувати до другого знака після коми або до третього знака, якщо після коми йде нуль.

До ключових параметрів квантитативних (кількісних) досліджень текстів також зараховують довжину речень. Як частина цілого тексту, речення неминуче несе додаткове навантаження, створюючи особливі синтаксичні малюнки, ритми, темп розвитку та інтонаційні структури твору. Наприклад, різкі зміни довжини надають висловлюванню нерівний і незграбний ритм, що має ефект динамічної напруги, незалежно від лексичного складу речення. І навпаки, розвиток тексту реченнями майже рівномірної довжини, близької до середньої, створює ефект плавного, ненав'язливого викладу [2].

Визначення квантитативних показників слів у текстах Л.Дашвар було реалізовано за алгоритмом: для кожної досліджуваної одиниці спочатку розраховано абсолютні та відносні частоти її значень, за яких було побудовано варіаційні ряди (табл. 1–3). Проаналізовано отримані варіаційні ряди та розраховано статистичні оцінки середньої частоти. Визначено середнє квадратичне відхилення, міру середньої

флуктуації частоти, стандартну помилку відхилення та відносну помилку. Розрахунки проводилися за стандартними формулами [6]. Під час проведення розрахунків речення будь якого типу рахувалось як одне. Слова, що пишуться через дефіс, вважались одним словом.

Середнє значення обчислюється за формулою:

$$\bar{x} = \frac{\sum x_i n_i}{\sum n_i}, \tag{1}$$

де  $x_i$  – значення варіанта,  $n_i$  – кількість появ варіанта у досліді,  $i$  – номер варіанта.

Середнє квадратичне відхилення розраховано за формулою:

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 n_i}{\sum n_i}}, \tag{2}$$

де  $x_i$  – значення варіанта,  $n_i$  – кількість появ варіанта у досліді,  $i$  – номер варіанта,  $\bar{x}$  – середнє значення.

Міру коливання середньої частоти визначали за формулою:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{\sum n_i}}, \tag{3}$$

де  $\sigma$  – середнє квадратичне відхилення,  $n_i$  – кількість появ варіанта у досліді,  $i$  – номер варіанта,  $\bar{x}$  – середнє значення.

Стандартну похибку визначають за формулою:

$$S_{\bar{x}} = \frac{\sigma}{\sqrt{\sum n_i - 1}}. \tag{4}$$

Відносну похибку обчислюють за формулою:

$$\epsilon = \frac{1,96\sigma_{\bar{x}}}{\bar{x}}, \tag{5}$$

де  $\sigma$  – середнє квадратичне відхилення,  $n_i$  – кількість появ варіанта у досліді,  $i$  – номер варіанта,  $\bar{x}$  – середнє значення,  $\sigma_{\bar{x}}$  – міра коливання середньої частоти.

Розрахунки статистичних показників були проведені в PyChartm – інтегрованому середовищі розробки для мови програмування Python.

Отримані результати було порівняно з аналогічними показниками інших українських авторів.

Дослідження довжини слів

Результати значень розрахунку абсолютних і відносних частот представлені у вигляді таблиць 1–3.

Таблиця 1

Розподіл довжини слів твору Л.Дашвар «Село не люди. Добити свідка»

Довжина слова	Кількість		Довжина слова	Кількість	
	абсолютна	відносна (%)		абсолютна	відносна (%)
1	8993	8,468	14	482	0,458
2	15038	14,303	15	133	0,126
3	8454	8,041	16	55	0,052
4	13088	12,448	17	28	0,027
5	12743	12,12	18	30	0,029
6	11918	11,335	19	22	0,21
7	10867	10,336	20	6	0,006
8	8737	8,31	21	10	0,01
9	6114	5,815	22	5	0,005
10	3815	3,628	23	2	0,002
11	2515	2,392	24	2	0,002
12	1350	1,284	25	1	0,001
13	823	0,783	26	1	0,001

Таблиця 2

Розподіл довжини слів твору Л.Дашвар «Галя без голови»

