

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Київський національний університет імені Тараса Шевченка
Навчально-науковий інститут філології
Кафедра української мови та прикладної лінгвістики

**СТИЛЕМЕТРІЯ ІДІОСТИЛІВ ОКСАНИ ЗАБУЖКО ТА СЕРГІЯ ЖАДАНА
(НА МАТЕРІАЛІ АВТОМАТИЧНО УКЛАДЕНИХ ЕЛЕКТРОННИХ
СЛОВНИКІВ ПРОЗОВИХ ТВОРІВ)**

Кваліфікаційна робота
на здобуття ОС «бакалавр»
студентки IV курсу
галузі знань 03 «Гуманітарні науки»,
спеціальність 035 «Філологія»
спеціалізації 035.10 «Прикладна
лінгвістика», ОПП «Прикладна
(комп'ютерна) лінгвістика
та англійська мова»
Юлії НЕСТЕРЕНКО
Науковий керівник:
к.філол.н., доц. кафедри
української мови
та прикладної лінгвістики
Оксана ЗУБАНЬ

«Допущено до захисту»
Протокол засідання кафедри
української мови та прикладної лінгвістики
протокол № 16 від «05» серпня 2026 року
Завідувач кафедри  (підпис)
к.філол.н., доц. **Сергій РІЗНИК**

ЗМІСТ

АНОТАЦІЯ	3
ВСТУП	5
РОЗДІЛ І.	8
ТЕОРЕТИЧНІ ЗАСАДИ ДОСЛІДЖЕННЯ ІДІОСТИЛІВ	8
1.1. Поняття ідіостилю в сучасному літературознавстві та мовознавстві	8
1.2. Ідіостиль у статистичному дослідженні: методи стилеметрії	10
1.3. Мовотворчість Оксани Забужко та Сергія Жадана у мовознавчих студіях	13
Висновки до розділу І	16
РОЗДІЛ ІІ.	18
АВТОМАТИЧНЕ УКЛАДАННЯ ЧАСТОТНИХ СЛОВНИКІВ ПРОЗОВИХ ТВОРІВ ОКСАНИ ЗАБУЖКО ТА СЕРГІЯ ЖАДАНА	18
2.1. Формування текстових вибірок: автоматична передобробка тексту	18
2.2. Електронні частотні словники: загальна характеристика	19
2.3. Визначення відносної похибки для статистичного дослідження	24
Висновки до розділу ІІ	26
РОЗДІЛ ІІІ.	28
МОРФОЛОГІЧНА СТАТИСТИЧНА МОДЕЛЬ СТИЛЕМЕТРИЧНОГО ДОСЛІДЖЕННЯ ІДІОСТИЛЮ ОКСАНИ ЗАБУЖКО ТА СЕРГІЯ ЖАДАНА	28
3.1. Полігони частот частин мови: порівняльний аспект	28
3.2. Зіставлення вибірок за смугою коливання середньої частоти частин мови	30
3.3. Стилеметричне індексування морфологічної структури текстів	33
Висновки до розділу ІІІ	35
ВИСНОВКИ	36
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	39
ДОДАТКИ	43
Додаток А. Текстовий корпус творів Оксани Забужко «Польові дослідження з українського сексу» та «Казка про калинову сопілку», використаний для стилеметричного дослідження.	43
Додаток Б. Текстовий корпус роману Сергія Жадана «Інтернат», використаний для статистичного аналізу.	43

Додаток В. База даних SQLite, що містить результати автоматичного аналізу текстового корпусу Оксани Забужко, зокрема частотні словники словоформ, лем та частин мови.	43
Додаток Г. База даних SQLite, що містить результати автоматичного аналізу текстового корпусу Сергія Жадана, зокрема частотні словники словоформ, лем та частин мови.	43
Додаток Ґ. Основний програмний код, розроблений для автоматичного укладання частотних словників, морфологічного аналізу текстів та формування баз даних.	43
Додаток Д. Програмний скрипт, розроблений мовою Python для автоматичного обчислення відносної похибки частотних показників частин мови, а також формування підсумкових таблиць результатів у форматі Microsoft Excel.	43
Додаток Е. Результати обчислення відносної похибки частотних показників частин мови для корпусів творів Оксани Забужко та Сергія Жадана.	43
Додаток Є. Програмний скрипт, розроблений мовою Python для автоматизованої побудови полігонів частот на основі варіаційних та інтервальних рядів частин мови.	43
Додаток Ж. Графічні матеріали, що містять побудовані полігони частот для різних частин мови та використовувалися під час стилеметричного аналізу.	44
Додаток З. Програмний код для обчислення довірчих інтервалів середньої частоти частин мови.	44
Додаток И. Графіки довірчих інтервалів середньої частоти частин мови, використані для зіставлення статистичних характеристик авторських корпусів.	44
Додаток І. Програмний скрипт, розроблений мовою Python для автоматизованого обчислення лінгвостатистичних та психолінгвістичних індексів текстів на основі морфологічного аналізу засобами бібліотеки Rymorphy3.	44
Додаток Ї. Результати автоматизованого обчислення лінгвостатистичних та психолінгвістичних індексів для корпусів творів Оксани Забужко та Сергія Жадана.	44

АНОТАЦІЯ

Дослідження, представлене у кваліфікаційній роботі, зосереджується на порівняльному аналізі ідіостилів Оксани Забужко та Сергія Жадана із застосуванням методів лінгвостатистики та комп'ютерного моделювання. Актуальність роботи визначається необхідністю впровадження об'єктивних підходів у стилеметричному аналізі сучасної художньої прози, що сприяє заміні традиційних описових методів аналізу точними кількісними характеристиками. Об'єктом дослідження українськомовні тексти. Приводом для аналізу стало вивчення морфологічних параметрів, які формують унікальну мовну ідентичність кожного автора.

Метою роботи є виконання стилеметричного аналізу ідіостилів Оксани Забужко та Сергія Жадана на базі морфолого-статистичної моделі тексту, визначеної автоматично. Для досягнення цієї мети було розроблено інструментарій на основі мови програмування Python, який забезпечує повний цикл обробки лінгвістичних даних: від етапів токенизації й лематизації до морфологічної розмітки (POS-тегування) та інтеграції результатів у бази даних SQLite. Наукова новизна роботи полягає у комплексному підході, що поєднує методи корпусної лінгвістики із математико-статистичними інструментами для перевірки репрезентативності вибірок. Особливо важливим є застосування порівняльного аналізу частотних розподілів частин мови із використанням довірчих інтервалів і побудовою полігонів частот, що вперше було реалізовано для ідіостилів Забужко та Жадана.

У ході дослідження встановлено, що ідіостиль Оксани Забужко характеризується значною долею абстрактної лексики та наявністю складних синтаксичних структур, що підтверджується розподілом дієслівних та сполучникових форм. Натомість мовна манера Сергія Жадана вирізняється динамічністю та схильністю до номінативного стилю, про що свідчать високі показники використання іменників і розмовної лексики. Обчислені стилеметричні індекси виявили статистично значущі відмінності в морфологічній будові текстів досліджуваних авторів. Розроблений у межах проекту програмний комплекс разом із базами даних становить практичну цінність для подальших досліджень у галузях автоматичної атрибуції текстів та розроблення авторських словників.

Ключові слова: ідіостиль, стилеметрія, частотний словник, комп'ютерна лінгвістика, Python, лінгвостатистика, Оксана Забужко, Сергій Жадан, корпус текстів.

ABSTRACT

The research presented in this qualification thesis focuses on a comparative analysis of the idiolects of Oksana Zabuzhko and Serhii Zhadan using methods of linguistic statistics and computer modeling. The relevance of the study is determined by the need to introduce objective approaches to the stylometric analysis of contemporary fiction, replacing traditional descriptive methods with precise quantitative characteristics. The object of the research is Ukrainian-language texts. The study is motivated by the investigation of morphological parameters that shape the unique linguistic identity of each author.

The aim of the thesis is to conduct a stylometric analysis of the idiolects of Oksana Zabuzhko and Serhii Zhadan based on an automatically generated morphological-statistical text model. To achieve this goal, a software toolkit was developed using the Python programming language, providing a complete cycle of linguistic data processing: from tokenization and lemmatization to morphological annotation (POS tagging) and integration of the results into SQLite databases. The scientific novelty of the study lies in its comprehensive approach, combining methods of corpus linguistics with mathematical and statistical tools for testing sample representativeness. Of particular importance is the application of a comparative analysis of part-of-speech frequency distributions using confidence intervals and frequency polygons, which has been implemented for the first time in the study of the idiolects of Zabuzhko and Zhadan.

The research findings indicate that Oksana Zabuzhko's idiolect is characterized by a significant proportion of abstract vocabulary and the presence of complex syntactic structures, as confirmed by the distribution of verbal and conjunctive forms. In contrast, Serhii Zhadan's linguistic style is distinguished by dynamism and a tendency toward a nominative style, evidenced by the high frequency of nouns and colloquial vocabulary. The calculated stylometric indices revealed statistically significant differences in the morphological structure of the texts produced by the authors under study. The software package developed within the project, together with the created databases, has practical value for further research in the fields of automatic text attribution and the development of author-specific dictionaries.

Keywords: idiolect, stylometry, frequency dictionary, computational linguistics, Python, linguistic statistics, Oksana Zabuzhko, Serhii Zhadan, text corpus.

ВСТУП

Актуальність дослідження. У сучасній лінгвістиці спостерігається зростання інтересу до аналізу ідіостилю як унікальної системи мовних особливостей авторів. Інтеграція методів комп'ютерної лінгвістики та стилеметрії відкриває нові можливості для об'єктивного і математично обґрунтованого аналізу текстів, надаючи змогу уникнути суб'єктивізму в оцінці стилістичних характеристик.

Дослідження індивідуальних стилів Оксани Забужко та Сергія Жадана становить важливу частину сучасної лінгвістичної науки, адже їхня творчість відображає ключові тенденції розвитку сучасної української літератури. Лінгвостилістичні особливості творів Оксани Забужко вивчали такі дослідники, як О. М. Капура й М. М. Максименко, які акцентували увагу на її лексичних, синтаксичних та концептуальних рисах. Аналогічно, творчість Сергія Жадана зацікавила В. С. Гриду, Г. І. Гримашевич й інших науковців, що аналізували специфіку його синтаксичних конструкцій, мовотворчість і художню образність.

Проте аналіз наукової літератури свідчить про відсутність у вітчизняному мовознавстві комплексних статистичних досліджень ідіостилів Оксани Забужко та Сергія Жадана. Зокрема, бракує праць, які б поєднували автоматичне створення частотних словників, морфологічне індексування та стилеметричне моделювання текстів. Усе це підкреслює актуальність обраної теми для бакалаврської роботи та аргументує необхідність проведення такого дослідження.

Стан дослідження проблеми. Теоретичну основу цього дослідження формують праці, присвячені ідіостилістиці (Л. Блумфільд, Д. Врайт, В. І. Волошук, Х. І. Дідух), лінгвостатистиці та складанню частотних словників (В. І. Перебийніс, Н. П. Дарчук), а також стилеметричному аналізу та модельним підходам до морфологічної структури тексту (С. Н. Бук, М. Едер, О. М. Зубань).

Мета дослідження. Здійснити стилеметричний аналіз ідіостилів Оксани Забужко та Сергія Жадана на базі морфолого-статистичної моделі тексту, визначеної автоматично.

Завдання дослідження:

1. Розробити теоретичну базу для вивчення ідіостилю та впровадження стилеметричних підходів у лінгвістику.
2. Побудувати корпуси текстів творів авторів, здійснити їх автоматичну предобробку і виконати морфологічну розмітку текстів за допомогою POS-тегування.
3. Розробити програмне забезпечення для автоматизованого створення частотних словників словоформ, лем та частин мови.

4. Обчислити відносну похибку статистичного дослідження для кожної частини мови.
5. Побудувати полігони частот кожної частини мови у двох текстових вибірках і здійснити порівняльний аналіз розподілу частот у текстах обох ідіостилів.
6. Побудувати смуги коливання абсолютних частот кожної частини мови в обох текстових вибірках і визначити довірчі інтервали розподілу абсолютних частот у зіставному порівнянні ідіостилів.
7. Автоматизувати обчислення статистичних морфологічних індексів в обох текстових вибірках, здійснити порівняння двох ідіостилів на базі сформованої морфологічної стилеметричної моделі.

Об'єкт дослідження: українськомовні тексти.

Предмет дослідження: морфологічна статистична структура тексту.

Методологічна основа дослідження базується на системно-структурному та лінгвостатистичному підходах, які дозволяють розглядати ідіостиль як індивідуальну сукупність мовних патернів автора. Ці патерни можна формалізувати та точно виміряти за допомогою цифрових методів.

Для вирішення поставлених завдань використовувалися такі **методи**: метод комп'ютерного моделювання мовних явищ, метод графічного унаочнення, статистичний метод, методика статистичного індексування. В автоматичному аналізі було використано такі бібліотеки: *re* для очищення тексту та роботи з регулярними виразами, *tokenize_uk* для токенизації українськомовних текстів, *pymorphy3* для морфологічного аналізу та визначення частин мови, *sqlite3* для створення й опрацювання баз даних, *collections* для підрахунку частот мовних одиниць, *pandas* для обробки та аналізу статистичних даних, *math* для виконання математичних розрахунків, пов'язаних із визначенням відносної похибки та довірчих інтервалів, а також *matplotlib* для побудови полігонів частот і графічного представлення результатів дослідження.

Матеріал дослідження: Основою дослідження стали прозові твори Оксани Забужко — «Польові дослідження з українського сексу» та «Казка про калинову сопілку», а також роман Сергія Жадана «Інтернат». Для виконання статистичного аналізу було створено електронні текстові корпуси загальним обсягом по 50 000 слововживань кожна.

Теоретична та практична значущість отриманих результатів: Теоретична цінність роботи полягає у вдосконаленні методології стилеметричного аналізу ідіостилю через інтеграцію лінгвостатистичних методів і автоматизованої обробки текстових даних. У практичному аспекті дослідження важливе тим, що його результатом стало розроблення програмного забезпечення для автоматичного створення частотних словників та морфологічного аналізу текстів українською мовою. Отримані напрацювання можуть слугувати основою для подальших

стилеметричних, лінгвостатистичних і корпусно-лінгвістичних досліджень, а також активно застосовуватись в освітньому процесі, зокрема у викладанні прикладної лінгвістики та технологій комп'ютерної обробки тексту.

Структура роботи: Кваліфікаційна робота включає вступ, три основні розділи, висновки, список використаних джерел і додаток з матеріалами дослідження. Загальний обсяг роботи становить 45 сторінок, а список використаних джерел містить 52 позиції.

РОЗДІЛ І.

