Issue 18, part 2, 2025

https://doi.org/10.23939/sisn2025.18.2.192

УДК 81.33

СТАТИСТИЧНИЙ ПРОФІЛЬ РОМАНУ Є. КУЗНЄЦОВОЇ «ДРАБИНА»

Олена Дзюба
Національний університет «Львівська політехніка»,
кафедра прикладної лінгвістики, Львів, Україна
E-mail: olena.dziuba.mflpl.2024@lpnu.ua, ORCID: 0009-0006-1429-146

У статті представлено результати комплексного лінгвостатистичного аналізу роману Є. Кузнєцової «Драбина» та його перекладу польською мовою. Дослідження проводилося за допомогою корпусного менеджера AntConc для визначення статистичних характеристик тексту. Цей програмний засіб використано для опрацювання корпусу, а саме створення частотних словників. Після формування частотних словників, у середовищі MS Excel здійснено лематизацію та визначення частин мови. Встановлено, що обсяг тексту оригіналу та перекладу практично ідентичний, однак оригінал вирізняється вищим багатством словника, більшою кількістю hapax legomena та меншою стереотипністю лексики порівняно з перекладом. Обчислено показник Юла, який відображає різноманітність словника, індекси винятковості, стереотипності та концентрації словника, ентропію, редунданцію тексту та інші показники. Також було застосовано кількісні алгоритми для визначення автоматизованого індексу читабельності (ARI), що свідчить про міру складності сприйняття тексту. Виявлено, що ARI у перекладі вищий (14 проти 12), що може свідчити про його дещо більшу складність для читання. Аналіз морфолого-статистичних характеристик підтверджує, що у словнику обох текстів домінують іменники та дієслова, тоді як у тексті домінують службові частини мови. Службові слова, маючи низький відсоток у словнику, активно функціонують у структурі мовлення, забезпечуючи зв'язність та синтаксичну організацію. Збереження психолінгвістичних статистичних параметрів (середній розмір речення, коефіцієнт дієслівності та логічної зв'язності) свідчить про адекватне відтворення авторського стилю в перекладі. Результати надають цінні відомості про кількісні характеристики художніх текстів і підкреслюють потенціал паралельних корпусів для порівняльних стилістичних, перекладознавчих і психолінгвістичних досліджень, пропонуючи надійну методологічну основу для подальшого вивчення взаємодії між лексичним вибором, синтаксичними моделями та психолінгвістичними параметрами в художньому перекладі.

Ключові слова – лінгвостатистика, статистичний профіль, ідіостиль, частотний словник, переклад

Постановка проблеми

Із розвитком комп'ютерних технологій значно розширилися межі лінгвістичних та перекладознавчих досліджень. Корпусна лінгвістика уможливила статистичний аналіз текстів, що відкрило нові шляхи для дослідження літератури і мови. Корпуси текстів слугують основою для статистичного аналізу, надаючи джерельну базу для теоретичних та практичних застосувань у лінгвістиці. Саме на основі корпусних даних реалізуються лінгвостатистичні підходи, що дають змогу виявляти закономірності у вживанні мовних одиниць, аналізувати частотність та встановлювати стилістичні особливості текстів.

Кількісний опис лексичного та морфологічного складу тексту дозволяє виявити не лише загальні закономірності функціонування мови, але й специфічні риси авторського стилю (ідіолекту), які важко встановити іншими методами аналізу (Buk, 2021).

Хоча статистичні методи є усталеними для аналізу ідіостилю та перекладу, статистичний профіль роману сучасної української письменниці Є. Кузнєцової «Драбина» (2023) та його польського перекладу потребує окремої розвідки. Зважаючи на відсутність деталізованого квантитативного аналізу цього твору, існує необхідність у застосуванні комплексу лінгвостатистичних показників для об'єктивної характеристики його лексичного складу та частиномовного розподілу.

Дослідження статистичного профілю роману, що передбачає виявлення та опис кількісних характеристик та закономірностей на різних рівнях (лексичному, синтаксичному, рівні всього тексту), належить до проблем, які можна розв'язати за допомогою корпусних технологій та лінгвостатистики.