Довжина слова	Кількість		Довжина слова	Кількість	
	абсолютна	відносна (%)		абсолютна	відносна (%)
1	2	3	4	5	6
1	9036	9,559	15	115	0,122
2	13280	14,049	16	65	0,069
3	7728	8,766	17	24	0,025

Закінчення табл. 2

1	2	3	4	5	6
4	12631	13,362	18	14	0,015
5	10720	11,341	19	10	0,011
6	10433	11,037	20	4	0,004
7	8898	9,413	21	2	0,002
8	7822	8,275	22	4	0,004
9	5464	5,78	23	2	0,002
10	3692	3,906	24	2	0,002
11	2298	2,431	25	2	0,002
12	1338	1,415	26	2	0,002
13	664	0,702	29	1	0,001
14	274	0,29	42	1	0,001

Таблиця 3

Розподіл довжини слів твору Л.Дашвар «Мати все»

Довжина слова	Кількість		Довжина слова	Кількість	
	абсолютна	відносна (%)		абсолютна	відносна (%)
1	5374	9,095	15	109	0,184
2	8612	14,575	16	85	0,144
3	4657	7,882	17	44	0,074
4	7481	12,661	18	30	0,051
5	6640	11,238	19	24	0,041
6	7117	12,045	20	10	0,017
7	5512	9,329	21	3	0,005
8	4691	7,939	22	5	0,008
9	3344	5,66	23	2	0,003
10	2368	4,008	24	1	0,002
11	1580	2,674	25	1	0,002
12	760	1,286	26	2	0,003
13	417	0,706	29	2	0,003
13	214	0,362	35	1	0,002

Аналізуючи отримані результати, можна стверджувати, що у творах найчастіше використовуються слова довжиною 2, 4, 5, 6 букв («Село не люди. Добити свідка»: 2 – 15038, 14,303 %; 4 – 13088, 12,448 %; 5 – 12743, 12,12 %; 6 – 11918, 11,335 %. «Галя без голови»: 2 – 13280, 14,049 %; 4 – 12631, 13,362 %; 5 – 10720, 11,341 %; 6 – 10433, 11,037 %. «Мати все»: 2 – 8612, 14,575 %; 4 – 7481, 12,448 %; 5 – 6640, 11,238 %; 6 – 7117, 12,042 %).

До середньої групи належать слова, довжина яких становить 1, 3, 7, 8, 9 букв («Село не люди. Добити свідка»: 7 – 10867, 10,336 %; 1 – 8993, 8,468 %; 8 – 8737, 8,31 %; 3 – 8454, 8,041 %; 9 – 6114, 5,815 %. «Галя без голови»: 1 – 9036, 9,559 %; 7 – 8898, 9,413 %; 8 – 7822, 8,275 %; 3 – 7728, 8,766 %; 9 – 5464, 5,78 %. «Мати все»: 7 – 5512, 9,329%; 1 – 5374, 9,095 %; 8 – 4691, 7,939 %; 3 – 4657, 7,882%; 9 – 3344, 5,66 %).

Частка всіх інших слів становить не більше 4 %.

Як видно з результатів проведеного дослідження, в усіх трьох творах частки слів за довжиною, що найчастіше використовуються та належать до середньої групи, практично збігаються.

Результати статистичних розрахунків довжини слів отриманих засобами мови Python:

Статистичні показники твору Л.Дашвар «Село не люди. Добити свідка»:

середнє значення = 5.28;  
середнє квадратичне = 2.401;  
міра коливання середньої частоти = 0.002;  
стандартна похибка = 0.137;  
відносна похибка = 0.001.

Статистичні показники твору Л.Дашвар «Галя без голови»:

середнє значення = 5.21;  
середнє квадратичне = 2.589;

міра коливання середньої частоти = 0.002;  
 стандартна похибка = 0.136;  
 відносна похибка = 0.001.

Статистичні показники твору Л.Дашвар «Мати все»:  
 середнє значення = 5.29;  
 середнє квадратичне = 2.795;  
 міра коливання середньої частоти = 0.004;  
 стандартна похибка = 0.137;  
 Відносна похибка = 0.001.

