

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

**Статистичний аналіз корпусу новітніх поетичних
та пісенних текстів воєнного часу**

Кваліфікаційна робота

освітнього ступеня «бакалавр»
за спеціальністю 035 «Філологія»,
спеціалізацією 035.10 «Прикладна
лінгвістика»,
галузі знань 03 «гуманітарні науки»
ОПП «Прикладна (комп'ютерна)
лінгвістика та англійська мова»
студентки IV курсу

Поліни КАЖАНОВОЇ

Науковий керівник:

доктор філософії, доц. Юлія
ЦИГВІНЦЕВА

Рецензент:

к.філол.н. Ірина САМОЙЛОВА

«Допущено до захисту»

Протокол № 11 засідання кафедри
української мови та прикладної лінгвістики

ННІФ від 01.06.2023

Завідувач кафедри _____ **Сергій Різник**

КИЇВ – 2023

ЗМІСТ

ВСТУП.....	3
РОЗДІЛ I. СТАТИСТИЧНИЙ АНАЛІЗ У ЛІНГВІСТИЦІ.....	7
1.1. Об'єкт, предмет та основні поняття статистичної лінгвістики.....	7
1.2. Методи статистичного аналізу корпусу текстів.....	11
1.3. Частотний словник та методика його укладання.....	13
Висновки до розділу 1.....	16
РОЗДІЛ II. ДОСЛІДЖЕННЯ КОРПУСУ НОВІТНІХ ПОЕТИЧНИХ ТА ПІСЕННИХ ТЕКСТІВ ВОЄННОГО ЧАСУ.....	18
2.1. Специфіка та параметризація корпусу.....	18
2.2. Корпусно-статистичний аналіз лексики поетичної мови.....	26
2.3. Створення частотного словника сучасної поетичної мови.....	29
2.4. Порівняльний аналіз довоєнних віршів та поезії, написаної під час війни.....	33
Висновки до розділу 2.....	43
ВИСНОВКИ.....	44
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	47
ДОДАТКИ.....	52

ВСТУП

Поезія — це «мистецтво художнього відображення дійсності в словесних образах» [39, с. 764]. Вірші відтворюють не лише настрої та емоції автора, а й його світогляд, реакцію на події, свідком яких той є.

Вони здатні передати всю складність внутрішнього світу особистості й привернути увагу до певної теми чи проблеми. Коли поет виражає свої почуття словами та структурою вірша, він запрошує читачів у подорож своєю душею. Проте поезія відображає не лише особисті переживання поета, а й суспільні настрої. Вона часто є голосом колективної свідомості, може надихати на боротьбу, гуртувати людей і бути рушієм змін.

Сьогодні вірші стають свідками російсько-української війни. Під час конфлікту поети висловлюють свої почуття та страхи щодо війни, страждань і руйнувань, які вона приносить. Вірші про війну можуть бути гірким і приголомшливим свідченням жахів війни. Вони також можуть відображати героїзм і співчуття до тих, хто страждає. Вірші про війну відтворюють історію та залишаються у спадок майбутнім поколінням гірким уроком пам'яті.

Вірші мають унікальну здатність бути понад часом, залишатися актуальними. Вони дозволяють читачам поглянути на емоції та настрої людей, а також на соціальні та історичні контексти, зрозуміти та оцінити складність людського досвіду та дають можливість ділитися ним та вчитися в інших. Вірші є потужним інструментом для передачі та збереження історії та культури. Вони допомагають зрозуміти світ і самих себе через об'єктивне та суб'єктивне читання.

Глибше зрозуміти зміст поетичного тексту допомагає вивчення його мови на всіх лінгвістичних рівнях. Сучасний лінгвістичний аналіз оперує багатьма методами, з-поміж яких — і квантитативні. Один із них, статистичний аналіз, є важливим методом прикладної лінгвістики. Він надає дослідникам інструменти для аналізу та інтерпретації лінгвістичних даних, допомагає проникнути у властивості мовних одиниць, зрозуміти різні аспекти мови. Застосовуючи

статистичні методи, лінгвісти отримують об'єктивні та кількісно обґрунтовані результати досліджень, можуть опрацювати значні обсяги текстів, виявляти повторювані шаблони і досліджувати, як мова змінюється в часі та просторі. Статистичний аналіз віршів употужнює літературознавчий та дозволяє не лише комплексно описати мову, а й краще зрозуміти образність та ідею поетичних творів.