Аналіз останніх досліджень та публікацій

Статистична лінгвістика є міждисциплінарною наукою, що вивчає кількісні характеристики та закономірності мови. Цей науковий напрям активно розвивався на межі XX–XXI ст. і має об'єктивні підстави для застосування у мовознавстві, оскільки дає змогу пізнати глибинні закони, які лежать в основі мови та мовлення.

Використання статистичних методів у дослідженнях зумовлене властивими мові кількісними ознаками, оскільки мова є складною ймовірнісною системою дискретних одиниць, яким притаманні як якісні, так і кількісні характеристики (Buk, 2021).

Застосування до вивчення мови математичного інструментарію із комп'ютерними технологіями сприяє переосмисленню наукового статусу лінгвістичної дисципліни. Результати, виявлені методами статистичної лінгвістики, плідно застосовують у багатьох сферах, зокрема: стилеметрії, літературознавстві, дослідженні ідіостилю письменника та авторській лексикографії. Цією проблематикою займаються безліч вітчизняних та зарубіжних учених.

Darchuk (1976) аналізувала статистичні характеристики лексики як відображення структури тексту та розробила підходи до кількісного опису мовних одиниць і механізмів їхнього функціонування.

Kamińska-Szmaj (1990) вивчала лексичні відмінності між функційними стилями польської мови за допомогою статистичних методів. Її праці показують, як статистичні параметри допомагають виявляти функційно-стильові особливості текстів і формують основу для автоматичної класифікації стилів.

Perebyinis (2002) систематизувала й адаптувала методи кількісної лінгвістики для аналізу українських текстів і показала, як статистичні дані можуть відображати характеристики мовлення. Її дослідження показали, що різниця між статистичною структурою текстів слугує критерієм для виявлення відмінностей між стилями письменників, причому стилерозрізнювальна потужність найвища на лексичному та синтаксичному рівнях.

Buk (2021) застосовувала лінгвостатистичні показники для дослідження авторського ідіолекту І. Франка шляхом створення Корпусу текстів великої прози І. Франка (КТ ВПФ), на основі якого провела комплексну лінгвостатистичну параметризацію ідіолекту письменника. На базі цього корпусу було укладено Частотні словники (ЧС) окремих творів, виявлено розподіл частин мови ВПФ, а також обчислено кількісне співвідношення між різними елементами тексту.

Загалом, в українській науці успішність застосування квантитативних методів демонструють також праці інших вчених, таких як М. Білинський, Л. Гіков, А. Потапенко, В. Широков та інших. Актуальність лінгвостатистики підтверджується її верифікаційністю, що гарантує надійність висновків.

Формулювання цілі статті

Це дослідження ґрунтується на комплексному лінгвостатистичному аналізі корпусу, здійсненому за допомогою корпусного менеджера AntConc, і передбачає виконання таких завдань: визначення набору релевантних лінгвостатистичних показників для кількісної характеристики тексту, встановлення морфолого-статистичних характеристик та кількісних відношень між частинами мови в оригіналі та перекладі, обчислення психолінгвістичних статистичних параметрів текстів, здійснення порівняльного аналізу оригіналу та перекладу на основі отриманих кількісних даних.

Виклад основного матеріалу

Лінгвостатистичний аналіз дає змогу виявити закономірності розподілу мовних одиниць, визначити їхню частотність, взаємозв'язки та тенденції функціонування у різних типах текстів. Саме тому статистичні підходи стають необхідним інструментом у сучасному мовознавстві, особливо в поєднанні з корпусними технологіями.

Одним із ключових інструментів статистичної лінгвістики є текстовий корпус. Під корпусом текстів ми розуміємо структуровані та філологічно компетентні мовні дані, представлені в автоматизованому вигляді. Вони призначені для розв'язання конкретних лінгвістичних завдань. Текстовий корпус – це певним чином організоване електронне зібрання писемних і усних текстів довільної природної мови, якому притаманні обов'язкові ознаки і яке призначене для здійснення наукового дослідження мови (Demska-Kulchytska, 2005).