Отримані показники доцільно порівняти з результатами оцінки стилів української мови на прикладі інших письменників [6]. У таблиці твори позначені як: 1 – «Село не люди. Вбити свідка», 2 – «Галя без голови», 3 – «Мати все».

Таблиця 4

Порівняльна таблиця довжини слова у творах українських авторів

Автор	Середнє значення	Середнє квадратичне відхилення	Відносна похибка
А.Головко	4,74	0,10	0,34
О.Гончар	5,41	0,07	0,021
О.Довженко	4,73	0,08	0,027
П.Панч	5,28	0,29	0,091
М.Стельмах	5,30	0,16	0,048
Ю.Яновський	5,06	0,13	0,041
Р.Іваничук, новели	5,06	2,83	0,000
Р.Іваничук, романи	5,25	2,93	0,000
Л.Дашвар 1	5,28	2,4	0,001
Л.Дашвар 2	5,21	2,59	0,001
Л.Дашвар 3	5,29	2,8	0,001

Середня довжина слова творів Л.Дашвар найближча до показників П.Панча, М.Стельмаха та романів Р.Іваничука. Щодо середнього квадратичного відхилення та відносної похибки, то ці показники близькі до значень новел та романів Р.Іваничука. На думку авторів, це пов'язано з близькістю років діяльності цих авторів. Результати порівняння показують, що статистичні показники творчості Л.Дашвар цілком узгоджуються з відповідними показниками творчості українських письменників. Показово, що значення показників творів письменниці також близькі, а значення відносної похибки загалом однакові. Це може бути використано як один з аргументів під час визначення авторства тексту.

#### Дослідження довжини речень

Розподіл довжини речень у творах Л.Дашвар.

Розподіл речень у творі «Село не люди. Добити свідка»:

кількість речень у творі = 5399;

абсолютне значення кількості речень відносно кількості слів:

{1: 87, 2: 173, 3: 211, 4: 234, 5: 233, 6: 207, 7: 231, 8: 181, 9: 203, 10: 209, 11: 193, 12: 164, 13: 175, 14: 159, 15: 169, 16: 139, 17: 127, 18: 118, 19: 114, 20: 117, 21: 94, 22: 108, 23: 83, 24: 88, 25: 73, 26: 74, 27: 89, 28: 76, 29: 65, 30: 63, 31: 63, 32: 53, 33: 52, 34: 50, 35: 39, 36: 42, 37: 31, 38: 44, 39: 46, 40: 32, 41: 36, 42: 31, 43: 19, 44: 46, 45: 26, 46: 19, 47: 14, 48: 25, 49: 27, 50: 18, 51: 21, 52: 12, 53: 22, 54: 24, 55: 22, 56: 15, 57: 14, 58: 10, 59: 15, >=60: 274}

відносне значення кількості речень відносно кількості слів (%):

{1: 1.611, 2: 3.204, 3: 3.908, 4: 4.334, 5: 4.316, 6: 3.834, 7: 4.279, 8: 3.352, 9: 3.76, 10: 3.871, 11: 3.575, 12: 3.038, 13: 3.241, 14: 2.945, 15: 3.13, 16: 2.575, 17: 2.352, 18: 2.186, 19: 2.112, 20: 2.167, 21: 1.741, 22: 2.0, 23: 1.537, 24: 1.63, 25: 1.352, 26: 1.371, 27: 1.648, 28: 1.408, 29: 1.204, 30: 1.167, 31: 1.167, 32: 0.982, 33: 0.963, 34: 0.926, 35: 0.722, 36: 0.778, 37: 0.574, 38: 0.815, 39: 0.852, 40: 0.593, 41: 0.667, 42: 0.574, 43: 0.352, 44: 0.852, 45: 0.482, 46: 0.352, 47: 0.259, 48: 0.463, 49: 0.5, 50: 0.333, 51: 0.389, 52: 0.222, 53: 0.407, 54: 0.445, 55: 0.407, 56: 0.278, 57: 0.259, 58: 0.185, 59: 0.278, >=60: 0.05 }

Середнє значення кількості слів у реченні = 20.953

Розподіл речень у творі «Галя без голови»:

кількість речень у творі = 3805;

абсолютне значення кількості речень відносно кількості слів:

{1: 43, 2: 62, 3: 94, 4: 114, 5: 117, 6: 118, 7: 116, 8: 122, 9: 134, 10: 110, 11: 106, 12: 100, 13: 103, 14: 114, 15: 100, 16: 83, 17: 87, 18: 100, 19: 90, 20: 89, 21: 68, 22: 74, 23: 80, 24: 66, 25: 68, 26: 61, 27: 65, 28: 56, 29: 58, 30: 58, 31: 49, 32: 39, 33: 41, 34: 43, 35: 46, 36: 40, 37: 47, 38: 40, 39: 32, 40: 35, 41: 39, 42: 30, 43: 32, 44: 18, 45: 27, 46: 19, 47: 25, 48: 17, 49: 25, 50: 22, 51: 14, 52: 17, 53: 13, 54: 20, 55: 18, 56: 17, 57: 13, 58: 18, 59: 12, >=60: 341}

відносне значення кількості речень відносно кількості слів:

{1: 1.13, 2: 1.629, 3: 2.47, 4: 2.996, 5: 3.075, 6: 3.101, 7: 3.049, 8: 3.206, 9: 3.522, 10: 2.891, 11: 2.786, 12: 2.628, 13: 2.707, 14: 2.996, 15: 2.628, 16: 2.181, 17: 2.286, 18: 2.628, 19: 2.365, 20: 2.339, 21: 1.787, 22: 1.945, 23: 2.102, 24: 1.735, 25: 1.787, 26: 1.603, 27: 1.708, 28: 1.472, 29: 1.524, 30: 1.524, 31: 1.288, 32: 1.025, 33: 1.078, 34: 1.13, 35: 1.209, 36: 1.051, 37: 1.235, 38: 1.051, 39: 0.841, 40: 0.92, 41: 1.025, 42: 0.788, 43: 0.841, 44: 0.473, 45: 0.71, 46: 0.499, 47: 0.657, 48: 0.447, 49: 0.657, 50: 0.578, 51: 0.368, 52: 0.447, 53: 0.342, 54: 0.526, 55: 0.473, 56: 0.447, 57: 0.342, 58: 0.473, 59: 0.315, >=60: 0.089}

Середнє значення кількості слів у реченні = 26.512

Розподіл речень у творі «Мати все»:

кількість речень у творі = 4012;

абсолютне значення кількості речень відносно кількості слів:

{1: 132, 2: 196, 3: 249, 4: 257, 5: 213, 6: 217, 7: 220, 8: 191, 9: 166, 10: 135, 11: 108, 12: 146, 13: 117, 14: 120, 15: 127, 16: 89, 17: 77, 18: 94, 19: 68, 20: 91, 21: 86, 22: 67, 23: 58, 24: 46, 25: 55, 26: 57, 27: 46, 28: 44, 29: 26, 30: 29, 31: 38, 32: 28, 33: 25, 34: 24, 35: 18, 36: 17, 37: 19, 38: 23, 39: 21, 40: 7, 41: 19, 42: 14, 43: 20, 44: 11, 45: 21, 46: 18, 47: 6, 48: 6, 49: 9, 50: 10, 51: 8, 52: 5, 53: 10, 54: 9, 55: 5, 56: 5, 57: 9, 58: 5, 59: 5, >=60: 70}

відносне значення кількості речень відносно кількості слів:

{1: 3.29, 2: 4.885, 3: 6.206, 4: 6.406, 5: 5.309, 6: 5.409, 7: 5.484, 8: 4.761, 9: 4.138, 10: 3.365, 11: 2.692, 12: 3.639, 13: 2.916, 14: 2.991, 15: 3.166, 16: 2.218, 17: 1.919, 18: 2.343, 19: 1.695, 20: 2.268, 21: 2.144, 22: 1.67, 23: 1.446, 24: 1.147, 25: 1.371, 26: 1.421, 27: 1.147, 28: 1.097, 29: 0.648, 30: 0.723, 31: 0.947, 32: 0.698, 33: 0.623, 34: 0.598, 35: 0.449, 36: 0.424, 37: 0.474, 38: 0.573, 39: 0.523, 40: 0.174, 41: 0.474, 42: 0.349, 43: 0.499, 44: 0.274, 45: 0.523, 46: 0.449, 47: 0.15, 48: 0.15, 49: 0.224, 50: 0.249, 51: 0.199, 52: 0.125, 53: 0.249, 54: 0.224, 55: 0.125, 56: 0.125, 57: 0.224, 58: 0.125, 59: 0.125, >=60: 0.018}