ТЕОРЕТИЧНІ ЗАСАДИ ДОСЛІДЖЕННЯ ІДІОСТИЛІВ

Поняття ідіостилію відіграє фундаментальну роль у сучасному літературознавстві, адже воно репрезентує унікальний мовний почерк автора. Ідіостиль охоплює широкий спектр лексичних, синтаксичних, семантичних і стилістичних характеристик, які разом формують специфічну інтонацію автора та його індивідуальний спосіб вираження думок. Аналіз ідіостилію сприяє глибшому усвідомленню концептуальної структури тексту й індивідуального світосприйняття митця.

У межах сучасної лінгвістики дослідження ідіостилію поступово виходить за рамки традиційного описового аналізу, перетворюючись на багатовимірне вивчення зв'язків між мовою та мисленням. У процесі цього аналізу кожна текстова одиниця автора розглядається не лише як елемент художньої форми, а як відображення індивідуальної картини світу, у якій звичні мовні засоби набувають нових, суб'єктивних значень. Така багаторівнева природа ідіостилію породжує потребу в інтеграції класичних методів філологічного аналізу з сучасними методиками прикладної лінгвістики. Це дозволяє не лише виокремлювати глибинні структури індивідуального авторського «Я», але й виразно окреслювати якісні й кількісні аспекти організації тексту.

1.1. Поняття ідіостилію в сучасному літературознавстві та мовознавстві

У лінгвістичних студіях останніх десятиліть дедалі більше уваги приділяється вивченню мови не лише як абстрактної системи, а як засобу самовираження мовної особистості. У цьому контексті поняття індивідуального стилю постає одним із ключових інструментів аналізу авторської мови, що відкриває широкі можливості для дослідження літературного тексту. Категорія ідіостилію, попри свою багатозначність, дедалі чіткіше окреслюється як особливий мовний світ конкретного автора, що формується на перетині лінгвістичних, психологічних, культурних і творчих чинників [Гриценко 1984]. Як зазначають Ю. О. Калимон, І. М. Кульчицький та І. О. Ліхнякевич, ідіостиль є складною, структурно єдиною системою засобів словесного вираження, яка відображає творчу самотність митця та його мовну картину світу [Калимон 2023].

Варто зауважити, що підґрунтям для розуміння ідіостилію є поняття ідіолекту. Ще класик американського структуралізму Л. Блумфільд зазначав, що загальна мова спільноти є певною абстракцією, оскільки в реальності кожен індивід володіє власним, специфічним набором мовних звичок. Автор зазначає, що ці індивідуальні відмінності в мовленні зумовлені унікальним досвідом кожної людини, що фактично робить мову кожного окремого мовця неповторною одиницею аналізу [Блумфільд 1973]. Розвиваючи цю думку, М. Култхард та А. Джонсон зазначають, що

ідіолект функціонує як унікальний «лінгвістичний відбиток пальця», де специфічні комбінації лексичного вибору та граматичних патернів дозволяють ідентифікувати автора тексту з високим ступенем точності. [Култхард 2007]. У сучасній англomовній лінгвістиці ці ідеї отримують подальший розвиток: зокрема, М. Култхард підкреслює, що ідіолект визначається сукупністю повторюваних мовних звичок, які формують індивідуальний мовний профіль автора та забезпечують можливість його розпізнавання в тексті. [Култхард 2004]. Разом із тим сучасні дослідження також уточнюють зміст цього поняття: Д. Врайт розглядає ідіолект як цілісну мовну систему індивіда, що охоплює всі рівні мовної організації та формується під впливом особистого мовного досвіду [Врайт 2018].

Дослідниця В. І. Волошук звертає увагу на «мовну прозорість» терміна «ідіолект» та визначає його як систему мовних засобів індивіда, що формується на основі засвоєння мови і розвивається в процесі життєдіяльності. Авторка підкреслює, що індивідуальний «почерк» митця виявляється вже в самому факті вибіркoвості та перевазі певних лексичних чи морфологічних одиниць [Волошук, 2008, с. 5]. Це знаходить підтвердження у працях Х. І. Дідух, де ідіостиль розглядається як «система мовностилістичних засобів, що відображають унікальну творчу манеру певної мовної особистості». [Дідух, 2012].

Особливого значення набуває розгляд ідіостилю як способу побудови авторської моделі світу. О. В. Фадєєва зазначає, що ідіостиль є формою репрезентації концептуальної та мовної картин світу автора. Через систему авторських номінацій митець створює унікальний «віртуальний всесвіт», де мовні одиниці наділяються суб'єктивними смислами та трансформуються відповідно до художнього задуму [Фадєєва 2019]. У цьому контексті варто також врахувати результати сучасних корпусних досліджень: М. Барлоу доводить, що індивідуальні мовні патерни (лексичні та синтаксичні) є відносно стабільними та можуть розглядатися як основа ідіостилю автора [Барлоу 2000].

О. П. Костецька вказує на те, що індивідуальне мовлення автора є об'єктом, де перетинаються лінгвістичний аналіз емоційно-експресивних явищ та літературознавче осмислення авторського світобачення [Костецька, 2014, с. 196]. Сучасне мовознавство розглядає ідіостиль як багаторівневу структуру. Зокрема, О. Андріяш підкреслює роль лексичного рівня (використання оказіоналізмів, архаїзмів) у визначенні меж ідіостилю [Андріяш, 2011, с. 36], а А. І. Корнієнко та А. С. Бугайова наголошують на використанні специфічних стилістичних доміант, які допомагають реалізувати індивідуальний художній задум у межах національної мови [Корнієнко, 2016, с. 36].

Важливо враховувати й те, що ідіостиль не є сталою категорією. У різні періоди творчості автор може змінювати ритм, синтаксис, тональність, переорієнтовуватись на нові теми чи жанри. Ці зміни не скасовують

наявність стилю, а навпаки — виявляють його живу, динамічну природу. Тому дослідження ідіостилу вимагає не лише методичної точності, а й глибокого чуття тексту, здатності побачити за мовними формами авторське бачення світу [Переломова, 2001, с. 93].

Отже, поняття ідіостилу в сучасному мовознавстві набуває широкого, багатовимірного значення. Воно поєднує формальні мовні риси з естетичними і психологічними чинниками, дозволяючи повніше розкрити мовну індивідуальність автора. Уміння розпізнати й інтерпретувати ідіостиль є надзвичайно важливим як для дослідника мови, так і для читача, зацікавленого в глибшому осмисленні літературного тексту.

1.2. Ідіостиль у статистичному дослідженні: методи стилеметрії

Сучасна лінгвістична парадигма розглядає ідіостиль як об'єкт, що піддається точному кількісному вимірюванню за допомогою методів стилеметрії. Застосування статистичного аналізу дозволяє дослідникам перейти від суб'єктивних інтерпретацій художнього тексту до виявлення об'єктивних, математично верифікованих закономірностей авторського мовлення. Квантитативний підхід забезпечує фіксацію унікальних параметрів ідіолекту, які виявляються через специфічну частотність мовних одиниць на різних рівнях текстової структури. У дослідженнях підкреслюється, що статистичні методи дають змогу виявити сталі характеристики авторського стилю. Зокрема, Ф. Мостеллер та Д. Валас доводять, що аналіз частотності функціональних слів дозволяє з високою точністю ідентифікувати автора тексту, оскільки такі мовні одиниці є незалежними від тематики й відображають глибинні особливості індивідуального стилю [Мостеллер 1966].

Розвиток кількісних методів у вітчизняному мовознавстві пов'язаний насамперед із працями українських лінгвостатистиків, які адаптували статистичний підхід до дослідження мовних явищ української мови. Зокрема, В. С. Перебийніс наголошує, що використання статистичних параметрів дозволяє об'єктивно розмежувати функціональні стилі та виявити індивідуальні особливості авторського письма, оскільки кожен текст має свою унікальну цифрову характеристику розподілу мовних одиниць. Дослідниця зазначає, що саме статистичний підхід відкриває шлях до точного моделювання мовної структури, мінімізуючи вплив дослідницького суб'єктивізму [Перебийніс 1967].

Важливим кроком у практичній реалізації цих ідей стало видання «Частотного словника сучасної української художньої прози» (1981). Укладачі словника зазначають, що автоматичне опрацювання тексту за допомогою ЕОМ дозволяє розкрити ті формальні риси мовних одиниць, на які може спиратися машина для виявлення закономірностей будови тексту та функціонування мови в мовленні [Перебийніс 1981].

Розвиваючи ці положення у праці «Частотні словники та їх використання» (1985), автори В. С. Перебийніс, М. П. Муравицька та Н. П.

Дарчук зазначають, що частотний словник постає не лише лексикографічним довідником, а насамперед потужним інструментом лінгвістичного аналізу. Дослідники зазначають, що за допомогою частотних характеристик стає можливим об'єктивне визначення спільних та індивідуальних рис авторських стилів. Зокрема, автори наголошують, що порівняльний аналіз словникового складу вибірок із текстів різних письменників дозволяє встановити ступінь лексичного зв'язку між творами та виокремити специфічне «ядро» лексики, притаманне конкретному автору. [Перебийніс, 1985, с. 160].

Новий етап у розвитку стилеметрії пов'язаний із переходом до аналізу мікроструктур мови, зокрема її морфемного рівня. О. М. Зубань зазначає, що комп'ютерне лексикографічне моделювання морфемної системи дозволяє досліджувати ідіостиль як систему частотних морфемних моделей. Дослідниця наголошує, що використання електронних морфемних словників і конкордансів дає змогу виявити індивідуальні уподобання автора у виборі словотворчих засобів, які часто є несвідомими, а отже — більш стійкими характеристиками ідіолекту. Авторка зазначає, що морфемна статистика є релевантним інструментом для ідентифікації авторства, оскільки вона відображає глибинні структурні закономірності побудови тексту [Зубань 2020].

В умовах сучасної цифровізації філологічних досліджень автоматизація перестає бути просто допоміжним технічним засобом, а перетворюється на повноцінний метод верифікації авторства. Використання комп'ютерних алгоритмів дозволяє досліднику «дистанціюватися» від тексту, усуваючи емоційне сприйняття художніх образів і фокусуючись на статистичних константах, які й складають жорстку структуру ідіостилу. У контексті світового досвіду дослідники зазначають, що роль стилеметричних методів у лінгвістиці останнім часом отримує все більше уваги в різних дисциплінах, зокрема у форензичній (судовій) лінгвістиці, де кількісний аналіз стає ключовим для ідентифікації особи [Нгуен 2021]. Зокрема, М. Кесемонт підкреслює, що стилеметрія стала міждисциплінарним інструментом, який поєднує лінгвістику, інформатику та статистику для дослідження авторського стилю і встановлення авторства текстів [Кесемонт, 2014 с. 2].

Сучасні лінгвістичні дослідження засвідчують зростання ролі кількісних методів у вивченні мови як складної системи. Як зазначає С. Н. Бук, мовна система характеризується не лише якісними, а й кількісними параметрами, причому ці характеристики «властиві усім рівням мовної системи, та особливо виразні для лексичного» [Бук, 2021 с. 166]. У цьому контексті статистична лінгвістика розглядається як один із важливих напрямів сучасних досліджень. Водночас розвиток кількісних підходів пов'язується з можливостями автоматизованого опрацювання текстів, зокрема створенням частотних словників і використанням комп'ютерних

технологій для аналізу мовних даних [Краснобаєва-Чорна, 2018, с. 94]. У сучасних закордонних дослідженнях також наголошується на значенні кількісного аналізу текстів для виявлення особливостей авторського стилю. Зокрема, підкреслюється, що застосування обчислювальних методів і статистичних моделей дозволяє виявляти індивідуальні мовні патерни автора та здійснювати їх порівняння на основі формалізованих параметрів тексту [Стамататос 2018], що підтверджує ефективність кількісного підходу в дослідженні ідіостилу.

Водночас сучасні дослідження конкретизують ці положення, демонструючи, що кількісні параметри тексту — обсяг словника, індекс різноманітності, кількість *hapax legomena*, індекс винятковості — дають змогу змодельовати статистичний профіль ідіостилу та об'єктивно описати індивідуальні особливості мовлення автора; при цьому підкреслюється, що «квантитативні характеристики... відіграють особливу роль у дослідженні ідіостилу письменника», а застосування статистичних методів передбачає квантифікацію мовного матеріалу та його аналіз як системи числових показників, що уможлиблює виявлення закономірностей функціонування мовних одиниць і більш точне окреслення рис авторського стилю [Шитик, 2024, с. 102]. Крім того, у сучасних українських дослідженнях підкреслюється, що формування лінгвостатистичного «портрета» ідіостилу передбачає систематизацію емпіричних даних про статистичну структуру тексту та їх зіставлення з типовими параметрами функціонального стилю мови, що дає змогу виявити індивідуальні риси авторського мовлення; відповідно, ідіостиль постає як сукупність кількісно вимірюваних характеристик тексту, об'єднаних у цілісну модель [Зубань, 2024, с. 223].

Додатково сучасні корпусно-лінгвостатистичні дослідження розширюють уявлення про ідіостиль як багатовимірну систему кількісних параметрів тексту. С. Н. Бук зазначає, що квантитативна параметризація великої прози автор демонструє внутрішню диференціацію творів одного письменника, а також що квантитативні характеристики словника і тексту конкретного письменника не гомогенні, а демонструють внутрішню диференціацію за різними параметрами, що підтверджує можливість виявлення індивідуального стилю через систему статистичних показників. [Бук, 2021, с. 4].

Особливого значення у межах стилеметричних досліджень набувають питання атрибуції та лінгвістичної експертизи текстів. Атрибуція розглядається як процес встановлення авторства на основі аналізу стійких мовних ознак, що формують індивідуальний стиль письменника. Зокрема, І. М. Кульчицький та К. І. Гнатовська зазначають, що корпусно-статистичний аналіз дозволяє виявити індивідуальні мовні параметри тексту та здійснювати зіставлення різних авторських корпусів з метою встановлення їх належності [Кульчицький, 2019, с. 969]. Водночас Н. Я.

Лотоцька підкреслює, що частотні характеристики та показники лексичного складу виступають об'єктивними критеріями ідентифікації авторського стилю та можуть ефективно застосовуватися у практиці лінгвістичної експертизи [Лотоцька, 2023, с. 66]. У сучасних англomовних дослідженнях ці підходи також підтверджуються: зокрема, М. Едер зазначає, що стилеметричні методи, засновані на частотному аналізі мовних одиниць, дозволяють з високою точністю встановлювати авторство текстів та виявляти індивідуальні стилістичні закономірності [Едер, 2010, с. 7]. Таким чином, поєднання кількісних методів, корпусного аналізу та лінгвістичної інтерпретації забезпечує високий рівень точності атрибуції та дозволяє розглядати ідіостиль як ефективний інструмент лінгвістичної експертизи.

У сучасній лінгвістиці атрибуції текстів застосування багатопараметричних стилеметричних моделей дозволяє аналізувати статистичні індекси та їх співвідношення, що дає змогу визначати ступінь подібності між текстами й встановлювати їх авторську належність. Зокрема, зіставлення числових значень статистичних параметрів і використання метрик відстані між текстами дає змогу виявити як спільні, так і відмінні риси мовної організації текстів та зробити обґрунтовані висновки щодо авторства [Дарчук, 2024 с. 235].