У межах цього дослідження корпусний підхід поєднано зі статистичними методами, що дало змогу кількісно описати мовні особливості тексту оригіналу й перекладу. Зокрема, було укладено частотні словники (ЧС) оригіналу та перекладу для подальшого квантитативного аналізу. Частотний словник відображає статистичну структуру тексту, надаючи кількісну інформацію про появу мовних одиниць. Для створення ЧС використано програму AntConc Е. Лоуренса, яка дає змогу опрацьовувати корпуси, зокрема, здійснювати пошук та підрахунок різних елементів тексту, аналізувати частотність та контекст вживання словоформ, словосполук та морфем, порівнювати вживаність словоформ у різних текстах тощо (Kotsuik & Prokip, 2016; Kapranov, 2021). Після формування частотних списків було здійснено лематизацію та визначення частин мови. Далі визначено ряд кількісних показників, які характеризують тексти оригіналу та перекладу (результати подано у табл. 1).

Таблиця 1

Статистична характеристика текстів оригіналу та перекладу

Показник	Оригінал	Переклад
Кількість слововживань	41 227	41 473
Кількість словоформ	10 839	10 597
Обсяг словника (кількість лем)	6242	5662
Багатство словника	0,15	0,14
Середня повторюваність слова у тексті	6,6	7,32
Показник Юла	3004,58	3586,92
Hapax legomena	3283	2842
Індекс винятковості тексту	0,08	0,07
Індекс винятковості словника	0,53	0,5
Високочастотна лексика у тексті	29 066	29 950
Високочастотна лексика у словнику	506	521
Індекс концентрації тексту	0,71	0,72
Індекс концентрації словника	0,08	0,09
Індекс концентрації Містріка	1,43	1,36
Індекс стереотипності словника	12,84	13,7
Ентропія	4,35	4,41
Редунданція тексту	0,5	0,49
Автоматизований індекс читабельності	12	14

Отримані результати дають змогу зіставити оригінал роману та його переклад за низкою кількісних параметрів.

Обсяг тексту (N) в оригіналі та перекладі практично ідентичний: 41 227 та 41 473 слововживань відповідно. Обсяг словника (W) та кількість словоформ (W_f) у перекладі дещо менші (5 662 леми та 10 597 словоформ) порівняно з оригіналом (6 242 і 10 839 відповідно).

Багатство словника (W/N) – це відношення кількості лем до кількості всіх слів у тексті. Цей показник є вищим в оригіналі (0,15 проти 0,14). Своєю чергою, середня повторюваність слова (N/W), що є величиною, оберненою до попередньої, вища в перекладі (7,32 проти 6,6), тобто перекладач частіше звертається до вже вжитих одиниць.

Показник Юла (K) використовують для оцінювання різноманітності та оригінальності словника у тексті. Більші значення K вказують на бідніший словниковий запас. Згаданий показник обчислюють за формулою (Kamińska-Szmaj, 1990):

$$K = \frac{\sum f_i^2 W_i - N}{N^2} \cdot 10^4, \quad (1)$$

де: f_i – частота; W_i – кількість різних слів (лем), що трапляються з цією частотою; N – загальна кількість слововживань у тексті. У перекладі цей показник становить 3586,92, тоді як в оригіналі – 3004,58. Отже, словник перекладу менш оригінальний і більш стандартизований.

Подібна тенденція простежується й у кількості hapax legomena (W_1) – слів, що трапляються лише один раз: в оригіналі їх 3283 (що становить 7 % тексту), у перекладі – лише 2842. Відповідно, індекс винятковості тексту (W_1/N) та словника (W_1/W) є вищими в оригіналі (0,08 та 0,53 проти 0,07 і 0,5). Це свідчить, що авторський текст вирізняється більшою кількістю унікальних слів, тоді як переклад тяжіє до вживання вже усталеної лексики.

Водночас у перекладі переважає високочастотна лексика: як у тексті (29 950 проти 29 066), так і в словнику (521 проти 506). Це зумовлює зростання індексів концентрації: у тексті (0,72 проти 0,71) та у словнику (0,09 проти 0,08). Отже, у перекладі помітніший вплив закону переваги Дьюї, суть якого полягає в тому, що мова й мовлення базуються на обмеженому наборі одиниць, які вживаються дуже часто, тоді як більшість слів є низькочастотними (Buk, 2021).

Індекс концентрації Містріка вимірює, яку частку тексту займають слова, що вживаються більше одного разу. В оригіналі цей показник є дещо вищим (1,43), ніж у перекладі (1,36). Загалом такі показники узгоджуються з характерною для художньої прози тенденцією до слабшої концентрації лексики (Kamińska-Szmaj, 1990).