Середнє значення кількості слів у реченні = 15.332

Результати дослідження показують, що у творах Л.Дашвар найчастіше зустрічаються речення, які мають від 11 до 15 слів. Середні значення кількості слів у реченні дозволяють стверджувати, що авторка у своїх текстах часто використовує досить довгі речення. Ця особливість може бути визначена як авторська риса і використовуватися для визначення авторства та стилю.

#### Аналіз використання різних типів речень у творах Л.Дашвар (табл. 4)

Дашвар Л. «Село не люди. Добити свідка»:

кількість окличних речень = 4383;  
кількість питальних речень = 2699;  
кількість речень розповідних = 8817;  
кількість речень = 15899;  
кількість речень з діалогу = 7977;  
кількість знаків "?! " = 448;  
кількість речень = за списком 8818.

Дашвар Л. «Галя без голови»:

кількість окличних речень = 3967;  
кількість питальних речень = 3177;  
кількість речень розповідних = 6333;  
кількість речень = 13477;  
кількість речень з діалогу = 6341;  
кількість знаків "?! " = 474.

Дашвар Л. «Мати все»:

кількість окличних речень = 1581;  
кількість питальних речень = 1388;  
кількість речень розповідних = 5738;  
кількість речень = 8707;  
кількість речень з діалогу = 4292;  
кількість знаків "?! " = 153.

Порівняльна таблиця типів речень у творах Л.Дашвар

Назва твору	Окличні		Питальні		Розповідні	
	абсолютне значення	відносне значення, %	абсолютне значення	відносне значення, %	абсолютне значення	відносне значення, %
«Село не люди. Добити свідка»	4383	27,67	2699	16,98	8817	55,00
«Галя без голови»	3967	29,44	3177	23,58	6333	47,00
«Мати все»	1581	18,16	1388	15,95	5738	65,90

Твори автора наповнені діалогами. Так у творі «Село не люди. Добити свідка» кількість речень з діалогу становить 50,18 %, «Галя без голови» – 47,05 %, «Мати все» – 49,30 %. Письменниця у своїх творах для передачі складної інтонації питання-оклику використовує комбінацію символів у вигляді «?!». Їх кількість становить у творі «Село не люди. Добити свідка» – 448 (2,8 %), «Галя без голови» – 474 (3,5 %), «Мати все» – 153 (1,7 %). Отримані показники дозволяють стверджувати про підвищену емоційність творів Л.Дашвар і можуть розглядатись як одні з основних ознак її авторства та стилю.

**Висновки та перспективи подальших досліджень.** Для творів Л.Дашвар, що ілюструють виконані дослідження, можна стверджувати про відповідність отриманих результатів показникам творчості видатних українських письменників. Відмінними ознаками текстів письменниці є переважне вживання слів розміром 2–6 букв, довгих (до 20 слів) речень, наявність великої кількості діалогів (близько 50 %), речень з емоційним забарвленням та складною інтонацією. У перспективі доцільно продовжити квантитативні дослідження україномовних текстів, збільшуючи кількість авторів, кількість творів, впроваджуючи нові методи та технології.