Статистичний аналіз досліджували в своїх наукових працях В. Перебийніс [34; 35], С. Бук [6; 7; 8], Д. Занько [18], Г. Ситар [37; 38], Ж. Краснобаєва-Чорна [24], В. Теслюк [40; 41]. Корпусно-статистичним дослідженням присвячені наукові розвідки Х. Борецької [4], Н. Дарчук та В. Сорокіна [11].

Актуальність статистичного аналізу корпусу новітніх поетичних та пісенних текстів воєнного часу зумовлена тим, що це дозволяє виявити та глибше проаналізувати мовну специфіку цих текстів, особливості поетичної мови, вивчити стильові уподобання та індивідуальні риси творчої манери поетів, виявити відмінності та схожості у використанні мовних засобів, стилістичних прийомів та тематиці, зрозуміти мовленнєву специфіку цього періоду тощо.

Новизна дослідження полягає в тому, що новітні поетичні та пісенні тексти воєнного часу як репрезентанти світовідчуття та мовної картини світу українців ще не були предметом глибокого та комплексного вивчення.

Метою роботи є корпусно-статистичне дослідження лексики поетичних текстів воєнного часу та поетичної мови загалом.

Відповідно до мети в роботі передбачено розв'язати такі **завдання**:

- визначити загальні теоретичні положення статистичного аналізу в лінгвістиці та його значення;
- пояснити та обґрунтувати важливість корпусів текстів у дослідженні мови;
- з'ясувати актуальність воєнних віршів в корпусі;
- проанотувати корпус досліджуваних текстів;

- проаналізувати статистично новітні поетичні та пісенні тексти воєнного часу;
- створити частотний словник сучасної поетичної мови;
- порівняти, як змінився стиль поезії одного автора під час воєнного стану та мирного життя;
- написати сайт, де будуть представлені всі вірші корпусу.

Об'єктом дослідження є новітні поетичні та пісенні тексти воєнного часу, а **предметом** — мовні особливості поетичних текстів та їх статистичні характеристики.

У процесі роботи було застосовано такі **методи**:

- метод ключових слів, який охоплює низку процедур;
- аналітичний, індукція, дедукція — для визначення сутності та головних завдань статистичного аналізу;
- метод попереднього опрацювання тексту — послідовність операцій із текстами для приготування їх до додання в корпус;
- метод контекстного аналізу — з'ясування особливостей вживання слів у конкретних контекстах;
- статистичний та лінгвістичний аналізи — для дослідження поетичної мови та інтерпретації результатів.

Матеріалом роботи є новітні поетичні та пісенні тексти воєнного часу, опубліковані у 2022-2023 роках.

Практичне значення роботи полягає у створенні корпусу новітніх поетичних та пісенних текстів та частотного словника на основі нього, які можна буде доповнювати та використовувати для дальших досліджень із прикладної лінгвістики, лексикології, лексикографії, літературознавства та інших дисциплін.

Структура роботи. Кваліфікаційна робота складається зі вступу, двох розділів (7 підрозділів), висновків до розділів, загальних висновків, списку використаних джерел, додатків.

У вступі обґрунтовано актуальність теми, сформульовано мету, завдання, об'єкт та предмет дослідження, окреслено теоретико-методологічну основу, новизну та практичне значення роботи, описано її структуру.

У першому розділі визначено основні теоретичні засади статистичного аналізу в лінгвістиці, окреслено його розвиток в Україні, описано методи статистичного аналізу корпусу текстів та методику укладання частотного словника.

Другий розділ присвячено параметризації корпусу новітніх поетичних та пісенних текстів воєнного часу, аналізу його специфіки, корпусно-статистичного аналізу лексики поетичної мови, особливостям створення частотного словника сучасної поетичної мови. Висновки узагальнюють теоретичні положення та містять стислий опис результатів аналізу мовної специфіки досліджуваних текстів.