Узагальнюючи викладене, слід зазначити, що кількісні методи є ефективним інструментом дослідження ідіостилу, оскільки дозволяють об'єктивно фіксувати та аналізувати частотні характеристики мовних одиниць і виявляти індивідуальні особливості авторського мовлення. У цьому контексті особливого значення набуває індексування частин мови, яке забезпечує систематизацію тексту за морфологічними категоріями та створює основу для подальшого статистичного аналізу. Розподіл частин мови виступає стабільним показником ідіостилу, що дозволяє моделювати структурні особливості тексту й підвищує точність стилеметричних досліджень та атрибуції авторства, поєднуючи морфологічний і квантитативний підходи в єдину аналітичну систему.

1.3. Мовотворчість Оксани Забужко та Сергія Жадана у мовознавчих студіях

Сучасний український літературний простір вирізняється розмаїттям стильових напрямів, авторських голосів та форм художнього самовираження. На цьому тлі постаті Оксани Забужко та Сергія Жадана займають особливе місце, адже їхня творчість не лише формує уявлення про новітню українську літературу, а й багато у чому її визначає. Вони репрезентують різні покоління, різні дискурсивні стратегії, різне бачення функції слова, але об'єднані глибоким відчуттям мовного простору як простору ідентичності, пам'яті, боротьби і творчого оновлення [Єрмоленко, 1999, с. 25].

Мовна творчість кожного з цих авторів є складним поєднанням лексичної насиченості, стилістичної сміливості, інтонаційної багатшаровості й глибинної наративної структури. Їхні тексти вирізняються не просто авторською манерою, а цілісною концепцією мовлення як художнього жесту, де кожне слово несе не лише змістове, а й емоційне, культурне, символічне навантаження [Кравець, 2004, с. 143]. У цьому контексті дослідження ідіостилу Забужко і Жадана передбачає не лише опис їхніх лексичних чи синтаксичних уподобань, а й уважне вслуховування в ритм, тон, підтекст їхнього письма.

Творчість Оксани Забужко — це приклад глибоко філософської, інтелектуалізованої літератури, де мова часто стає засобом роздуму, деконструкції, саморефлексії. У її текстах поєднуються есеїстична відстороненість і щоденникова щирість, фольклорні алюзії і модерністська руйнація наративної цілісності [Ткаченко, 2003 с. 111]. О. М. Капура зазначає, що прозові твори Оксани Забужко мають характерну «словесність», мовну виразність у відображенні явищ буття, а серед визначальних рис її письма дослідниця виокремлює інтертекстуальність, фрагментарність, цитатність та автобіографічність [Капура, 2023, с. 111]. В той час як, М. М. Максименко вказує, що мовний феномен письменниці вирізняється «незвичністю, нешаблонністю лексики та образних засобів, вільним поводженням з стильовими нормами літературного словотворення» [Максименко, 2015, с. 5].

Натомість Сергій Жадан працює з іншим типом мовного простору. Його стиль динамічний, урбаністичний, експресивний, у ньому відчувається ритм вулиці, енергетика прямої мови, мовна свобода, що межує з авантюризмом. Проза й поезія Сергія Жадана глибоко занурені в соціальний контекст, але при цьому вони позбавлені публіцистичної прямолінійності — автор зберігає образну щільність, метафоричність, мовну несподіваність [Коломієць, 2019, с. 211]. Його тексти часто вибудовуються на зіткненні низького і високого, іронічного й сакрального, що теж формує яскраву ідіостильову матрицю. Должикова зазначає, що мовна організація творів Сергія Жадана поєднує поетичні, розмовні та книжні елементи, що забезпечує відтворення мовної картини епохи та індивідуального стилю автора [Должикова, 2013 с. 131]. В. С. Грида вважає, що ідіостиль Сергія Жадана формується як система мовних засобів, що відображає індивідуальну картину світу автора та його художнє бачення дійсності [Грида, 2023 с. 33].

Мовна тканина творів Оксани Забужко передбачає розгорнуту семантичну гру, багаторівневу інтертекстуальність, схильність до культурного кодування змісту [Арешенков, 2007, с. 33]. Її мова вирізняється комплексністю синтаксичних конструкцій, де навіть незначні відхилення від стандартних норм стають навмисним засобом впровадження нових смислових відтінків. Такий підхід дозволяє

перетворити кожен фрагмент тексту на своєрідну «мовну сцену», в якій кожна пауза, кожен розділовий знак набуває символічного значення [Коломієць, 2019, с. 29]. В. В. Карасьов і О. П. Левченко зазначають, що ідіостиль Оксани Забужко визначається інтеграцією суб'єктивного голосу та об'єктивованої розповіді, орієнтацією на абстракцію і концептуалізацію та підвищеною стилістичною виразністю [Карасьов, 2024, с. 242].. У іншому дослідженні ці ж автори підкреслюють, що тексти Забужко характеризуються більшою складністю, що проявляється у довжині речень і слів та структурі мовлення [Карасьов, 2023, с. 1].

Сергій Жадан, у свою чергу, орієнтується на створення тексту як діалогу з сучасним світом, де мовна пластика формується за принципом змінності та адаптивності до контексту реального спілкування. Його твори вирізняються динамічністю і короткочасною емоційною інтенсивністю, що досягається через використання простих, але водночас експресивних мовних конструкцій. У його текстах відчутна увага до розмовної мови, до інтонаційної природності, до того, як звучить слово на вулиці, в реальній комунікації. Така уявна простота — оманлива, адже за нею криється висока міра стильової керованості [Ставицька, 2009, с. 8]. Г. І. Гримашевич зазначає, що однорідні члени речення є ключовим засобом організації синтаксису в текстах Жадана, причому домінування присудків забезпечує динаміку та експресивність мовлення [Гримашевич, 2025, с. 134].

Порівнюючи два підходи до мовної творчості, можна відзначити, що Оксана Забужко віддає перевагу структурованості, що базується на збереженні традиційних літературних норм у поєднанні з інноваціями, тоді як Сергій Жадан вдається до більш експериментального розкладу мовного простору, що дозволяє йому вільно маневрувати між жанрами та стилями [Пономарів, 2000 с. 89]. З одного боку, це дозволяє Оксані Забужко створювати тексти, насичені символами та алюзіями, що мають потужну інтерпретаційну базу, з іншого — Сергій Жадан завойовує читача живою енергією та безпосередністю мовного сприйняття. Кононенко зазначає, що художній дискурс Сергія Жадана характеризується поєднанням різностильових елементів, зокрема жаргонізмів, діалектизмів і мовної гри, що свідчить про новаторський характер його текстотворення [Кононенко, 2018, с. 17] .

Ключовим аспектом дослідження є також врахування контекстуального впливу, коли мова твору інтегрується в ширшу систему соціокультурних, історичних і політичних факторів. Наприклад, автори активно реагують на події свого часу, що знаходить відбиток у виборі мовних засобів. У творчості Оксани Забужко це може бути інтерпретовано як прагнення зберегти культурну пам'ять і створити символічну репрезентацію історичного досвіду, тоді як у Сергія Жадана така адаптація є відгуком на сучасні соціальні трансформації, що постійно змінюються. Це дозволяє не

лише простежити еволюцію авторського стилю, а й встановити закономірності, які підтверджують взаємозв'язок між мовною практикою та суспільним контекстом [Селіванова, 2008, с. 405]. М. А. Крупка зазначає, що після появи роману «Польові дослідження з українського сексу» жіночий досвід перестає бути табуйованим у культурі [Крупка, 2007, с. 73]. Це підтверджують слова А. Лісничої, що у творчості Оксани Забужко реалізуються феміністичні ідеї через образи самотійних і незалежних жінок [Лісничка, с. 2].

Окрім цього, аналіз творчості охоплює й виявлення внутрішньої мовної поліфонії, що характеризується використанням різних мовних реєстрів, стилістичних прийомів і навіть жанрових комбінацій. Автори часто застосовують інтертекстуальні цитати, фрагменти фольклору, діалектизми, що робить текст багатшаровим та різноманітним. Це відкриває можливість не лише описувати мовні одиниці, але й виявляти синергетичні взаємозв'язки, які формують особливості авторського мовлення. В. В. Карасьов і О. П. Левченко зазначають, що аналіз метафор у текстах Оксани Забужко дозволяє дослідити культурну специфіку та особливості художнього мислення авторки [Карасьов, 2023, с. 436]. М. Антошина та В. Зайцева зазначають, що метафора в текстах Жадана є ключовим елементом образного мислення, який формує індивідуальну авторську концепцію світу [Антонишина, 2023, с. 1]. А. Черниш вказує, що мова в романі «Ворошиловград» постає як складна адаптивна система, що реагує на соціокультурні зміни та формує художні концепти [Черниш, 2024, с. 14].

Отже, аналіз мовотворчості Оксани Забужко та Сергія Жадана засвідчує наявність індивідуально маркованих ідіостильових систем, що реалізуються через різнорівневу організацію мовних засобів — від лексики й синтаксису до інтонаційної та концептуальної структури тексту. Ідіостиль Оксани Забужко характеризується інтелектуалізованістю, інтертекстуальністю та ускладненою синтаксичною організацією, тоді як ідіостиль Сергія Жадана вирізняється динамічністю, експресивністю та орієнтацією на розмовну стихію мовлення. Водночас обидва автори демонструють здатність трансформувати мовні ресурси відповідно до художнього задуму та соціокультурного контексту, що підтверджує їхню роль як репрезентантів сучасного літературного процесу та носіїв унікального мовного бачення світу.

Висновки до розділу I

Аналіз теоретичних засад вивчення ідіостилу у сучасному мовознавстві свідчить про суттєве розширення цього поняття порівняно з традиційним розумінням стилю як сукупності мовних засобів. Ідіостиль постає складною багатокомпонентною системою, яка охоплює мовні, когнітивні, культурні й естетичні аспекти, що відображають унікальну авторську картину світу. Він функціонує не лише як засіб вираження думок, але і як

інструмент самоідентифікації, через який автор розкриває своє світосприйняття завдяки індивідуальному добору, комбінації й трансформації мовних ресурсів.

Упровадження кількісних методів аналізу, зокрема стилеметрії, значно розширює межі дослідження ідіостилю, забезпечуючи його вивчення на основі наукової точності. Виявлення індивідуальних рис авторської мови здійснюється з використанням таких статистичних інструментів, як частотний аналіз, індекси лексичної різноманітності та інші кількісні параметри. Ці підходи дають змогу виявити приховані закономірності текстової структури, які часто залишались поза увагою в рамках традиційних якісних методів. Особливу роль відіграє аналіз розподілу частин мови, який слугує корисним інструментом для систематизації мовного матеріалу і допомагає точно ідентифікувати особливості ідіостилю. Завдяки цьому стає можливим створення моделей мовного профілю авторів із високим рівнем точності.

Дослідження зосереджується на мовнотилістичних стратегіях двох видатних сучасних українських письменників — Оксани Забужко та Сергія Жадана. Їхні літературні тексти наочно демонструють два яскраво виражені й водночас контрастні підходи до організації мовного матеріалу. Стель Забужко вирізняється інтелектуальною насиченістю, складною синтаксичною архітектонікою та глибоким використанням інтертекстуальних зв'язків. Натомість стиль Жадана відзначається емоційною насиченістю, енергійністю й опорою на елементи розмовної народної мови. Незважаючи на суттєві відмінності, обидва письменники формують унікальні ідіостилі, що слугують виразником їхнього неповторного світосприйняття і художніх концепцій.

Перші результати дослідження підтверджують, що ідіостиль є одним із найскладніших явищ у мовознавстві, яке проявляється на всіх рівнях текстової організації — від морфології й лексики до синтаксису чи концептуальної структури. Для його глибокого аналізу необхідне поєднання якісних і кількісних методів. Такий інтегрований підхід дозволяє як описати характерні риси авторського стилю, так і точно виміряти їхні кількісні характеристики. Загалом це відкриває широкі перспективи для вдосконалення стилеметричних досліджень і поглибленого вивчення мовної індивідуальності.

РОЗДІЛ II.

АВТОМАТИЧНЕ УКЛАДАННЯ ЧАСТОТНИХ СЛОВНИКІВ ПРОЗОВИХ ТВОРІВ ОКСАНИ ЗАБУЖКО ТА СЕРГІЯ ЖАДАНА

2.1. Формування текстових вибірок: автоматична передобробка тексту

Одним із найважливіших етапів аналізу ідіостилі є створення репрезентативного текстового корпусу, який дозволяє об'єктивно й статистично обґрунтовано зіставляти мовні особливості різних авторів. У межах цього дослідження було відібрано твори, що найточніше відображають характерні риси індивідуального стилю Оксани Забужко та Сергія Жадана.

Для вивчення творчості Оксани Забужко обрано роман «Польові дослідження з українського сексу» та повість «Казка про калинову сопілку». Ці твори виділяються складною синтаксичною організацією, високим рівнем інтелектуальної насиченості, а також багатим використанням культурологічної та філософської лексики. Натомість у рамках аналізу стилю Сергія Жадана був обраний роман «Інтернат», що вирізняється радикально іншим підходом до побудови тексту. Його стиль виразний, динамічний, наповнений розмовною лексикою, короткими синтаксичними конструкціями, численними вставними елементами та емоційно забарвленими висловами, які формують його неповторну літературну манеру.

На першому етапі було створено електронний корпус текстів загальним обсягом 50 тисяч слововживань. Корпус переведено у формат .txt, сумісний із програмним забезпеченням для автоматизованої обробки текстів. Для коректного відображення українських символів використовувалося кодування UTF-8.

замахалась ти, золотце. ох, замахалась. зовсім до ручки дійшла, третій місяць у м'язах трем не вшухає, вранці, прокидаючись, а надто тепер, коли прокидаєшся, сама, перше, що відчуваєш, своє прискорене серцебиття, якого нічим не заспати, гаразд, хоч спити уже без транквілізаторів, і оті страшні приступи сухої блювоти, які змагали були ночами, нагадують про себе хіба як, чистячи зуби, заглибоко засовуєш шітку, коротким нудотним вивертом, несвідомою клітинною пам'яттю про власне тупе улягання нетайним, від першої хвилини, попервах іще хоч збуднопристрасно вишептаним, а по кількох тижнях уже просто сухим, розпорядчим звертанням, візьми в ротика. глибше візьми. глибше, ну. суха ложка рот дере, еге ж, оце воно, попервах і вона ще пробувала якось порозумітися, розтлумачити, що їй теж би, своєю чергою, спершу дечого хотілося, і не тільки там, крім статевих органів, тебе в мені більше нічого не цікавить. і, якщо в тебе були якісь плани на сьогоднішній вечір, то не вадило б дати мені про це знати до того, як я пішла спати, замість сидіти шкробати свою графіку, і взагалі, я, знаєш, не люблю роздягатися сама. добре, весело обіцяв він, ми тобі завтра таку увертюру закатаєм. завтра, одначе, так ніколи й не

Рис.2.1. Фрагмент очищеного тексту Оксани Забужко

Для того, щоб перейти до автоматизованої обробки тексту спочатку реалізовувалось очищення даних вручну. Було видалено розділові знаки, числа, технічні символи, зайві пробіли та інші елементи, які не мали значущого впливу на формування статистичного профілю мовлення автора. Всі слова були переведені до нижнього регістру, що унеможливило дублювання словоформ через відмінності у використанні великих і малих літер. Після цього оброблений текст проходив через процес токенізації — розбиття суцільного тексту на окремі слововживання. Для цього була застосована бібліотека `tokenize_uk` [54], оптимізована для української мови.