Індекс стереотипності словника ($I_{stereot.}$), який запропонував Ян Містрік, показує, наскільки словник тексту схильний до повторів і стандартності, на відміну від індексів різноманітності чи оригінальності:

$$I_{stereot.} = \frac{N - W_1}{W_{i>1}}. \quad (2)$$

Цей показник становить (12,84 для оригіналу і 13,7 для перекладу) і демонструє, що переклад містить більше повторюваних слів і менше унікальних одиниць, ніж авторський текст.

Ентропія (H_1) показує ступінь непередбачуваності появи слова в тексті. Чим вище ентропія, тим важче «вгадати», яке слово з'явиться далі. Її можна обчислити за формулою (Kamińska-Szmaj, 1990):

$$H_1 = \ln N - \frac{\sum W_i f_i \ln W_i f_i}{N}. \quad (3)$$

Показники ентропії практично збігаються: 4,35 в оригіналі та 4,41 у перекладі. Це означає, що рівень непередбачуваності в доборі слів у двох текстах майже однаковий.

Редунданція тексту (R_1) показує, наскільки слова в тексті залежні одне від одного (Kamińska-Szmaj, 1990):

$$R_1 = 1 - \frac{H_1}{\ln W}. \quad (4)$$

Показники редунданції також схожі (0,5 проти 0,49).

Автоматизований індекс читабельності (ARI) вказує на міру складності сприйняття тексту читачем; це співвідношення знаків у слові та кількості речень. Обчислюється за формулою: ARI = 4,71 * C / W + 0,5 * W / S – 21,43; де C – це кількість символів у тексті, W – кількість слів (пробілів) і S – кількість речень. Значення ARI можуть варіюватися від 6 до 18 і утворювати 12 рівнів складності. Нецілі результати завжди округлюються до найближчого більшого цілого числа (Nohovska, 2022). У перекладі ARI вищий (14 проти 12). Можна зробити висновок, що переклад є дещо складнішим для читання.

Також розраховано морфолого-статистичні характеристики – підрахунок кількості одиниць кожної частини мови у словнику та тексті. Аналіз їхнього кількісного співвідношення дає змогу виявити специфічні риси ідіолекту автора та особливості мовного стилю тексту. Як зазначає Н. Дарчук, тексти розповідного та описового типу зазвичай ґрунтуються на іменниковій основі. Іменники роблять опис статичним, лаконічним та стислим. Прикметники слугують для підкреслення ознак предметів та явищ. Натомість для текстів, що описують діяльність людини та різні процеси, характерна більша кількість дієслів (Darchuk, 1976).

Розподіл слововживань та лем за частинами мови подано у таблиці 2.

Таблиця 2

Розподіл слововживань та лем за частинами мови

Частина мови	Слововживання		Леми	
	Оригінал	Переклад	Оригінал	Переклад
Іменники	11 251 (27,29 %)	11 977 (28,88 %)	2372 (38,00 %)	2394 (42,28 %)
Дієслова	8222 (19,94 %)	8890 (21,44 %)	2116 (33,90 %)	1769 (31,24 %)
Займенники	5621 (13,63 %)	5305 (12,79 %)	52 (0,83 %)	44 (0,78 %)
Прикметники	2391 (5,80 %)	2430 (5,86 %)	991 (15,88 %)	903 (15,95 %)
Прислівники	4130 (10,02 %)	3188 (7,69 %)	528 (8,46 %)	385 (6,80 %)
Числівники	336 (0,81 %)	338 (0,81 %)	42 (0,67 %)	46 (0,81 %)
Сполучники	3355 (8,14 %)	3239 (7,81 %)	28 (0,45 %)	40 (0,71 %)
Прийменники	4268 (10,35 %)	4774 (11,51 %)	48 (0,77 %)	39 (0,69 %)
Частки	1463 (3,55 %)	1189 (2,87 %)	34 (0,54 %)	15 (0,26 %)
Вигуки	190 (0,46 %)	143 (0,34 %)	31 (0,50 %)	27 (0,48 %)
Всього	41 227	41 473	6242	5662

Аналіз частиномовного складу на рівні слововживань показує загальну подібність українського оригіналу та польського перекладу, проте з певними закономірними відмінностями. В обох текстах найбільшу частку становлять іменники та дієслова. У польському перекладі простежується більша частка дієслів і прийменників. Для українського тексту натомість властива більша кількість прислівників і часток, які традиційно виконують функцію вираження модальних та експресивних відтінків. Відмінності між підкорпусами незначні (переважно у межах 1–2 %).