У додатку 1 подано посилання на корпус, у додатку 2 — на таблицю віршів Романа Коляди, у додатку 3 — на таблицю зі статистичним та тематичним аналізом корпусу текстів, у додатку 4 – на таблицю частотності слів, у додатку 5 — на макет сайту із віршами, який буде доопрацьовано надалі.

Загальний обсяг роботи становить 52 сторінки.

РОЗДІЛ І. СТАТИСТИЧНИЙ АНАЛІЗ У ЛІНГВІСТИЦІ

1.1. Об'єкт, предмет та основні поняття статистичної лінгвістики

Статистична лінгвістика (лінгвістична статистика) — це розділ квантитативної лінгвістики, наука, яка вивчає мову за допомогою методів статистики, або «займається виявленням та описом кількісних характеристик та закономірностей мови та мовлення» [7]. Вона працює над аналізом природної мови з використанням математичних моделей і статистичних методів. Статистична лінгвістика вивчає різні лінгвістичні явища, такі як граматики, лексика, семантика, структура речень і тексту, і використовує методи статистичного аналізу для виявлення закономірностей у мові. Приклади завдань, які вирішує статистична лінгвістика, включають машинний переклад, розпізнавання мовлення, класифікацію тексту, тональний аналіз, вилучення інформації тощо.

Результати статистичної лінгвістики широко використовують в галузі комп'ютерної лінгвістики, яка передбачає розробку алгоритмів і програм для обробки природної мови за допомогою комп'ютерів. З розвитком інтернету статистична лінгвістика стала особливо важливою, оскільки вона допомагає обробляти та аналізувати велику кількість текстової інформації, необхідної багатьом додаткам, таким як пошукові системи, соціальні мережі, машинний переклад тощо.

Статистична лінгвістика також використовується в інших сферах, таких як маркетинг і реклама, де її можна використовувати для аналізу настроїв споживачів і визначення найбільш ефективних способів привернути увагу до продукту.

Однією з важливих праць українських науковців на тему статистичного аналізу є посібник «Статистичні методи для лінгвістів» (2013) Валентини Перебийніс [34]. У цій праці авторка розглядає питання статистичного аналізу, досліджуючи та пояснюючи передумови правильної організації статистичного

дослідження. Вона пропонує методи та показники статистичного аналізу для виявлення стилістичних рис та характерних особливостей поезії.

Необхідним для нашого дослідження є посібник із лінгвістичного аналізу тексту «Лінгвістичний аналіз тексту» (2008) Ірини Кочан [23]. У ньому описано основні аспекти лінгвістичного аналізу тексту та подано методологічні підходи до аналізу текстових структур. У своєму посібнику І. Кочан розглядає різні рівні аналізу тексту, такі як фонетичний, морфологічний, синтаксичний та семантичний. Вона описує основні концепції та методи, що використовуються в лінгвістичному аналізі тексту, а також наводить приклади застосування цих методів до різних типів текстів. Посібник є важливим джерелом для вивчення різних аспектів лінгвістичного аналізу тексту.

Також у цьому контексті варто згадати навчально-методичний посібник «Основи статистичної лінгвістики» (2008) [7] та статті Соломії Бук, зокрема «Статистичні характеристики роману Івана Франка “Основи суспільності” (на основі частотного словника твору)» [8], яка ілюструє практичні інструменти здійснення статистичного аналізу. У своїй роботі ми орієнтувалися на розроблену С. Бук методичку комплексного квантитативного опису творів.

Об'єктом дослідження статистичної лінгвістики є природна мова та різні мовні явища, такі як вимова, слова, граматики, семантика тощо. Статистична лінгвістика використовує математичні моделі та статистичні методи для вивчення мовних явищ, аналізу великих обсягів текстової інформації та виявлення мовних закономірностей.

Корпус текстів є однією з основних тем дослідження статистичної лінгвістики. Корпус — це велика колекція текстів певною мовою, яка використовується для вивчення мовних явищ і розробки статистичних моделей. За допомогою корпусу можна дослідити частоту слів і фраз, лексичні особливості текстів, проаналізувати структуру речень і тексту, вивчити граматичні елементи тощо.

Іншим об'єктом статистичної лінгвістики є мовні моделі, які використовуються для опису та прогнозування мовних явищ. Існує багато

різних типів мовних моделей, включаючи моделі n-грам, приховані моделі Маркова, нейронні мережі тощо.