#### Список використаної літератури:

1. *Альтман Г.* Мода та істина в лінгвістиці / *Г.Альтман* // Проблеми квантитативної лінгвістики. – Чернівці : Рута, 2005. – 109 с.
2. *Бук С.Н.* Основи статистичної лінгвістики : навч.-метод. посіб. / *С.Н. Бук* ; відп. ред. проф. *Ф.С. Бацевич*. – Л. : Видавничий центр ЛНУ ім. І.Франка, 2008. – 124 с.
3. *Бук С.Н.* Статистичні характеристики лексики основних функціональних стилів української мови: спроба порівняння / *С.Н. Бук* // Лексикографічний бюлетень. – 2016. – № 13. – С. 166–170.
4. *Кригін М.Ю.* Дослідження інформаційно-статистичних властивостей українського тексту / *М.Ю. Кригін* // ІПММС НАНУ. – 2000. – № 1. – С. 120–127.
5. *Крупа М.* Лінгвістичний аналіз тексту / *М.Крупа*. – Тернопіль : Підручники і посібники, 2005. – 416 с.
6. *Кульчицький І.М.* Дослідження довжини речення та слова у творах Романа Іванчука / *І.М. Кульчицький* // Вісник Національного університету «Львівська політехніка». Серія : Інформаційні системи та мережі. – 2017. – № 872. – С. 139–148.
7. *Кульчицький І.М.* Окремі аспекти квантитативних досліджень української мови / *І.М. Кульчицький* // Україна модерна. – 2019. – № 27. – С. 73–96. DOI: 10.3138/ukrainamoderna.27.073.
8. *Перебийніс В.І.* Математична лінгвістика : навчальний посібник / *В.І. Перебийніс*. – Київ : Вид. центр КНЛУ, 2014. – 125 с.
9. *Перебийніс В.І.* Статистичні методи для лінгвістів : посібник / *В.І. Перебийніс*. – Вид. 2, випр. і допов. – Вінниця : Нова Книга, 2013. – 176 с.
10. *Де Сосюр Ф.* Курс загальної лінгвістики / *Ф. де Сосюр* ; пер. з фр. *А.Корнійчук, К.Тищенко*. – К. : Основи, 1998. – 324 с.
11. *Штефан С.В.* Статистичні методи досліджень : тексти лекцій для студентів Інституту журналістики / *С.В. Штефан* [Електронний ресурс]. – Режим доступу : <http://journalib.univ.kiev.ua/navch/StatMetodyDoslid.pdf>.
12. *Шухова О.О.* Квантитативні характеристики порядку слів у сучасній українській розмовній мові / *О.О. Шухова* // Лінгвістичні дослідження. – 2020. – Вип. 52. – С. 3–11 [Електронний ресурс]. – Режим доступу : <https://dspace.hnpu.edu.ua/server/api/core/bitstreams/eb292d05-d4da-41cd-85a9-efd06e1f7552/content>.

#### References:

1. Altman, G. (2005), «Moda ta istyna v lnhvistytsi», *Problemy kvantytatyvnoi lnhvistyky*, Ruta, Chernivtsi, 109 p.
2. Buk, S.N. (2008), *Osnovy statystychnoi lingvistyky, navch.-metod. posib.*, in Batsevych, F.S. (ed.), *Vydavnychiy tsentr LNU im. I.Franka, L.*, 124 p.
3. Buk, S.N. (2016), «Statystychni kharakterystyky leksyky osnovnykh funktsionalnykh styliv ukrainskoi movy: spoba porivniannia», *Leksykohrafichnyi biuletyn*, No. 13, pp. 166–170.
4. Kryhin, M.Yu. (2000), «Doslidzhennia informatsiino-statystychnykh vlastyvostei ukrainskoho tekstu», *IPMMS NANU*, No. 1, pp. 120–127.
5. Krupa, M. (2005), *Lnhvistychnyi analiz tekstu*, Pidruchnyky i posibnyky, Ternopil, 416 p.