Аналіз на рівні лем показав, що в обох словниках домінують іменники (38,00 % в оригіналі та 42,28 % у перекладі), і це зрозуміло, оскільки вони слугують для найменування об'єктів та подій. На другому місці – дієслова (33,90 % і 31,24 % відповідно), які що забезпечують динамічність тексту, позначаючи дії або стани. Іменники створюють відчуття мовної нерухомості, тоді як дієслова відповідають за динамічний розвиток подій. На третьому й четвертому місцях – прикметники та прислівники, функція яких – окреслювати ознаки предметів, явищ та понять: прикметники становлять близько 16 % в обох текстах, прислівники – 8,46 % в оригіналі та 6,80 % у перекладі. Частки, сполучники, числівники та вигуки представлені в межах 0,26–0,83 %. Загалом розподіл лем у перекладі зберігає основні тенденції оригіналу.

У словнику обох текстів домінують іменники та дієслова. Проте в тексті їхні частки є помітно нижчими. Службові частини мови, навпаки, характеризуються високою частотністю у тексті при низькому відсотку у словнику. Це підтверджує, що замкнені і відносно небагаті класи функційних слів активно працюють у структурі мовлення, забезпечуючи зв'язність і синтаксичну організацію тексту.

На основі цих даних можна обчислено кількісні відношення між частинами мови в оригіналі та перекладі (табл. 3).

Таблиця 3

Кількісні відношення між частинами мови у оригіналі та перекладі

Показник	Оригінал	Переклад
Ідекс епітетизації	4,71	4,93
Індекс дієслівних означень	0,5	0,36
Ступінь номінальності	1,37	1,35
Індекс прономіналізації	0,07	0,03
Індекс модальності	0,04	0,03
Індекс субстантивності	0,27	0,29

Індекс епітетизації відображає кількість іменників, що припадає на один прикметник: чим більший цей показник, тим менш насиченим є текст означеннями (Kamińska-Szmaj, 1988). В оригіналі цей індекс становить 4,71, а у перекладі – 4,93. Це означає, що в перекладі на один прикметник припадає 4,93 іменника, а в оригіналі – 4,71 іменника. Індекс дієслівних означень визначає кількість прислівників на одне дієслово. В оригіналі він дорівнює 0,5, у перекладі – 0,36. Ступінь номінальності, який показує, скільки іменників припадає на одне дієслово, в оригіналі становить 1,37, а в перекладі – 1,35.

Індекс прономіналізації – це відношення кількості особових займенників до обсягу словоживань тексту. Він відображає ступінь кореферентності (TextAttributor, 2024). В оригіналі цей показник становить 0,07, тоді як у перекладі – 0,03. Така різниця пояснюється особливостями польської мови: у ній особові займенники у ролі підмета часто опускаються, оскільки на особу виконавця дії вказує закінчення дієслова.

Індекс модальності визначається як відношення кількості часток до кількості слів у тексті. Він свідчить про ступінь емотивності й експресивності мовлення (TextAttributor, 2024). В оригіналі цей індекс дорівнює 0,04, у перекладі – 0,03. Це означає, що переклад характеризується дещо меншою кількістю часток.

Індекс субстантивності виражає відношення кількості іменників до загального обсягу словоживань тексту. Він показує насиченість тексту іменниками й одночасно відображає його статичність (TextAttributor, 2024). В оригіналі цей показник становить 0,27, а в перекладі – 0,29. Це означає, що у перекладі іменників відносно більше, ніж в оригіналі, і текст набуває більш статичного характеру.

Також пораховано ряд психолінгвістичних статистичних показників (табл. 4).

Таблиця 4

Психолінгвістичні показники у оригіналі та перекладі

Показник	Оригінал	Переклад
Середній розмір речення	17,06	16,93
Коефіцієнт Трейгера	3,44	3,66
Коефіцієнт опредметненої дії	0,73	0,74
Коефіцієнт дієслівності	0,20	0,21
Коефіцієнт логічної зв’язності	1,05	1,09
Коефіцієнт емболії	0,04	0,03

У досліджуваному матеріалі середній розмір речення в оригіналі становить 17,06 слова, а в перекладі – 16,93. Різниця є мінімальною, що дозволяє говорити про збереження ритму й темпу авторського мовлення.