Для забезпечення достовірності статистичного аналізу кожен текстовий корпус обмежувався обсягом у 50 000 слововживань, які розділялися на 50 рівних підвбірок, кожна з яких містила по 1 000 слововживань. Цей метод дозволив оцінити стабільність частотних характеристик у межах різних сегментів тексту і здійснити порівняльний аналіз мовних моделей авторів.

Важливим етапом попередньої обробки був морфологічний аналіз, проведений з використанням бібліотеки `rumorphy3` [55]. Вона автоматично визначала леми та частини мови кожного слова, що дало змогу скласти докладні частотні словники за словоформами, лемами та граматичними категоріями. У випадках, коли система не могла ідентифікувати певні слова, їхній аналіз здійснювався вручну. Морфологічне індексування відіграло вирішальну роль у дослідженні, оскільки саме вивчення частин мови дало змогу виявити приховані закономірності у стилістиці аналізованих авторів.

Усі отримані результати були збережені в базі даних SQLite. Були створені окремі таблиці для зберігання інформації про словоформи, леми, частини мови, а також проміжна таблиця, яка містила їхній взаємозв'язок разом із граматичними категоріями та частотними характеристиками кожної підвбірки (див. Додатки В і Г). Така організація бази даних забезпечила високу ефективність систематизації мовного матеріалу й стала основою для подальшого стилOMETричного аналізу індивідуальних мовних особливостей Оксани Забужко та Сергія Жадана.

2.2. Електронні частотні словники: загальна характеристика

У рамках проведеного дослідження було сформовано чотири основні таблиці в базі даних: БД «словоформи», БД «леми», БД «част_мови» та БД «intermediate_data». Кожна з таблиць виконувала певну функцію в процесі створення частотних словників і підготовки матеріалу для стилеметричного аналізу.

У результаті автоматизованої обробки текстового корпусу були сформовані електронні частотні словники прозових творів Оксани Забужко та Сергія Жадана. Процес створення словників здійснювався із застосуванням програмного забезпечення, розробленого мовою Python (див. Додаток Г), з результатами, збереженими у базах даних SQLite (див. Додаток В,Г). Такий підхід дозволив впорядкувати значний масив мовного матеріалу й створив можливості для його подальшого статистичного аналізу.

словоформа	gen_freq	підв_1	підв_2	підв_3	підв_4	підв_5	підв_6	підв_7	підв_8	підв_9	підв_10	підв_11	підв_12	підв_13	підв_14	підв_15	підв_16	підв_17	підв_18	підв_19	підв_20
Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр
голос	29	0	1	2	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0
лиш	29	0	1	0	1	0	0	0	0	1	1	1	1	0	1	0	0	1	0	0	0
при	29	0	0	1	0	1	1	1	1	1	0	2	1	3	0	1	0	0	1	0	0
ними	29	0	0	0	0	0	0	0	0	0	0	1	2	2	0	2	0	0	1	0	0
оленка	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
об	28	1	2	1	1	0	0	1	0	0	0	0	0	1	0	0	1	0	1	2	0
мала	28	2	0	0	0	0	1	0	1	1	0	0	0	4	1	2	0	1	0	0	0
трохи	28	1	0	2	0	0	0	1	0	0	2	0	0	0	1	0	0	0	1	1	0
собою	28	0	0	1	1	0	0	1	2	0	0	0	2	0	0	1	0	0	1	1	0
лед	28	0	0	0	0	0	0	0	3	1	2	0	0	0	1	1	0	0	1	1	0
нема	27	0	2	0	1	2	1	1	0	2	3	0	0	2	0	1	1	0	0	2	0
від	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ось	26	1	0	0	0	0	0	2	1	1	0	2	2	0	1	0	0	1	2	3	0
правда	26	1	0	2	0	2	0	0	0	0	0	0	0	0	0	3	0	1	0	1	0
геть	26	1	0	0	0	1	1	0	0	0	1	0	1	0	1	0	0	1	0	1	0
лий	26	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	2	1	2	0
під	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
одного	25	1	0	0	1	1	0	0	0	0	0	0	0	0	1	2	0	2	0	1	0
вперше	25	0	1	2	1	0	0	1	0	1	1	1	1	0	0	0	0	2	1	0	0
господи	25	0	1	0	0	1	5	1	0	1	0	0	1	1	0	1	0	0	0	3	0
щастю	25	0	0	1	3	0	1	0	2	2	0	0	1	0	1	0	1	0	0	1	0
яку	25	0	0	1	0	0	1	0	0	0	2	0	1	1	0	0	0	0	0	0	0
яка	25	0	0	0	0	0	0	0	0	2	1	0	0	0	0	3	0	0	0	0	0

Рисунок 2.1. Таблиця «словоформи» для твору Оксани Забужко

Таблиця «словоформи» містить лексеми в тій формі, у якій вони представлені в тексті оригіналу. Для кожної словоформи автоматично визначалася загальна частота її використання (gen_freq), а також кількість вживань у межах кожної відібраної текстової підвибірки (підв_1, підв_2, ..., підв_50). Такий підхід дозволив ідентифікувати як загальні закономірності частотності, так і динаміку функціонування мовних одиниць у різних фрагментах творів.

лема	gen_freq	підв_1	підв_2	підв_3	підв_4	підв_5	підв_6	підв_7	підв_8	підв_9	підв_10	підв_11	підв_12	підв_13	підв_14	підв_15	підв_16	підв_17	підв_18	підв_19	підв_20
Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр
не	1307	23	21	21	20	15	24	24	30	26	21	27	19	23	38	13	26	29	26	17	0
паша	1174	20	14	16	30	16	30	17	23	25	14	22	19	27	27	29	31	17	22	32	0
на	1093	22	20	21	23	26	27	17	20	21	21	15	28	18	24	19	28	22	23	21	0
і	937	12	18	36	21	19	27	26	34	39	28	23	23	36	21	20	33	8	11	18	0
він	704	12	8	13	8	18	18	23	17	20	8	9	15	28	15	13	14	11	10	23	0
що	679	11	13	20	6	5	15	14	17	17	16	8	24	17	11	19	25	14	9	12	0
в	663	16	15	11	8	19	16	20	16	12	12	12	16	12	2	15	12	18	16	15	0
й	580	9	13	13	4	16	15	14	10	18	15	25	13	15	9	14	8	17	9	11	0
є	570	16	12	15	11	11	8	19	8	2	9	7	11	11	10	7	11	20	9	10	0
у	517	10	8	6	10	10	10	10	8	7	10	6	8	7	15	8	10	15	7	5	0
до	475	13	17	8	17	9	13	11	6	11	6	8	6	10	13	8	12	9	7	13	0
за	448	10	7	10	6	15	7	7	8	8	7	6	9	7	4	14	11	11	16	4	0
вони	390	2	4	18	6	9	6	12	15	8	4	14	22	4	7	4	2	10	5	2	0
як	386	5	2	11	8	5	4	8	9	9	9	8	7	0	5	8	4	9	9	10	0
а	352	3	5	6	2	5	2	8	6	15	11	7	10	8	18	10	4	6	4	11	0
його	321	3	3	10	8	3	13	7	7	13	3	6	8	7	7	5	8	4	4	15	0
це	315	5	6	2	9	0	5	7	6	3	5	6	5	8	4	9	3	5	2	8	0
так	292	7	6	12	5	8	5	5	13	4	6	5	7	2	13	8	5	3	2	4	0
говорити	285	2	0	4	7	3	8	12	7	8	0	6	1	6	8	2	6	4	1	6	0
але	280	6	8	4	5	7	1	5	3	2	5	10	3	11	6	4	9	6	8	5	0
бути	273	2	5	1	2	2	7	5	3	9	7	5	2	2	7	5	3	7	1	2	0
малий	246	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
тут	243	1	3	5	4	3	2	9	4	6	3	2	7	11	5	9	6	5	2	9	0

Рисунок 2.2. Таблиця «леми» для твору Сергія Жадана

Таблиця «леми» була побудована шляхом морфологічної нормалізації словоформ. Усі граматичні варіанти одного і того ж лексемного елемента об'єднувалися навколо спільної базової форми. Наприклад, форми «говорив», «говорила» та «говорить» узагальнювалися у вигляді леми

«говорити». Це дало змогу зменшити вплив граматичної варіативності та провести більш ґрунтовний аналіз лексичних особливостей індивідуального стилю письменників.

	частина	gen_freq	підв_1	підв_2	підв_3	підв_4	підв_5	підв_6	підв_7	підв_8	підв_9	підв_10	підв_11	підв_12	підв_13	підв_14	підв_15	підв_16	підв_17
	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр
1	NOUN	16600	319	304	337	355	343	305	296	317	353	374	305	350	331	339	316	351	368
2	VERB	5951	110	117	101	114	106	111	131	109	90	105	115	101	118	108	127	126	109
3	PRCL	4899	113	107	126	101	131	142	103	126	101	110	124	103	106	110	135	108	108
4	NPRO	4197	102	94	80	90	59	78	79	92	72	73	91	93	85	105	79	67	76
5	CONJ	3962	69	110	73	86	79	76	80	81	83	63	86	95	74	68	75	86	60
6	PREP	3820	76	81	74	70	75	81	76	74	73	72	67	65	75	73	65	82	58
7	ADJF	3726	76	72	69	66	97	91	84	78	88	75	85	75	80	63	68	74	91
8	ADVB	3563	68	50	85	63	53	54	87	63	69	70	67	61	62	78	70	52	72
9	GRND	2426	51	48	40	38	40	37	49	47	50	48	40	40	52	41	48	41	45
10	NUMR	642	10	16	11	14	10	21	14	9	10	8	16	12	15	10	9	7	10
11	Без частини мови	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Рис. 2.3. Таблиця «част_мови» для твору Оксани Забужко

Окремою складовою бази даних стала таблиця «част_мови», сформована автоматично за допомогою бібліотеки `ru morphology3`. У ході морфологічного аналізу кожна словоформа отримувала граматичне маркування, на основі чого вона класифікувалася за відповідними категоріями частин мови.

word	lemma	pos	total_frequency	freq_01	freq_02	freq_03	freq_04	freq_05	freq_06	freq_07	freq_08	freq_09	freq_10	freq_11	freq_12	freq_13	freq_14	freq_15
Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр
про	про	NOUN	104	1	0	2	1	1	0	1	3	2	1	0	0	3	0	0
себе	себе	NPRO	104	1	1	0	1	3	4	1	8	1	3	0	7	6	3	2
нічого	ніщо	NPRO	103	1	1	3	1	1	3	0	4	2	2	1	0	0	4	0
хтось	хтось	NPRO	96	2	3	7	4	1	1	3	1	1	0	3	5	1	1	8
каже	казати	VERB	95	1	0	0	3	0	1	3	0	1	2	1	0	6	5	2
навіть	навіть	PRCL	94	3	4	3	2	3	1	1	2	2	2	2	3	3	2	0
дивиться	дивитися	VERB	94	2	1	2	1	0	6	3	2	5	2	4	2	5	3	2
вже	вже	ADVB	93	2	5	3	1	0	0	2	1	2	3	4	4	0	2	0
пашу	паша	NOUN	93	2	2	2	4	4	5	0	2	7	1	2	0	2	7	2
паші	паша	NOUN	92	3	0	2	2	3	3	2	2	3	2	1	4	2	1	3
розуміє	розуміти	VERB	91	0	1	0	1	0	1	0	1	1	0	2	3	5	0	2
стоїть	стояти	VERB	89	1	4	4	4	5	0	0	2	2	0	0	2	1	2	1
немає	немати	VERB	89	1	1	1	2	0	1	1	0	6	0	1	3	1	4	2
хоча	хоча	PRCL	88	2	4	3	1	2	2	3	3	2	0	1	2	2	0	0
їх	вони	NPRO	88	0	1	2	2	2	1	2	2	5	1	2	5	2	2	1
було	бути	VERB	87	1	3	0	0	0	1	1	2	2	1	3	1	0	0	1
лише	лише	PRCL	86	2	1	4	3	1	1	0	1	4	2	4	2	2	3	0
де	де	NPRO	83	3	2	1	3	0	2	0	0	1	5	1	4	0	3	4
ним	він	NPRO	82	3	0	0	0	5	0	2	1	1	1	1	1	1	0	1
потім	потім	NPRO	82	0	3	0	4	0	1	1	7	1	0	1	1	3	1	4
сам	сам	NPRO	79	2	0	0	0	1	1	2	4	0	2	2	1	1	3	1
знову	знову	ADVB	79	0	1	0	0	0	5	1	3	3	0	3	2	1	4	3
двері	двері	NOUN	78	4	2	0	3	2	0	2	7	0	0	1	3	0	1	2

Рисунок 2.4. Таблиця «леми» для твору Сергія Жадана

Таблиця «`intermediate_data`» (проміжна таблиця) відігравала ключову роль у впорядкуванні результатів автоматичного аналізу. Вона містила розгорнуті дані про кожну мовну одиницю: словоформу, відповідну лему, граматичну категорію та загальну частоту її вживання в тексті. Окрім цього, у таблиці наявні частоти появи кожної одиниці в межах окремих текстових підвбірок, що дозволяло відстежувати специфіку її розподілу в різних частинах твору. Завдяки інтеграції лексичних, морфологічних і статистичних характеристик, ця таблиця стала основою для створення підсумкових частотних словників і проведення подальших стилеметричних розрахунків.

У результаті автоматичного аналізу було використано в автоматичному морфологічному аналізі таку парадигму частин мови: NOUN (іменники й власні назви), VERB (дієслова), NPRO (займенники), PRCL (частки та вигуки), CONJ (сполучники), ADVB (прислівники), ADJF (прикметники та дієприкметники), PREP (прийменники), GRND (дієприслівники) і NUMR (числівники). Окрім того, система створила категорію «Без частини мови», до якої тимчасово було віднесено мовні одиниці, що вимагали додаткової ручної перевірки.