6. Kulchytskyi, I.M. (2017), «Doslidzhennia dovezhyny rechennia ta slova u tvorakh Romana Ivanychuka», *Visnyk Natsionalnoho universytetu «Lvivska politekhnika»*. Seriya. *Informatsiini systemy ta merezhi*, Vydavnytstvo Lvivskoi politekhniki, Lviv, No. 872, pp. 139–148.
7. Kulchytskyi, I.M. (2019), «Okremi aspekty kvantytatyvnykh doslidzhen ukraïnskoi movy», *Ukraina moderna*, No. 27, pp. 73–96, doi: 10.3138/ukrainamoderna.27.073.
8. Perebyinis, B.I. (2014), *Matematychna linhvistyka*, navchalnyi posibnyk, Vyd. tseñtr KNLU, Kyiv, 125 p.
9. Perebyinis, V.I. (2013), *Statystychni metody dlia linhvistiv*, posibnyk, vyd. 2, vypr. i dopov., Nova Knyha, Vinnytsia, 176 p.
10. De Sosiur, F. (1998), *Kurs zahalnoi lingvistyky*, Translated by fr. Korniichuk, A. and Tyshchenko, K., Osnovy, K., 324 p.
11. Shtefan, S.V., *Statystychni metody doslidzhen*, teksty leksii dlia studentiv Instytutu zhurnalistyky, [Online], available at: <http://journlib.univ.kiev.ua/navch/StatMetodyDoslid.pdf>
12. Shukhova, O.O. (2020), «Kvantytatyvni kharakterystyky poriadku sliv u suchasniï ukraïnskiiï rozmovniï movi», *Linhvistychni doslidzhennia*, Issue 52, pp. 3–11, [Online], available at: <https://dspace.hnpu.edu.ua/server/api/core/bitstreams/eb292d05-d4da-41cd-85a9-efd06e1f7552/content>

**Окунькова Оксана Олексіївна** – старший викладач кафедри комп'ютерної інженерії та кібербезпеки Державного університету «Житомирська політехніка».

<https://orcid.org/0009-0004-0093-0694>.

Наукові інтереси:

- інформаційні та комп'ютерні технології в освіті.

**Варганова Діна Олександрівна** – старший викладач кафедри комп'ютерних технологій у медицині та телекомунікаціях Державного університету «Житомирська політехніка».

<https://orcid.org/0000-0002-8712-027X>.

Наукові інтереси:

- використання інформаційних технологій в освіті та науці.

**Дячук Ольга Юріївна** – старший викладач кафедри комп'ютерної інженерії та кібербезпеки Державного університету «Житомирська політехніка».

<https://orcid.org/0000-0002-6996-4700>.

Наукові інтереси:

- комп'ютерні мережі;
- кібербезпека;
- архітектура комп'ютера.

**Колощук Марія Сергіївна** – асистент кафедри комп'ютерної інженерії та кібербезпеки Державного університету «Житомирська політехніка».

<https://orcid.org/0009-0001-5825-2054>.

Наукові інтереси:

- комп'ютерні мережі;
- кібербезпека;
- архітектура комп'ютера.

**Okunkova O.O., Varhanova D.O., Dyachuk O.Yu., Koloshchuk M.S.**

#### **Quantitative research of Luco Dashvar's works**

Quantitative linguistics is a powerful tool that allows us to study language using mathematical methods and statistical analysis. These methods allow transforming qualitative linguistic data into quantitative ones, which makes it possible to conduct a more objective and in-depth analysis of textual data.

The article uses descriptive statistics methods such as mean value, standard deviation, measure of fluctuation of the average frequency, standard and relative error. To describe the main characteristics of the texts, the article uses methods of calculating the frequency of words, length and type of sentences, saturation of the text with dialogues, etc.

The possibility of using mathematical statistics methods in linguistics, the main directions of their application and implementation of the obtained results for the study of textual information are described. For example, determining the authorship of a text and analysing emotional colouring.

The author substantiates the relevance of conducting new research on existing and new collections of texts, which makes it possible to both confirm the previously identified statistical parameters and regularities of the text structure and identify new structural features of the modern Ukrainian language. The research presented in the article is the basis of the statistical typology of texts and languages.

The Python programming language was used to automate the calculation of the analysis results. Python is one of the most popular tools for text mining due to its flexibility and large number of libraries. Python can be used to automate many routine tasks related to natural language processing. The quantitative indicators are obtained as a result of research on the works of the contemporary Ukrainian writer L.Dashvar.

**Keywords:** statistics; frequency; variation series; absolute value; relative value; quantitative research.

Стаття надійшла до редакції 26.09.2024.