Коефіцієнт Трейгера виражає відношення кількості дієслів до кількості прикметників у тексті та вказує на емоційну стабільність/нестабільність мовця (TextAttributor, 2024).. Цей показник становить 3,44 в оригіналі та 3,66 у перекладі. Отже, обидва тексти зберігають подібне співвідношення дієслів і прикметників, що свідчить про відсутність істотних відмінностей у рівні емоційної стабільності.

Коефіцієнт опредметненої дії визначається як відношення кількості дієслів до кількості іменників у тексті й характеризує рівень соціалізованості, синтаксичної завершеності/незавершеності висловлювання (TextAttributor, 2024). В оригіналі цей коефіцієнт дорівнює 0,73, а в перекладі – 0,74. Незначна різниця свідчить про збереження перекладом пропорцій між дієслівними та номінативними конструкціями, а також загальної синтаксичної організації тексту.

Коефіцієнт дієслівності (агресивності) – це відношення кількості вживань дієслів і дієслівних форм (дієприкметників і дієприслівників) до загальної кількості всіх слововживань (Zasiekin, 2012). Значення практично збігаються (0,20 в оригіналі та 0,21 у перекладі), що вказує на адекватне відтворення перекладачем динаміки й емоційної насиченості оригіналу.

Коефіцієнт логічної зв’язності – це співвідношення загальної кількості службових слів (сполучників і прийменників) і загальної кількості речень:

$$K_{\text{лог. зв.}} = \frac{N_{\text{служб. сл.}}}{3N_{\text{реч.}}}. \quad (5)$$

За величин у межах одиниці забезпечується гармонійний зв’язок службових слів і синтаксичних конструкцій (Zasiekin, 2012). У перекладі цей показник трохи вищий (1,09 проти 1,05 в оригіналі), що свідчить про дещо активніше використання перекладачем засобів синтаксичного зв’язку. Водночас показники залишаються в межах гармонійного співвідношення, тож сприйняття тексту не зазнає спотворень.

Коефіцієнт емболії («засміченості») мовлення – це відношення загальної кількості ембол (вигуки та частки) до загальної кількості слів. Коефіцієнт емболії відображає специфіку вербального інтелекту та емоційного стану мовця чи автора. Окрім того, цей показник може слугувати індикатором загальної культури мовлення (Zasiekin, 2012). В оригіналі коефіцієнт емболії становить 0,04, тоді як у перекладі – 0,03. Така різниця не є значною й свідчить про збереження перекладачем культури мовлення та стилістичних особливостей авторського тексту.

Отже, проведений аналіз демонструє, що текст перекладу загалом відповідає тексту оригіналу: спостережувані відмінності між показниками незначні й не впливають на якість сприйняття тексту реципієнтами.

Висновки

Проведений лінгвостатистичний аналіз дав змогу комплексно описати кількісні та структурні параметри роману Є. Кузнєцової «Драбина» та його польського перекладу, що відображають індивідуальні особливості авторського стилю. Основні показники свідчать, що текст перекладу загалом відповідає тексту оригіналу, а спостережувані відмінності між показниками є незначними й не впливають істотно на якість сприйняття тексту реципієнтами.

Незважаючи на збереження основних рис, у перекладі спостерігається тенденція до більшої статичності (вищий індекс субстантивності) та дещо зниженої експресивності (нижчий індекс модальності).

Перспективи подальших студій у цьому напрямі полягають у проведенні якісного аналізу лексико-семантичного дослідження домінант ідіостилю твору та їхній відповідників у перекладі. Створення дослідницького паралельного корпусу дає змогу оптимізувати подальші дослідження та відкриває перспективи для комплексного вивчення ідіостилю сучасних українських авторів і їхніх перекладів, забезпечуючи об’єктивні дані для контрастивної лінгвістики та перекладознавства.

Список літератури

Buk, S. N. (2021). *Corpus-lexicographic and linguistic-statistical measurements of Ivan Franko’s major prose works: dictionary and text*. (Doctoral dissertation, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine).