Наповнення кожної граматичної категорії здійснювалося на основі принципів автоматичного морфологічного аналізу. У категорію NOUN належали назви об'єктів, осіб, явищ та абстрактних понять. До категорії VERB зараховували як особові форми дієслів, так і інфінітиви. У категорію ADJF входили прикметники, що позначали характеристики об'єктів, а до NPRO належали різновиди займенників. Прислівники класифікувалися в категорію ADVB, тоді як службові частини мови згрупувалися у PREP (прийменники), CONJ (сполучники) і PRCL (частки). Оскільки вигуки не формували окремої категорії у роботі аналізатора, вони включалися до групи PRCL разом із частками. Додатково система виділяла такі категорії, як GRND (дієприслівники) і NUMR (числівники).

	частина	gen_freq	підв_1	підв_2	підв_3	підв_4	підв_5	підв_6	підв_7	підв_8	підв_9	підв_10	підв_11	підв_12	підв_13	підв_14	підв_15	підв_16	підв_17
	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр
1	NOUN	16600	319	304	337	355	343	305	296	317	353	374	305	350	331	339	316	351	368
2	VERB	5951	110	117	101	114	106	111	131	109	90	105	115	101	118	108	127	126	109
3	PRCL	4899	113	107	126	101	131	142	103	126	101	110	124	103	106	110	135	108	108
4	NPRO	4197	102	94	80	90	59	78	79	92	72	73	91	93	85	105	79	67	76
5	CONJ	3962	69	110	73	86	79	76	80	81	83	63	86	95	74	68	75	86	60
6	PREP	3820	76	81	74	70	75	81	76	74	73	72	67	65	75	73	65	82	58
7	ADJF	3726	76	72	69	66	97	91	84	78	88	75	85	75	80	63	68	74	91
8	ADVB	3563	68	50	85	63	53	54	87	63	69	70	67	61	62	78	70	52	72
9	GRND	2426	51	48	40	38	40	37	49	47	50	48	40	40	52	41	48	41	45
10	NUMR	642	10	16	11	14	10	21	14	9	10	8	16	12	15	10	9	7	10
11	Без частини мови	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Рис. 2.5. Структура бази даних частотних словників Оксани Забужко

Дослідження частотної лексики в творчості Оксани Забужко, проведене на основі її художнього мовлення, виявило домінуючу роль іменників (16600 слововживань; 33,40%). Це свідчить про високий рівень предметності та концептуальної насиченості авторського стилю. Значні показники частотності було зафіксовано також для дієслів (5951 слововживань; 11,97%), часток і вигуків (4899; 9,86%), займенників (4197; 8,44%) і сполучників (3962; 7,87%). Помітна пріоритетність використання іменникових конструкцій тісно пов'язана зі складною синтаксичною побудовою текстів та інтелектуалізованим характером прози письменниці.

	частина	gen_freq	підв_1	підв_2	підв_3	підв_4	підв_5	підв_6	підв_7	підв_8	підв_9	підв_10	підв_11	підв_12	підв_13	підв_14	підв_15	підв_16	підв_17
	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр	Фільтр
1	NOUN	14806	338	321	274	347	321	306	279	264	278	306	288	267	240	273	291	319	307
2	VERB	10481	188	171	212	203	167	201	203	196	201	172	201	188	211	222	225	195	197
3	NPRO	5196	77	92	118	81	90	109	120	113	109	87	77	135	158	99	118	78	95
4	PRCL	4442	83	86	90	77	62	83	74	86	94	86	84	81	90	112	79	79	80
5	CONJ	3955	68	81	109	56	72	87	94	95	112	105	92	97	104	87	94	92	68
6	ADVB	3731	66	72	65	70	70	76	76	87	82	73	110	84	55	69	83	81	71
7	ADJF	3415	93	85	54	93	113	54	58	72	59	92	84	64	69	64	36	59	79
8	PREP	2699	57	55	50	51	71	51	67	48	40	53	36	59	52	55	54	60	71
9	GRND	539	12	12	11	7	10	4	6	21	9	8	13	11	9	14	10	16	12
10	NUMR	441	11	15	14	9	15	15	17	11	9	16	9	9	5	7	13	12	12
11	Без частини мови	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Рис. 2.6. Структура бази даних частотних словників Сергія Жадана

У корпусі текстів Сергія Жадана також виявлено домінування іменників (загалом 14 806 уживань; 29,79%). Водночас дієслова зустрічаються значно частіше, порівняно з іншими авторами, – їх зафіксовано 10 481 випадок; 21,09%. Це, ймовірно, свідчить про підвищену динамічність авторського мовлення, більшу увагу до зображення процесуальності та акцент на розмовну комунікацію. Займенники посідають третю позицію за вживаністю (5196; 10,45%), а частки, вигуки (4442; 8,94%) та сполучники (3955; 7,96%) також мають помітне значення в структурі його мови. Така граматична композиція підкреслює унікальні риси стилю Жадана, формуючи його характерний ритм і естетику художнього тексту.

Порівняльна діаграма вибірок О.Забужко та С.Жадана

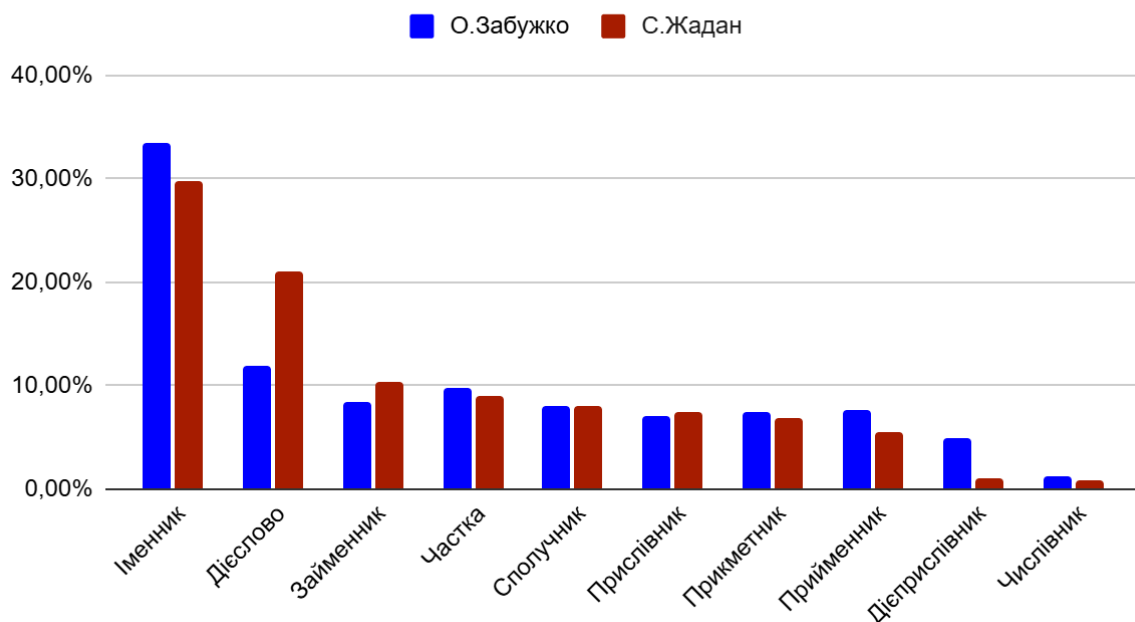


Рис. 2.7. Порівняльна діаграма відсоткового розподілу частин мови у вибірках творів Оксани Забужко та Сергія Жадана

Аналіз морфологічної будови текстів обох авторів виявив наявність як спільних рис, так і стилістичних відмінностей. Хоча в вибірках обох письменників переважають іменники, у творах Оксани Забужко їхня частота значно вища, ніж у текстах Сергія Жадана, що вказує на більшу

концептуальність і предметність її художнього світу. Натомість суттєвою відмінністю є частота вживання дієслів: у Жадана вони трапляються майже вдвічі частіше, що свідчить про динамічність, насиченість подіями та активність процесів у його прозі.

Також помітна суттєва розбіжність у вживанні дієприслівників. У текстах Оксани Забужко їх майже в п'ять разів більше, ніж у творах Сергія Жадана, що вказує на складнішу синтаксичну структуру її речень та багатство описів супутніх дій. Водночас частота використання сполучників, прислівників, прикметників, часток і займенників виявилася доволі схожою, що свідчить про наявність спільних рис у граматичній організації художнього мовлення обох авторів.

Електронні частотні словники стали ключовим інструментом кількісного аналізу мовних даних, уможливлючи перехід від описових методів до статистичного моделювання авторського стилю. В цьому контексті індексування частин мови стає надзвичайно актуальним, адже воно не лише увиразнює граматичні пріоритети тексту, а й допомагає розкрити унікальні принципи мовної організації художнього світу кожного автора.

2.3. Визначення відносної похибки для статистичного дослідження

Для оцінки надійності отриманих частотних показників у межах дослідження була проведена перевірка за допомогою розрахунку відносної похибки. Цей показник дозволяє визначити рівень статистичної стійкості частот окремих мовних одиниць у текстовій вибірці. Особливу значущість відносна похибка має у стилеметричному аналізі, де частотні характеристики високо-, середньо- та низькочастотних одиниць мають різний ступінь статистичної достовірності.

Розрахунок відносної похибки здійснювався автоматизовано за допомогою програми, написаної мовою Python (див. Додаток Д). Для цього використовувалася формула: $\epsilon = 1,96 / \sqrt{f}$, де ϵ — відносна похибка, f — частота мовної одиниці, а 1,96 — коефіцієнт, що відповідає 95% довірчому рівню. Усі розрахунки були збережені в окремій Excel-таблиці для частин мови (див. Додаток Е).

Частина мови	Частота	Відносна похибка	Частина мови	Частота	Відносна похибка
Вибірка - "Забужко ПДУК, КПКС"			Вибірка - "Жадан Інтернат"		
NOUN	16600	0,015	NOUN	14806	0,016
VERB	5951	0,025	VERB	10481	0,019
PRCL	4899	0,028	NPRO	5196	0,027
NPRO	4197	0,03	PRCL	4442	0,029
CONJ	3962	0,031	CONJ	3955	0,031
PREP	3820	0,032	ADVB	3731	0,032
ADJF	3726	0,032	ADJF	3415	0,034
ADVB	3563	0,033	PREP	2699	0,038
GRND	2426	0,04	GRND	539	0,084
NUMR	642	0,077	NUMR	441	0,093
Без частини мови	0	0	Без частини мови	0	0

Рис. 2.8. Розрахунок відносної похибки для частин мови у вибірках Оксани Забужко та Сергія Жадана

Для кожної частини мови було окремо визначено відносну похибку. У корпусі текстів Оксани Забужко найвищий рівень частотності демонструють іменники (NOUN), які зафіксовані в кількості 16 600 вживань, при цьому відносна похибка становить 0,015. Дієслова (VERB) були використані 5 951 раз із похибкою, рівною 0,025. Частки та вигуки (PRCL) мають частоту вживання, рівну 4 899, із відповідною похибкою 0,028. У вибірці текстів Сергія Жадана також було розраховано ці показники: іменники (NOUN) зустрічаються 14 806 разів із похибкою 0,016; дієслова (VERB) мають частотність 10 481 і похибку 0,019; займенники (NPRO) використано 5 196 разів із похибкою 0,027.

Діаграма порівнянь відносної похибки



Рис. 2.9. Порівняння відносної похибки частотних показників частин мови в текстових корпусах Оксани Забужко та Сергія Жадана

Аналіз показав, що для більшості частин мови значення відносної похибки залишаються на низькому рівні, варіюючись у діапазоні від 0,015 до 0,040. Це свідчить про високу надійність отриманих частотних характеристик з точки зору статистики. Найменші значення спостерігаються у таких категоріях, як іменники (NOUN), дієслова (VERB), частки (PRCL), займенники (NPRO), сполучники (CONJ) і прийменники (PREP), оскільки вони є найуживанішими у текстах. У той же час, для менш розповсюджених частин мови, таких як дієприслівники (GRND) та числівники (NUMR), похибка є значно вищою через їхню обмежену представленість у корпусах.

Узагальнені дані свідчать про те, що показники похибок двох авторських корпусів мають схожі величини. Однак у текстах Сергія Жадана для ряду

частин мови відзначено трохи вищі значення похибки, що може вказувати на нижчу стабільність частотних характеристик у порівнянні з корпусом творів Оксани Забужко.

Висновки до розділу II

У другому розділі дослідження виконано практичний етап, присвячений автоматизованому укладанню частотних словників прозових творів Оксани Забужко і Сергія Жадана. Для цього було сформовано репрезентативний корпус текстів, який включав роман Оксани Забужко "Польові дослідження з українського сексу", повість "Казка про калинову сопілку" та роман Сергія Жадана "Інтернат". Тексти попередньо підготовлено для автоматизованої обробки шляхом перетворення у формат .txt, очищення від службових символів, нормалізації регістру, токенизації та морфологічного аналізу. Стандартизованість та точність обробки забезпечено завдяки використанню бібліотек re, tokenize_uk, pymorphy3, sqlite3 і програмного середовища Python.

У ході дослідження створено електронні частотні словники словоформ, лем та частин мови, а також сформовано інтегровану таблицю результатів аналізу. Важливим аспектом дослідження стало морфологічне індексування текстів, яке дозволило здійснити не лише кількісний аналіз мовних одиниць, а й поглиблене вивчення граматичної структури авторського стилю. У результаті роботи виділено десять основних граматичних категорій: іменники, дієслова, займенники, частки, вигуки, сполучники, прислівники, прикметники, прийменники, дієприслівники та числівники. Аналіз частин мови дав змогу виявити ключові граматичні домінанти у текстах та розкрити структурні особливості ідіостилу кожного автора.

Кількісне зіставлення текстів Забужко і Жадана виявило суттєві відмінності в їхній граматичній організації. У творчості Оксани Забужко помітна висока частотність іменникових конструкцій, що свідчить про складну синтаксичну організацію її текстів і наголошує на інтелектуалізованому характері письма. У творах Сергія Жадана домінують дієслівні форми, які відображають динамічність його мовлення та процесуальність комунікації. Таким чином, кількісні характеристики граматики стають важливими маркерами індивідуального авторського стилю.

Статистична надійність результатів була підтверджена шляхом розрахунку відносної похибки для високочастотних мовних одиниць і частин мови. Це підкреслює ефективність автоматизованих методів у стилеметричних дослідженнях. Розроблена система електронного морфологічного аналізу та частотних словників створила основу для глибокого статистичного аналізу ідіостилу Оксани Забужко і Сергія Жадана. Це дозволило перейти від традиційного описового аналізу до моделювання структурних особливостей їхнього авторського мовлення.

РОЗДІЛ III.

МОРФОЛОГІЧНА СТАТИСТИЧНА МОДЕЛЬ СТИЛЕМЕТРИЧНОГО ДОСЛІДЖЕННЯ ІДІОСТИЛЮ ОКСАНИ ЗАБУЖКО ТА СЕРГІЯ ЖАДАНА

3.1. Полігони частот частин мови: порівняльний аспект

Одним із ключових напрямів сучасних стилеметричних досліджень є застосування статистичних методів для аналізу унікальних характеристик авторського стилю. Морфологічний рівень художнього тексту відіграє надзвичайно важливу роль у формуванні ідіостилю автора, оскільки співвідношення граматичних категорій дає змогу простежити закономірності мовної структури тексту і виявити кількісні особливості авторського мовлення. У цьому дослідженні було здійснено побудову морфологічної статистичної моделі на основі частотного аналізу вживання частин мови у прозових творах Оксани Забужко та Сергія Жадана.