Окрім того, статистична лінгвістика допомагає розв'язувати різні інші лінгвістичні завдання, такі як машинний переклад, розпізнавання мови, класифікація тексту, тональний аналіз, пошук інформації та багато інших. Кожне з цих завдань вимагає різних методів і моделей, але вони базуються на статистичних.

Предметом статистичної лінгвістики є вивчення природної мови за допомогою статистичних методів і математичних моделей. Статистична лінгвістика використовує корпусні тексти та інші джерела даних природної мови для побудови статистичних моделей мови для опису та прогнозування різноманітних мовних явищ.

У науковій літературі статистичну лінгвістику часто розуміють як близьке поняття до математичної лінгвістики, комбінаторної лінгвістики, комп'ютерної лінгвістики, стилеметрії тощо. Ці поняття не є синонімами, хоча іноді перетинаються в предметах свого розгляду. Стилеметрія — галузь, “що виявляє та аналізує кількісні характеристики певного функціонального стилю мови чи мовлення певних письменників” [7] — є розділом статистичної лінгвістики. А статистична лінгвістика є важливою галуззю комп'ютерної лінгвістики, яка займається розробкою алгоритмів та програм для обробки природної мови комп'ютерами.

Статистична лінгвістика пов'язана з різними мовознавчими дисциплінами, бо за допомогою її методів можна аналізувати різні мовні рівні, а також із лінгвістичним аналізом тексту, літературознавством та іншими філологічними і суміжними дисциплінами.

Основними для статистичної лінгвістики є такі поняття:

- Корпус — це набір текстів, які використовуються для вивчення природної мови. Корпуси можна складати з різноманітних джерел, включаючи літературні твори, новинні статті, записи розмов тощо.

- Частота — це кількість разів, коли слово або фраза зустрічається в корпусі. Частота використовується для визначення найпоширеніших слів і фраз у мові.
- Морфологія — це наука про будову слів у мові, зокрема про їх форми та варіанти. Морфологія є важливим аспектом статистичної лінгвістики, тому що це допомагає в розробці алгоритмів автоматичної обробки тексту, таких як сегментація слів і визначення частини мови.
- Синтаксис — наука про будову речення в мові. Статистична лінгвістика використовує статистичні моделі для аналізу синтаксичних структур у тексті, таких як дерева речень.
- Моделювання мови — це розробка математичних моделей, які описують природну мову та її явища. Мовне моделювання можна використовувати для створення систем автоматичного перекладу, аналізу настрою тексту, класифікації тексту тощо
- N-грами — це послідовності з n слів, які зазвичай використовуються в статистичній лінгвістиці для опису мови та її моделей. Наприклад, диграма — це послідовність з двох слів, а триграма — це послідовність з трьох слів.
- Алгоритми машинного навчання — це методи обробки даних, які використовуються для навчання комп'ютерних програм аналізу та інтерпретації природної мови. Алгоритми машинного навчання можна використовувати для створення систем розпізнавання мови, автоматичної класифікації тексту тощо.
- абсолютна та відносна частотність вживання лінгвістичних одиниць (букв, фонем, складів, слів, речень), їх класів (голосних, приголосних, частин мови) і сполучень (буквосполучень, словосполучень);
- закон переваги, закон Ціпфа-Мандель-брота, закон Менцерата (Альтманна) [7].

Ці та інші поняття та можливості роблять статистичну лінгвістику функціональною дисципліною для аналізу, моделювання та розуміння природної мови.

1.2. Методи статистичного аналізу корпусу текстів

Методи статистичного аналізу текстових корпусів є важливими інструментами статистичної лінгвістики. Одним із методів статистичного аналізу є частотний аналіз. Він базується на визначенні частотності слів і словосполучень у текстовому корпусі. Для цього текст розбивають на окремі слова та фрази, а їх входження в корпус підраховуються. За допомогою такого підходу можна ідентифікувати найпоширеніші слова та фрази в мові, а також визначити семантичний зв'язок між ними.

Іншим важливим методом для статистичного аналізу є морфологічний аналіз. У його основі лежить вивчення будови слова та його форми. Для цього текст розбивають на окремі слова, які потім аналізують за допомогою алгоритмів. Цей метод дозволяє визначити словоформу, частину мови та синтаксичні зв'язки між словами в реченні.