Darchuk, N. P. (1976). Statistical characteristics of vocabulary as a reflection of text structure. *Linguistic studies*, 97-102.

Demska–Kulchytska, O. (2005). *Fundamentals of the national corpus of the Ukrainian language*. Kyiv: Institute of Ukrainian language, National Academy of Sciences of Ukraine.

Kamińska-Szmaj, I. (1988). Parts of speech in the dictionary and text of five functional styles of written Polish (based on frequency dictionary data). *Bulletin of the Polish Linguistic Society*, 127-136.

Kamińska-Szmaj I. (1990). *Lexical differences between functional styles of written Polish: statistical analysis based on frequency dictionary data*. Wrocław: Wrocław University Press.

Kapranov Ya. V. (2021). AntConc Corpus Manager and its capabilities for determining the frequency of keywords in different languages. In *Knowledge engineering as a factor in intercultural cooperation between Ukraine, Japan, China, and South Korea: II International Scientific and Practical Conference* (pp. 100-102). Kyiv: KNLU Publishing Center.

Kotsuik L. M., & Prokip L. I. (2016). A corpus-based study of the terminology system of surgery. *Scientific Bulletin of the Lesya Ukrainka Eastern European National University. Series: Philological studies*, 6(331), 15-25.

Kuznietsova, Ye. (2023). *The Ladder*. Lviv: The Old Lion Publishing House.

Kuznietsowa, Ye. (2024). *The Ladder* (I. Boruszkowska, Trans.). Kraków: Znak Publishing House. (Original work published 2023)

Nohovska, S. H. (2022). Peculiarities of English language toning drinks advertising texts reproduction in Ukrainian: corpus-based experimental research. *Transcarpathian Philological Studies*, 22(2), 182-189. https://doi.org/10.32782/tps2663-4880/2022.22.2.33

Perebyinis, V. I. (2002). *Statistical methods for linguists.* Vinnytsia: Nova Knyha.

TextAttributor. (2024). *Methods*. Mova.info. Retrieved October 3, 2025, from http://ta.mova.info/methods

Zasiekin S. V. (2012). *Psycholinguistic universals of literary translation.* Lutsk: Lesya Ukrainka Volyn National University.

STATISTICAL PROFILE OF YEVHENIIA KUZNIETSOVA'S NOVEL DRABYNA

Olena Dziuba

Lviv Polytechnic National University,
Applied Linguistics Department, Lviv, Ukraine
E-mail: olena.dziuba.mflpl.2024@lpnu.ua, ORCID: 0009-0006-1429-146

The article presents the results of a comprehensive linguistic and statistical analysis of Yevheniia Kuznetsova's novel Drabyna (The Ladder) and its Polish translation. The study was conducted using the corpus manager AntConc to determine the statistical characteristics of the texts. This software tool was employed for corpus processing, namely for the compilation of frequency dictionaries. After generating the frequency lists, lemmatization and part-of-speech tagging were carried out in MS Excel. It was established that the volume of the original text and the translation is nearly identical; however, the original displays a higher degree of lexical richness, a greater number of hapax legomena, and lower lexical stereotypy compared to the translation. Yule's characteristic was calculated to reflect vocabulary diversity, along with indices of exclusiveness, stereotypy, and lexical concentration, as well as text entropy, redundancy, and other quantitative parameters. Quantitative algorithms were also applied to determine the Automated Readability Index (ARI), which indicates the degree of textual complexity. It was found that the ARI score is higher in the translation (14 versus 12), suggesting that the translated text is somewhat more difficult to read. The morpho-statistical analysis confirms that both texts' vocabularies are dominated by nouns and verbs, while function words prevail within the text structure. Despite their relatively low percentage in the vocabulary, function words play an active role in ensuring syntactic cohesion and textual connectivity. The preservation of psycholinguistic statistical parameters (such as average sentence length, verb density coefficient, and logical coherence index) indicates an adequate reproduction of the author's style in the translation. The results provide valuable insights into the quantitative characteristics of literary texts and highlight the potential of parallel corpora for comparative stylistic, translational, and psycholinguistic studies, offering a robust methodological framework for further research on the interaction between lexical choice, syntactic patterns, and psycholinguistic parameters in literary translation.

Keywords - statistical linguistics, statistical profile, idiolect, frequency dictionary, translation