Полігон частот, як один із графічних інструментів подання статистичних показників, дозволяє чітко відстежити особливості розподілу мовних одиниць та визначити ступінь їх стабільності у межах тексту. Використання такого інструменту в стилеметричних дослідженнях сприяє переходу від описового аналізу до створення кількісних моделей авторського стилю, що допомагає ідентифікувати морфологічні ознаки художнього мовлення. Для побудови варіаційних та інтервальних рядів використовувалися Google Таблиці (див. Додаток Є), а візуалізація отриманих статистичних даних у вигляді полігонів частот реалізовувалася автоматизовано за допомогою програмного забезпечення, створеного мовою Python (див. Додаток Ж).

Використання варіаційного ряду забезпечує можливість детально проаналізувати характер розподілу абсолютних значень частотності в межах обраної вибірки. Для кожної частини мови визначалися мінімальні та максимальні частотні значення, розмах варіації, кількість інтервалів та їхня величина. На наступному етапі створювалися інтервальні ряди, які дали змогу згрупувати частоти за певними діапазонами і виявити тенденції статистичної концентрації значень. Результати аналізу слугували основою для побудови полігонів частот, що забезпечили візуалізацію статистичних закономірностей у функціонуванні різних частин мови.

Найбільш інформативними для дослідження виявилися іменники та дієслова, адже саме ці частини мови демонструють найвищі частоти вживання і водночас мають найбільшу цінність для диференціації стилістичних особливостей авторів.

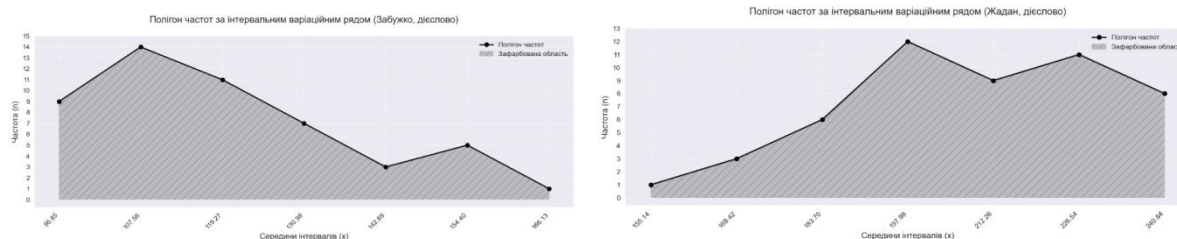


Рис. 3.1. Полігони частот інтервальних варіаційних рядів дієслів у творах Оксани Забужко та Сергія Жадана

Аналіз полігонів демонструє помітні відмінності у способі розподілу дієслівних частот. У випадку полігону тексту Оксани Забужко простежується поступове зменшення частот від центральних інтервалів до крайніх значень, що вказує на більш збалансований і рівномірний розподіл дієслівних форм у її мовленні. Водночас полігон Сергія Жадана вирізняється наявністю двох чітких локальних максимумів, а також різкими перепадами між суміжними інтервалами. Подібна графічна структура свідчить про значну неоднорідність дієслівного розподілу, акцентуючи на підвищеній динамічності стилю автора.

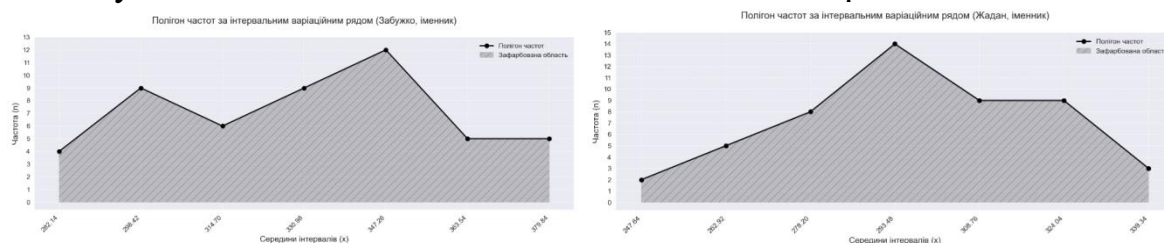


Рис. 3.2. Полігони частот інтервальних варіаційних рядів іменників у творах Оксани Забужко та Сергія Жадана

Полігони іменників обох авторів демонструють подібну загальну структуру, що підкреслює важливу роль номінативних конструкцій у проаналізованих текстах. Однак на полігоні Оксани Забужко найвища частотна концентрація фокусується в одному з центральних інтервалів, після чого спостерігається доволі різкий спад. Натомість тексти Сергія Жадана характеризуються більш плавним зміщенням між інтервалами та менш відчутною концентрацією в межах одного частотного діапазону. Це вказує на ширший спектр застосування іменникових одиниць у творчості Жадана та більшу різноманітність у їхньому розподілі.

Статистичний аналіз охопив і інші граматичні категорії. Займенники (NPRO), прийменники (PREP) та сполучники (CONJ) проявили відносно стабільний характер функціонування з помірними коливаннями частотних значень. Ця стала тенденція підкреслює синтаксичну збалансованість художнього тексту та передбачуваність використовуваних службових частин мови.

Помітно вираженіший рівень частотних змін було зафіксовано у часток (PRCL), вигуків, прикметників (ADJF) і прислівників (ADV). Варіаційні та інтервальні ряди цих категорій демонструють локальні сплески

частотності, які можуть свідчити про нерівномірність використання стилістично-експресивних або описових засобів мови. У зв'язку з тим, що вигуки були враховані разом із частками, отримані дані вказують також на емоційну насиченість художнього тексту.

Дієприслівники (GRND) та числівники (NUMR), навпаки, відзначаються найнижчими частотними показниками та високою варіативністю. Це обумовлено обмеженим застосуванням відповідних граматичних категорій у досліджуваних текстах.

Морфологічне статистичне моделювання дало змогу виявити, що найбільш виразну стилерозрізнявальну цінність мають іменники та дієслова. Саме ці частини мови найбільш ефективно репрезентують відмінності між індивідуальними стилями Оксани Забужко та Сергія Жадана. На мовлення Забужко характерною є вища концентрація іменникових конструкцій у поєднанні з рівномірним розподілом інших граматичних категорій. У свою чергу, стиль Жадана вирізняється більш активним використанням дієслівних форм, що сприяють посиленню динамізму морфології його текстів.

3.2. Зіставлення вибірок за смугою коливання середньої частоти частин мови

Ключовим етапом статистичного аналізу морфологічної структури художнього тексту є аналіз смуги коливання середньої частоти частин мови. На відміну від традиційного підходу, що обмежується порівнянням абсолютних частотних показників, застосування довірчих інтервалів дозволяє глибше оцінити надійність отриманих результатів. Це, у свою чергу, сприяє визначенню меж можливих варіацій середніх значень у межах досліджуваних вибірок. Подібний методологічний підхід набуває особливого значення у сфері стилеметричних досліджень, оскільки він дозволяє не лише оцінити середню частоту використання мовних одиниць, але й проаналізувати ступінь стабільності їхнього функціонування у структурі авторського мовлення. Усі необхідні статистичні розрахунки для визначення середньої частоти частин мови, обчислення стандартного відхилення та побудови довірчих інтервалів здійснювалися автоматизовано за допомогою спеціалізованого програмного скрипту, створеного мовою Python (див. Додаток 3).

Для кожної частини мови було визначено середню частоту вживання шляхом аналізу п'ятдесяти підвибірок, кожна з яких містила по 1000 слововживань. На основі отриманих емпіричних даних було обчислено стандартне відхилення та побудовано довірчі інтервали для трьох рівнів значущості: 68,3 %, 95,5 % та 99,7 %. Ці інтервали відповідають правилу трьох сигм у нормальному розподілі: $\mu \pm \sigma$ включає приблизно 68,3 % усіх можливих значень, $\mu \pm 2\sigma$ охоплює близько 95,5 %, а $\mu \pm 3\sigma$ поширюється на майже 99,7 % значень вибірки. Менша ширина довірчого інтервалу

відносно середнього значення вказує на більшу стабільність відповідної мовної одиниці.

Частина мови	Вибірка	μ (сер.част)	σ (сігма)	ДІ (68,3 %)	ДІ (95,5 %)	ДІ (99,7 %)
Іменник	Забужко	332.00	29.22	[302.78; 361.22]	[273.57; 390.43]	[244.35; 419.65]
	Жадан	296.12	23.45	[272.67; 319.57]	[249.22; 343.02]	[225.76; 366.48]
Дієслово	Забужко	119.02	17.94	[101.08; 136.96]	[83.13; 154.91]	[65.19; 172.85]
	Жадан	209.62	22.45	[187.17; 232.07]	[164.72; 254.52]	[142.27; 276.97]
Частка	Забужко	97.98	44.22	[53.76; 142.20]	[9.54; 186.42]	[-34.68; 230.64]
	Жадан	88.84	12.02	[76.82; 100.86]	[64.80; 112.88]	[52.77; 124.91]
Займенник	Забужко	83.94	16.06	[67.88; 100.00]	[51.82; 116.06]	[35.77; 132.11]
	Жадан	103.92	21.65	[82.27; 125.57]	[60.62; 147.22]	[38.97; 168.87]
Сполучник	Забужко	79.24	16.01	[63.23; 95.25]	[47.23; 111.25]	[31.22; 127.26]
	Жадан	79.10	16.10	[63.00; 95.20]	[46.90; 111.30]	[30.80; 127.40]
Прийменник	Забужко	76.40	19.52	[56.88; 95.92]	[37.35; 115.45]	[17.83; 134.97]
	Жадан	53.98	8.55	[45.43; 62.53]	[36.88; 71.08]	[28.32; 79.64]
Прикметник	Забужко	74.52	13.38	[61.14; 87.90]	[47.75; 101.29]	[34.37; 114.67]
	Жадан	68.30	17.84	[50.46; 86.14]	[32.62; 103.98]	[14.77; 121.83]
Прислівник	Забужко	71.26	13.13	[58.13; 84.39]	[45.00; 97.52]	[31.87; 110.65]
	Жадан	75.00	11.32	[63.68; 86.32]	[52.35; 97.65]	[41.03; 108.97]
Дієприслівник	Забужко	48.52	14.46	[34.06; 62.98]	[19.61; 77.43]	[5.15; 91.89]
	Жадан	10.98	4.20	[6.78; 15.18]	[2.58; 19.38]	[-1.63; 23.59]
Числівник	Забужко	12.84	3.96	[8.88; 16.80]	[4.93; 20.75]	[0.97; 24.71]
	Жадан	8.80	04.08	[4.72; 12.88]	[0.64; 16.96]	[-3.44; 21.04]

Рис. 3.3. Порівняльна таблиця середніх частот, сігми та довірчих інтервалів

Порівняльна таблиця ілюструє результати статистичного аналізу використання частин мови у вибірках текстів Оксани Забужко і Сергія Жадана. Результати аналізу свідчать, що іменники та дієслова мають найвищі середні частоти, оскільки вони становлять основу морфологічної структури художнього тексту. Водночас для менш уживаних категорій, таких як дієприслівники та числівники, характерне вище значення стандартного відхилення відносно середньої частоти, що свідчить про підвищену варіативність їх використання. Побудовані довірчі інтервали забезпечують можливість оцінити рівень стабільності кожної граматичної категорії та провести порівняння особливостей їх функціонування в ідіостилях обох авторів.

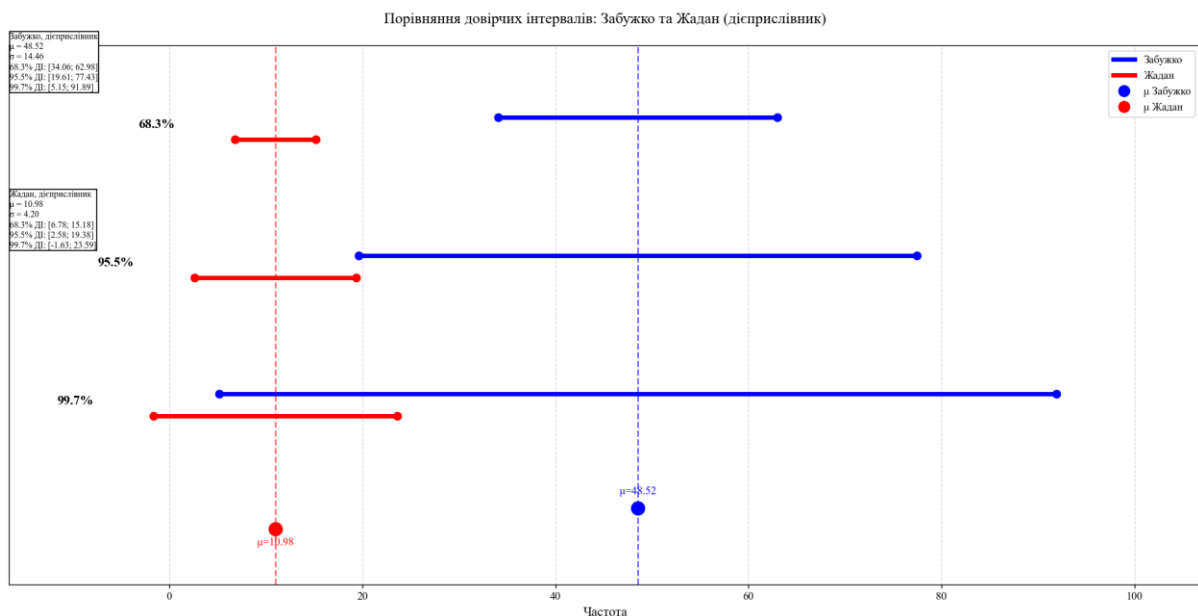


Рис. 3.4. Довірчі інтервали середньої частоти дієприслівників

Аналіз довірчих інтервалів дієприслівників вказує про їхню підвищену статистичну мінливість як граматичної категорії. Відміну від дієслів, корпус творів Оксани Забужко демонструє вищі показники в цій категорії. Зокрема, середня частота дієприслівників у її текстах значно перевищує аналогічний показник у працях Сергія Жадана, що свідчить про складнішу синтаксичну організацію її мови. Широкі довірчі інтервали вказують на нерівномірність розподілу цієї граматичної категорії в текстах та залежність її частотності від стилістичних і композиційних характеристик окремих фрагментів творів.

Окрім цього, дієприслівники відіграють ключову роль у створенні динамізму оповіді та забезпечують додаткові смислові нюанси дії. Саме тому їх використання суттєво залежить від конкретних комунікативних завдань автора, які також впливають на розширення довірчих інтервалів. Схожа ситуація спостерігається і щодо категорії дієслів, що також демонструють подібний розподіл у текстах.