Ще одним необхідним інструментом для статистичного методом є синтаксичний аналіз. Він заснований на вивченні будови речення в мові. Для цього текст розбивають на речення, які потім аналізують за допомогою синтаксичних алгоритмів. Цей метод дозволяє визначити будову речень, синтаксичні зв'язки й залежності між словами.

Іншим важливим методом статистичного аналізу є мовне моделювання. В його основі лежить створення математичних моделей, що описують мову та її явища. Для цього використовуються різні статистичні методи, такі як n-грами та алгоритми машинного навчання. Мовне моделювання дозволяє створювати системи автоматичного перекладу, аналізувати настрої тексту тощо. Загалом, методи статистичного аналізу текстових корпусів відіграють важливу роль у статистичній лінгвістиці та необхідні для розуміння й аналізу мови. Окрім того,

ці методи є основою для різноманітних інноваційних технологій, таких як голосові помічники, машинний переклад і розпізнавання мовлення.

Статистичні методи застосовують і для виявлення особливостей функціональних стилів мови та рис стилю окремих авторів, у дешифруванні історичних писемностей, у глотохронології, у вирішенні завдань стенографії, авторської та часової атрибуції твору, для лінгвістичного моніторингу функціонування мови у конкретному типі дискурсу, для контент-аналізу, визначення типів мовних помилок, нових слів та значень, художніх засобів, тематичних груп лексики, особливостей використання в текстах певних граматичних форм чи синтаксичних конструкцій тощо [7].

Однак, як і будь-який метод, статистичний аналіз текстових корпусів має свої обмеження. Наприклад, він не може врахувати контекст, який може вплинути на значення слова чи фрази. Крім того, він не завжди може визначити значення слова чи фрази, що може призвести до неправильного розуміння тексту.

А проте, незважаючи на свої обмеження, методи статистичного аналізу корпусів є потужними інструментами для вивчення мови та створення різноманітних сучасних додатків. Наприклад, голосові помічники, такі як Siri або Alexa, використовують методи статистичного аналізу корпусу, щоб розуміти запити користувачів і відповідати на них. Ці програми можна використовувати для пошуку інформації в інтернеті, нагадування про важливі зустрічі чи завдання, керування домашніми пристроями тощо.

Незважаючи на всі переваги статистичного аналізу текстових корпусів, його використання також пов'язане з багатьма етичними проблемами. Наприклад, його часто використовують для аналізу великих обсягів особистих даних, які можуть порушувати конфіденційність користувачів, як-от публікації в соціальних мережах або електронні листи.

Загалом, методи статистичного аналізу текстових корпусів відіграють важливу роль у різних сферах, пов'язаних з мовою та комунікацією. Вони дозволяють нам краще розуміти мову та всі правила, пов'язані з нею,

створювати нові технології та покращувати якість спілкування. Однак їх використання також вимагає відповідального підходу та дотримання етичних принципів.

Важливим аспектом статистичного аналізу текстових корпусів є також вибір відповідних методів та засобів аналізу даних. Наприклад, для визначення частоти слів і виразів можна використовувати такі методи частотного аналізу, як TF-IDF (інверсний термін документа) або списки частоти слів. Для визначення семантичної близькості між словами та виразами можна використовувати векторні семантичні методи, такі як Word2Vec або GloVe.

Іншим важливим аспектом статистичного аналізу текстових корпусів є обробка та попередня обробка даних. Це охоплює усунення шуму, видалення стоп-слів та лематизацію. Ці процеси допомагають усунути непотрібні елементи та спростити дані, тим самим підвищуючи якість аналізу та точність результатів.

Отже, статистичний корпусний аналіз має широкий спектр застосувань і може використовуватися в різних галузях, таких як традиційна лінгвістика, комп'ютерна лінгвістика, обробка природної мови, штучний інтелект тощо. Він може покращити розуміння мови, допомогти створити нові технології та підвищити якість спілкування. Однак його використання вимагає відповідального підходу та дотримання етичних принципів, щоб уникнути порушення конфіденційності та захисту даних.