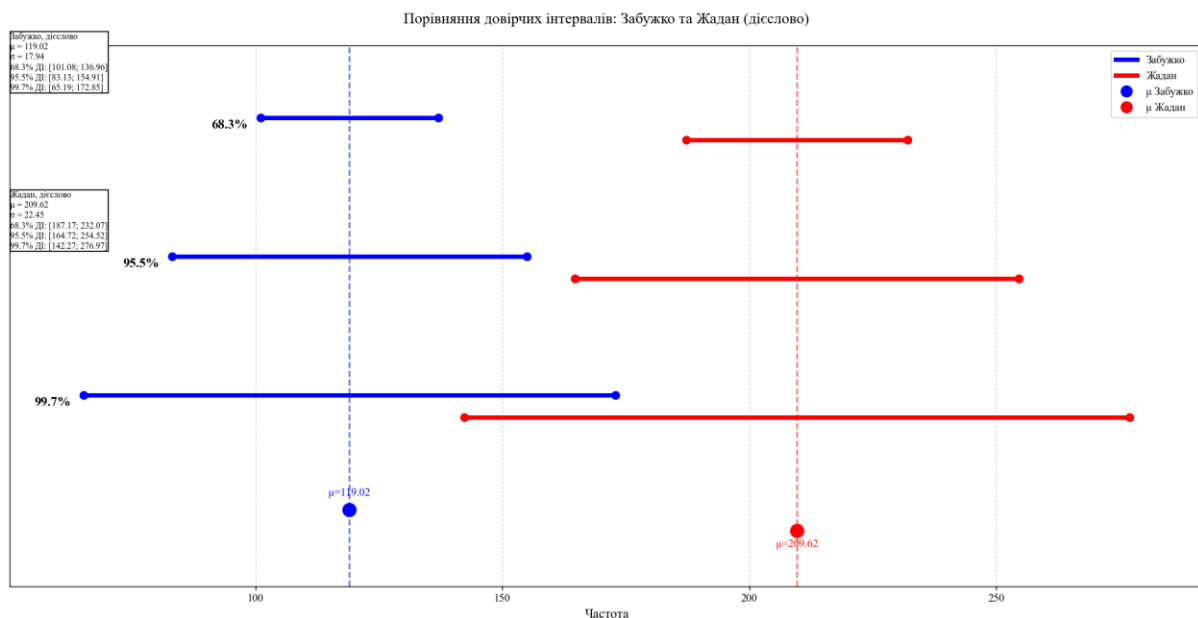


Рис. 3.5. Довірчі інтервали середньої частоти дієслів

Дослідження довірчих інтервалів дієслів у текстах Оксани Забужко та Сергія Жадана демонструє значні відмінності між їх творами. Як показує графік, корпус Жадана характеризується суттєво вищою середньою частотою дієслів порівняно з творами Забужко. Важливо зазначити, що діапазони інтервалів для цих вибірок практично не накладаються один на одного, що вказує на наявність статистично значущої розбіжності у стилістичних підходах авторів. Такі результати наочно ілюструють динамічність мовлення Жадана, яке виражається через високу концентрацію процесуальних конструкцій і активний розвиток дії у його текстах.

Слід особливо зазначити, що значна ширина довірчих інтервалів для дієприслівників та дієслів не свідчить про недоліки у вибірці чи низьку репрезентативність матеріалу. Навпаки, ці показники відображають реальні чинники функціонування зазначених категорій у художньому стилі. Аналіз цих даних дозволяє більш точно відобразити структурні особливості ідіостилу авторів.

Порівняльний аналіз довірчих інтервалів різних граматичних категорій виявив закономірний зв'язок між частотністю вживання категорії та ступенем її статистичної стабільності. Граматичні одиниці, які вирізняються високою частотністю — зокрема, іменники, дієслова, прикметники та службові частини мови, — демонструють компактніші довірчі інтервали та нижчу варіативність середньої частоти. Натомість категорії з меншою частотністю проявляють значно ширші інтервали коливань, що супроводжується вищим ступенем статистичної мінливості.

3.3. Стилеметричне індексування морфологічної структури текстів

Для аналізу та порівняння ідіостилів Оксани Забужко і Сергія Жадана було застосовано стилеметричну модель у рамках системи TextAttributor.

У цій моделі використано широкий спектр лінгвостатистичних індексів, які дозволяють дослідити лексичні, граматичні та психолінгвістичні характеристики текстів через кількісні показники.

Процес розрахунку стилеметричних індексів здійснювався автоматизовано за допомогою програмного коду, написаного мовою Python (див. Додаток І). На першому етапі відбувалася токенізація тексту та його очищення з використанням бібліотеки `tokenize\_uk`. Далі з допомогою морфологічного аналізатора `rutmorphu3` автоматично визначали частини мови для кожної словоформи. На основі цих даних отримували базові кількісні характеристики тексту: число слововживань, кількість речень, обсяг словникового запасу і кількість гапаксів.

Лексичні, граматичні та психолінгвістичні індекси обчислювалися через співвідношення окремих мовних параметрів. Зокрема, індекс багатства визначався як співвідношення обсягу словника до кількості слововживань. Індекс унікальності тексту розраховувався як відношення кількості гапаксів до загального обсягу тексту, а індекс унікальності словника — як співвідношення гапаксів до обсягу словника. Ступінь номінальності оцінювався шляхом порівняння кількості іменників із кількістю дієслів. Індекс епітетизації рахувався як співвідношення іменників до прикметників, а індекс дієслівних означень — як відношення прислівників до дієслів. Аналогічно визначалися індекси прономіналізації, модальності та субстантивності — через частку займенників, часток і іменників відповідно від загального числа слововживань.

Коефіцієнт Трейгера вираховувався як відношення дієслів до прикметників, коефіцієнт матеріалізованої дії — як співвідношення дієслів до іменників, а коефіцієнт агресивності — як відсоток дієслів щодо загальної кількості слововживань.

ПОРІВНЯЛЬНА ТАБЛИЦЯ		
Показник	ЖаданІнтернат.txt	Забужко ПДУК,КПКС.txt
Кількість слів	53818	52408
Кількість речень	5574	1037
Обсяг словника	11652	19373
Індекс багатства	0.2165	0.3697
Індекс винятковості тексту	0.1320	0.2831
Індекс винятковості словника	0.6096	0.7659
Індекс іменних означень	4.3232	3.1724
Індекс дієслівних означень	0.3588	0.4055
Ступінь номінальності	1.4117	2.0102
Індекс прономіналізації	0.1039	0.0927
Індекс модальності	0.0894	0.0887
Індекс субстантивності	0.2953	0.2921
Індекс динамічності	0.7502	0.3619
Індекс зв'язності мовлення	1.2865	7.3153
Коефіцієнт Трейгера	3.0623	1.5781
Коефіцієнт опредметненої дії	0.7083	0.4975
Коефіцієнт агресивності	0.2092	0.1453

Рис. 3.6. Порівняльна таблиця лінгвостатистичного аналізу для текстів Оксани Забужко та Сергія Жадана.

Порівняння лексичних характеристик текстів виявило істотні відмінності між ідіостилями Оксани Забужко й Сергія Жадана. Індекс багатства словника у творах Забужко дорівнює 0,3697, що суттєво перевищує показник Жадана — 0,2165, і свідчить про більшу лексичну різноманітність її творів. Такі ж тенденції спостерігаються й у випадку індексу винятковості тексту (0,2831 проти 0,1320) та індексу винятковості словника (0,7659 проти 0,6096), які демонструють більшу кількість унікальних слів і вищу індивідуалізацію мовлення Забужко. Таким чином, отримані дані дають підстави стверджувати, що її проза вирізняється багатшим словниковим складом і більшою лексичною оригінальністю.

Аналіз граматичних індексів також підкреслює значні розбіжності між авторами. Тексти Оксани Забужко показали вищий рівень номінальності (2,0102 проти 1,4117), більший індекс дієслівних означень (0,4055 проти 0,3588) і суттєво вищий індекс зв'язності мовлення (7,3153 проти 1,2865). Це свідчить про складнішу синтаксичну побудову її текстів, їхню багатослівність та орієнтованість на описово-аналітичний стиль. У той самий час, показники Жадана виявили вищий індекс іменних означень (4,3232 проти 3,1724) і більший рівень прономіналізації (0,1039 проти 0,0927), що підкреслює його схильність до конкретного викладу думок і активного використання елементів розмовної мови. Параметри модальності (0,0894 та 0,0887) і субстантивності (0,2953 та 0,2921) у двох авторів майже не відрізняються, що свідчить про схожу присутність відповідних категорій у їхніх текстах.

Психолінгвістичний аналіз також підтвердив суттєві відмінності між стилістикою авторів. У творах Сергія Жадана індекс динамічності досягає 0,7502 проти 0,3619 у Забужко. Така ж перевага спостерігається і для коефіцієнта Трейгера (3,0623 проти 1,5781), коефіцієнта опредметненої дії (0,7083 проти 0,4975) та коефіцієнта агресивності (0,2092 проти 0,1453). Ці показники свідчать про те, що проза Жадана базується на інтенсивнішому використанні дієслівних конструкцій, що надає їй більшої динаміки та виразності. Загалом стилеметричний аналіз дозволяє зробити висновок: у той час як Оксана Забужко виділяється лексичним багатством, високим рівнем зв'язності й номінальності мовлення, роботи Сергія Жадана вирізняються динамічністю, процесуальністю та емоційною насиченістю.

Висновки до розділу III

У третьому розділі проведено морфологічне статистичне моделювання ідіостилів Оксани Забужко та Сергія Жадана, застосовуючи методи лінгвостатистики, стилеметрії та комп'ютерного аналізу текстів. Для обробки корпусів текстів було створено спеціалізоване програмне забезпечення на основі мови програмування Python, яке автоматизувало визначення частин мови, побудову частотних характеристик та розрахунок статистичних показників. Завдяки цьому вдалося перейти від традиційного

якісного аналізу мовлення до кількісного, забезпечивши отримання об'єктивних даних про морфологічну структуру текстів обох авторів.

Дослідження полігонів частот виявило, що найважливішими категоріями для розмежування стилів є іменники та дієслова. Тексти Оксани Забужко мають вищу концентрацію іменникових конструкцій і демонструють більш рівномірне співвідношення частин мови, що свідчить про їхню описову, аналітичну природу, а також концептуальну насиченість її художнього мовлення. Натомість роботи Сергія Жадана відзначаються значно більшою активністю дієслів та їхньою варіативністю, що надає його текстам динамічного характеру і підкреслює процесуальність наративу.

Оцінка довірчих інтервалів середньої частоти окремих частин мови дозволила визначити рівень статистичної стабільності різних морфологічних категорій. Найбільш стабільними виявилися іменники, прикметники, займенники та службові слова, тоді як дієприслівники та числівники показали вищий ступінь варіативності. Цей аналіз також підтвердив суттєві відмінності між авторськими корпусами, особливо в аспекті використання дієслівних конструкцій, що дозволяє вважати ці категорії вагомими маркерами авторського стилю.

Результати стилеметричного аналізу показали, що проза Оксани Забужко характеризується високими показниками лексичного багатства, унікальності словника, номінальності та зв'язності мовлення. Це підкреслює інтелектуальний та рефлексивний характер її письма. Натомість творчість Сергія Жадана відзначається підвищеним рівнем динамічності, високим коефіцієнтом Трейгера, інтенсивністю дійових конструкцій і емоційною напругою. Ці особливості вказують на те, що домінуючими компонентами його текстів є дія, рух та експресія. Результати дослідження підтвердили ефективність стилеметрії у визначенні специфічних рис авторського стилю та довели доцільність використання морфологічних статистичних показників як достовірного інструменту для атрибуції художніх текстів.

ВИСНОВКИ

У рамках кваліфікаційної роботи було проведено всебічне стилеметричне дослідження ідіостилів Оксани Забужко та Сергія Жадана, використовуючи морфологічну статистичну модель тексту. Теоретичний аналіз спеціалізованих наукових праць підтвердив, що ідіостиль становить багаторівневу систему мовних засобів, яка відображає унікальну мовну картину світу автора. Сучасні підходи стилеметрії дозволяють вивчати цю систему не лише на якісному рівні, але й на кількісному, завдяки застосуванню математичних та статистичних інструментів для виявлення закономірностей авторського письма.

У ході дослідження були сформовані корпуси текстів творів Оксани Забужко («Польові дослідження з українського сексу», «Казка про калинову сопілку») та Сергія Жадана («Інтернат»), обсягом по 50 тисяч слововживань кожен. Використовуючи спеціально розроблене програмне забезпечення на основі мови Python, була виконана автоматична токенизація текстів, їх морфологічне тегування, формування частотних словників словоформ, лем і частин мови, а також створені бази даних для подальшого статистичного аналізу. Це дало можливість забезпечити високу точність, швидкість та об'єктивність у роботі з лінгвістичним матеріалом.

Результати аналізу електронних частотних словників свідчать про домінування іменників у текстах обох авторів, хоча їх функціональна роль різниться. Стиль Оксани Забужко характеризується високою долею іменників та дієприслівників, що вказує на виражену концептуальність, деталізованість і складну синтаксичну побудову її текстів. Натомість твори Сергія Жадана демонструють значно більшу частотність дієслів, що є показником динамічності мовлення, орієнтації на розвиток дій, руху та подій. Проведені статистичні розрахунки підтвердили надійність частотних характеристик і репрезентативність вибірок текстових даних.

Побудова полігонів частот та довірчих інтервалів дозволила встановити специфіку розподілу частин мови в обох корпусах. Згідно з отриманими даними, найбільш стабільними граматичними категоріями є іменники, прикметники, займенники та службові слова. Водночас дієприслівники та числівники мають найвищий рівень варіативності через їх низьку частотність і нерівномірний розподіл у художніх текстах. Аналіз довірчих інтервалів також виявив суттєві відмінності в морфологічній структурі текстів обох авторів.

Стилеметричне індексування дозволило кількісно описати особливості мовної манери кожного з письменників. У творах Оксани Забужко спостерігається значна лексична багатогранність, унікальність словника, переважання номінативності та високий рівень когезії тексту. Це свідчить про інтелектуалізований, рефлексивний і глибоко аналітичний характер її творчості. З іншого боку, тексти Сергія Жадана виділяються вищими

показниками динамічності, коефіцієнтів Трейгера, опредметнення дії та рівня експресивної напруженості, що підкреслює енергійність і акцент на процесуальності його художньої мови.

Проведене дослідження підтвердило дієвість методів комп'ютерної лінгвістики, лінгвостатистики та стилеметрії у вивченні індивідуально-авторських характеристик художнього тексту. Розроблене програмне забезпечення та отримані результати мають потенціал для використання в подальших дослідженнях, пов'язаних з автоматичною атрибуцією текстів, корпусною лінгвістикою, комп'ютерною стилеметрією, а також створенням електронних словників. Виконана робота засвідчила, що морфологічні статистичні показники є ефективним і надійним інструментом для моделювання та порівняння ідіостилів сучасних українських письменників.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Андріяш, О. Ідіостиль Анатолія Мойсієнка : мовний аспект // Дивослово. — 2011. — № 9. — С. 36–39.
2. Антошина, М., Зайцева, В. Метафора в художньому дискурсі Сергія Жадана // Ukrainian sense. — 2023. — № 1. — С. 111–123.
3. Арешенков, Ю. О. Стилїстика української мови : конспект лекцій. — Кривий Ріг, 2007.
4. Бук, С. Корпусно-лексикографічні та лінгвостатистичні виміри великої прози Івана Франка. — Київ, 2021.
5. Бук, С. Статистичні характеристики лексики основних функціональних стилів української мови // Лексикографічний бюлетень. — 2006. — Вип. 13. — С. 166–170.
6. Волошук, В. І. Індивідуальний авторський стиль, ідіолект, ідіостиль: питання термінології // Філологія. Мовознавство. — 2008. — № 79. — С. 5–8.
7. Грида В.С. Ідіостиль Сергія Жадана (на матеріалі роману “Інтернат”). Харків, ХНПУ імені Г.С. Сковороди. 2023.
8. Гриценко П. Ідіолект і текст // Лінгвостилїстика: об’єкт – стиль, мета – оцінка : зб. наук. пр. — К., 2007. — С. 16–43.
9. Дідух Х. І. Ідіостиль як відображення авторської картини світу. Філологічні науки. Серія «Риторика і стилїстика». 2012. URL: https://www.rusnauka.com/15_NNM_2012/Philologia/2_111114.doc.htm
10. Должикова Т. І. Своєрідність художнього мовомислення Сергія Жадана / Т. І. Должикова. Лінгвістика. 2013. № 2. С. 131-136. URL: http://nbuv.gov.ua/UJRN/ling_2013_2_20
11. Єрмоленко С. Я. Нариси з української словесності. Стилїстика та культура мови. — К. : Довіра, 1999. — 431 с.
12. Зубань, О. (2020). Комп’ютерне лексикографічне моделювання морфемної системи української мови. ВПЦ "Київський університет". <https://doi.org/10.17721/978-966-439-819-7>
13. Зубань, О., & Кривенок, В. (2024). Від слова до цифри: лінгвостатистичний «портрет» ідіостилю Степана Бандери. Українське мовознавство, 1(54), С. 222–254. [https://doi.org/10.17721/um/54\(2024\).222-254](https://doi.org/10.17721/um/54(2024).222-254)
14. Калимон Ю. О., Кульчицький І. М., Ліхнякевич. І. О. Ідіолект, ідіостиль, індивідуальний стиль. Тотожне чи різне? Науковий вісник Східноєвропейського національного університету імені Лесі Українки 5, Луцьк, 2014. С. 226-229
15. Капура О. М. Ознаки постмодернізму у прозових творах Оксани Забужко. Науковий вісник Міжнародного гуманітарного університету. Сер.: Філологія. № 60 Том 1. Одеса, 2023. С. 115.
16. Коломієць І. І. Стилїстика української мови : навчально-методичний посібник для студентів факультету української філології / І. І.

- Коломієць. — Умань : Візаві, 2019. — 240 с. — Режим доступу: <https://dspace.udpu.edu.ua/bitstream/123456789/11466/1/Stylistyka%20Ukrainskoi%20Movy.pdf> (дата звернення: 11.04.2025).
17. Кононенко В. Український новостиль: лінгвопоетика художнього дискурсу Сергія Жадана / В. Кононенко. Українознавчі студії. 2018. № 19. С. 17-31. URL: http://nbuv.gov.ua/UJRN/Us_2018_19_4
 18. Корнієнко А. І., Бугайова А. С. Ідіостиль автора: мовно-літературознавчий аспект / Науковий вісник Міжнародного гуманітарного університету. Серія : Філологія. Вип. 25(1). Миколаїв, 2016 С. 36-38.
 19. Костецька, О. П. (2014). Індивідуальне мовлення автора як об'єкт лінгвістики та підходи до його дослідження. Наукові записки [Національного університету «Острозька академія»]. Серія: Філологічна, 49, Київ. 2014. С. 196–199.
 20. Кравець Л. В. Стилїстика української мови: Практикум. — К. : Вища школа, 2004. — 199 с.
 21. Краснобаєва-Чорна Ж. В. Квантитативні методи в лінгвістиці: новітні тенденції / Науковий вісник Дрогобицького державного педагогічного університету імені Івана Франка. Сер. : Філологічні науки (мовознавство). 2018. № 9. С. 93-97. URL: http://nbuv.gov.ua/UJRN/nvddpufm_2018_9_23
 22. Крупка М. Художні версії любовних драм у сучасному прозовому дискурсі (романи О. Забужко, І. Карпи, С. Пиркало). Слово і час. 2007. № 7. С. 73-79.
 23. Кульчицький І. М., Гнатовська К. І. Дослідження ідіолекту письменника: корпусно-статистичний підхід / Молодий вчений. 2019. № 11(2). С. 968-970. URL: http://nbuv.gov.ua/UJRN/molv_2019_11%282%29_143
 24. Лісничка А. Феміністичні ідеї у творчості Оксани Забужко. «Казка про калинову сопілку» як вияв феміністичних ідей. - 7 с.
 25. Лотоцька Н. Я., Мох Р. Н. Специфіка ідіолекту Марії Матіос: статистичний профіль. Закарпатські філологічні студії. Ужгород, 2023. Вип. 29. Т. 1. С. 64–70.
 26. Перебийніс В. С. (Ред.). (1967). Статистичні параметри стилів. Наукова думка.
 27. Перебийніс, В., Муравицька, М., Дарчук, Н. (1985). Частотні словники та їх використання. Наукова думка.
 28. Переломова О. С. Ідіостиль Валерія Шевчука : дис. ... канд. філол. наук : 10.02.01 / О. С. Переломова; Сумський державний університет. — Суми, 2002. — 177 с.
 29. Пономарів О. Д. Стилїстика сучасної української мови : підручник / О. Д. Пономарів. — 3-тє вид., перероб. і допов. — Тернопіль : Навчальна книга – Богдан, 2000. — 248 с.

- 30.Селіванова О. О. Сучасна лінгвістика: напрями і проблеми. — Полтава : Довкілля-К, 2008. — 712 с.
- 31.Ставицька Л. О. Про термін “ідіолект” // Українська мова. — 2009. — № 4. — С. 3–17.
- 32.Ткаченко О. О. Мовна особистість і лінгвокультурологія: навч. посібник. — К. : Центр навчальної літератури, 2003. — 240 с.
- 33.Фадєєва О. В. Ідіостиль як відображення авторської картини світу у творах гібридного жанру. Вчені записки ТНУ імені В. І. Вернадського. Серія: Філологія. Соціальні комунікації. Том 30 (69). № 2 Ч. 1. Кременчук, 2019. С. 138–145.
- 34.Феномен творчості Оксани Забужко : інтелект-реліз / авт. укладач М. М. Максименко. – Полтава, 2015 . – 24 с.
- 35.ЧС СУХП – Частотний словник сучасної української художньої прози / за ред. В. С. Перебийніс. Київ: Наукова думка, 1981. Т. 1, 2.
- 36.Шитик Л., Щербина Т. Методи квантитативної лінгвістики в дослідженні ідіостилю письменника (на матеріалі творів Люко Дашвар). *Linguistic Bulletin*. 2024. Вип. 37. С. 102–111.
- 37.Barlow, M. *Individual usage: a corpus-based study of idiolects*. University of Auckland.
- 38.Bloomfield L. *Language* London, 1973. 566 p.
- 39.Chernysh A. *Linguosynergetic Vector of Studying Contemporary Ukrainian Discourse (On the Material of “Voroshilograd” by Serhiy Zhadan*. Sumy State University, Ukraine, 2024.
- 40.Coulthard M. R. *Author Identification, Idiolect and Linguistic Uniqueness*. 2004. С. 431–447
- 41.Coulthard M. R., Johnson A. *An introduction to forensic linguistics: language in evidence*. London: Routledge. 2007.
- 42.Darchuk, N., Zuban, O., Robeiko, V., Tsyhvintseva, Y., Sorokin, V., & Sazhok, M. (2024). The system for automatic stylometric analysis of Ukrainian media texts TextAttributor 1.0 (techniques, means, functionality). *Acta Linguistica Lithuanica*, 91, С. 224–247. <https://journals.lki.lt/actalinguisticalithuanica/article/view/2328>
- 43.Eder M. *Does Size Matter? Authorship Attribution, Small Samples, Big Problem*. Krakow, Poland. 2010.
- 44.Hrymashevych H. I. Homogeneous members of a sentence in the syntactic organization of prose texts from the collection "Arabesques" by Serhiy Zhadan. *Zhytomyr Ivan Franko State University Journal. Philological Sciences*. Vol. 1 (104)
- 45.Karasov V., Levchenko O. ANGER Conceptual Metaphors in Literary Texts by Oksana Zabuzhko and Halyna. Pahutyak. *COLINS* (2). 2023. P. 436–448.

46. Karasov V., Levchenko O. Quantitative markers of idiolect: stylometric analysis of Oksana Zabuzhko's literary texts. *Актуальні питання гуманітарних наук*. Вип. 91, том 1, 2025
47. Karasov V., Levchenko O. Statistical profile of O. Zabuzhko's idiostyle. Lviv Polytechnic National University, 2024.
48. Kestemont, M. Towards a Better Understanding of Burrows's Delta in Literary Authorship Attribution. Erlangen, Germany. 2014. 79-88 p.
49. Nguyen, T.-N., & Dinh, D. (2021). A Stylometric Investigation of Linguistic Styles Based on a Vietnamese Corpus. *Open Journal of Social Sciences*, 9, 74-87. <https://doi.org/10.4236/jss.2021.912006>
50. *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, Vol. 34, No. 2. (1966), pp. 277-279.
51. Stamatatos, E. Authorship Attribution Using Text Distortion *Journal of the Association for Information Science and Technology*, Karlovassi, Samos, Greece, 2018, 69(3).
52. Wright, D. *Idiolect*. Oxford Bibliographies in Linguistics. New York: Oxford University Press. 2018.
53. **re (Regular Expression Operations)** — стандартний модуль Python для роботи з регулярними виразами: <https://docs.python.org/3/library/re.html>
54. **tokenize_uk** — бібліотека для токенизації українськомовних текстів: <https://github.com/lang-uk/tokenize-uk>
55. **pymorphy3** — бібліотека морфологічного аналізу: <https://github.com/no-plagiarism/pymorphy3>
56. **sqlite3** — стандартний модуль Python для роботи з базами даних SQLite: <https://docs.python.org/3/library/sqlite3.html>
57. **collections** — стандартний модуль Python, що містить спеціалізовані структури даних (Counter, defaultdict тощо): <https://docs.python.org/3/library/collections.html>

ДОДАТКИ

Додаток А. Текстовий корпус творів Оксани Забужко «Польові дослідження з українського сексу» та «Казка про калинову сопілку», використаний для стилеметричного дослідження.

https://drive.google.com/file/d/1sLTM1S4F8D1DJ0zKcrr_g7D6cu0fqGQW/view?usp=sharing

Додаток Б. Текстовий корпус роману Сергія Жадана «Інтернат», використаний для статистичного аналізу.

https://drive.google.com/file/d/1lhREqdXjKdbxG3bH2lm9jmw1pJlW_paL/view?usp=drive_link

Додаток В. База даних SQLite, що містить результати автоматичного аналізу текстового корпусу Оксани Забужко, зокрема частотні словники словоформ, лем та частин мови.

https://drive.google.com/file/d/1Atv_ktsN1ibe7NjcgclQpx8y1SHidunx/view?usp=sharing

Додаток Г. База даних SQLite, що містить результати автоматичного аналізу текстового корпусу Сергія Жадана, зокрема частотні словники словоформ, лем та частин мови.

https://drive.google.com/file/d/14Vkpwl1m7rLpl_yUDtoWLvTa2K-AoPBFo/view?usp=sharing

Додаток І. Основний програмний код, розроблений для автоматичного укладання частотних словників, морфологічного аналізу текстів та формування баз даних.

https://drive.google.com/file/d/1Zh9q70CcZR_5k6RUm_tHdvXbmACsJNyC/view?usp=drive_link

Додаток Д. Програмний скрипт, розроблений мовою Python для автоматичного обчислення відносної похибки частотних показників частин мови, а також формування підсумкових таблиць результатів у форматі Microsoft Excel.

https://drive.google.com/file/d/1Z-63WTwW10fufl0ITqCwGFGEK_1Svom/view?usp=drive_link

Додаток Е. Результати обчислення відносної похибки частотних показників частин мови для корпусів творів Оксани Забужко та Сергія Жадана.

https://docs.google.com/spreadsheets/d/1_h1sGy0kecinzG4LCtSP7QQxnQahdX44/edit?usp=drive_link&oid=101273868934437848362&rtpof=true&sd=true

Додаток Є. Програмний скрипт, розроблений мовою Python для автоматизованої побудови полігонів частот на основі варіаційних та інтервальних рядів частин мови.

https://drive.google.com/drive/folders/16l2M36nm8HBTpCAuKoXsCugCNXqwV0ef?usp=drive_link

Додаток Ж. Графічні матеріали, що містять побудовані полігони частот для різних частин мови та використовувалися під час стилеметричного аналізу.

https://docs.google.com/spreadsheets/d/15gUatehpPsis9-78-oOwwhj5uOWwEKfY/edit?usp=drive_link&oid=101273868934437848362&rtpof=true&sd=true

Додаток З. Програмний код для обчислення довірчих інтервалів середньої частоти частин мови.

https://drive.google.com/file/d/1BQ-s7PrLL597AzQp11gsd4N41aJWqRpF/view?usp=drive_link

Додаток И. Графіки довірчих інтервалів середньої частоти частин мови, використані для зіставлення статистичних характеристик авторських корпусів.

https://docs.google.com/document/d/1lOBXB2Hc0ZPm9WTW5ZgDyzfhuSZ-Do24/edit?usp=drive_link&oid=101273868934437848362&rtpof=true&sd=true

Додаток І. Програмний скрипт, розроблений мовою Python для автоматизованого обчислення лінгвостатистичних та психолінгвістичних індексів текстів на основі морфологічного аналізу засобами бібліотеки Rymorphy3.

https://drive.google.com/file/d/1Gt4m8kuF1xYTdxQeY4cYgghyvnQmJYCd/view?usp=drive_link

Додаток Ї. Результати автоматизованого обчислення лінгвостатистичних та психолінгвістичних індексів для корпусів творів Оксани Забужко та Сергія Жадана.

https://drive.google.com/file/d/1N2jTI3kyf9-0SsmumVHCnZ4YF9M-Bles/view?usp=drive_link