1.3. Частотний словник та методика його укладання

Частотний словник у прикладній лінгвістиці — це інструмент, який містить інформацію про частоту вживання слів або мовних одиниць у певному корпусі текстових даних. Це список слів або виразів у порядку зменшення частоти. Частотні словники дозволяють лінгвістам та іншим дослідникам отримати кількісну інформацію про частотність слів і виразів у конкретному мовному матеріалі.

Частота може бути визначена як абсолютна кількість повторень слова чи виразу в корпусі, або як відносне значення, виражене у відсотках, або як значення, нормоване до розміру корпусу.

Словники частоти вживання дозволяють дослідникам ідентифікувати слова, фрази або лексичні одиниці, які найчастіше вживаються в мові. Їх можна використовувати для вивчення лексичного складу, лексичної варіативності, аналізу та порівняння текстів за лексичним складом.

Частотні словники також можуть допомогти визначити ключові слова та терміни, які є найбільш релевантними для конкретного корпусу тексту або для певної галузі знань.

Окрім того, частотні словники можна використовувати для порівняння частоти слів у різних мовах, територіальних чи соціальних діалектах. Це дає змогу дослідникам здійснювати порівняльний аналіз і вивчати лінгвістичні відмінності та подібності з погляду використання мови в тексті. У прикладній лінгвістиці частотні словники можна використовувати для розробки методів вивчення другої мови, комп'ютерної обробки природної мови, машинного перекладу та інших завдань, пов'язаних із мовою. Частотність слів і виразів може слугувати основою для визначення важливості та пріоритетності лексичних елементів при розробці алгоритмів і моделей обробки текстових даних.

Наприклад, на платформі MOVA.info [42] представлені декілька частотних словників:

- частотний словник сучасної української поетичної мови;
- частотний словник українського публіцистичного стилю – 2004;
- частотні словники корпусу;
- частотний словник художньої прози;
- частотний словник сучасної української публіцистики;
- частотний словник наукового стилю тощо.

Частотні словники можна створюватися на основі різних типів текстових корпусів. Корпус може включати письмові тексти, такі як книги, статті, блоги,

новини тощо, а також запис усного мовлення в аудіо- чи відеоформаті. Для створення частотного словника необхідно зібрати досить великий і репрезентативний корпус, щоб охопити різноманітні мовні структури та жанри.

Створення частотного словника складається з кількох кроків.

Першим кроком є збір тексту, що представляє цільову мову. Корпус повинен бути достатньо великим і різноманітним, щоб включати тексти різних типів, стилів і тематики.

Після збору корпусу текстові дані потребують попередньої обробки. Цей етап охоплює видалення певних символів, знаків пунктуації, цифр і тексту в нижньому регістрі. Ці кроки допоможуть забезпечити послідовність даних і уникнути помилок підрахунку.

Потім текст обробляється за допомогою алгоритму токенізації, який розбиває текст на окремі слова чи фрази. Цього можна досягти за допомогою різноманітних методів, включаючи розмежування пробілів або використання складніших алгоритмів поділу тексту.

Після токенізації відбувається лематизація, тобто зведення слова до його основної форми, або лемми. Лематизація сприяє поєднанню різних форм одного і того ж слова та зменшує обсяг словника. Для цього використовують морфологічні аналізатори або словники.

Після токенізації та лематизації необхідно обчислити частоту кожного слова в тексті чи корпусі. Цього можна досягти шляхом створення словника, у якому кожне слово зіставляється з кількістю разів, коли воно зустрічається в тексті чи корпусі. Частота лексеми може бути виражена як абсолютне значення (кількість повторень) або відносне значення (частота у відсотках від загальної кількості слів). Обидва варіанти можна використовувати для аналізу тексту та виявлення найбільш вживаних слів.

Останнім кроком створення частотного словника є розташування слів у порядку зменшення частоти. Це дозволяє виділяти найпоширеніші слова та визначати ключові елементи мови чи типу тексту.

Отриманий частотний словник можна використовувати для різних цілей у прикладній лінгвістиці. Частотний словник також може бути основою для створення лексичних ресурсів, таких як лексико-статистичні бази даних, електронні словники або засоби автоматичної обробки мови.

Однак слід зазначити, що частотні словники мають свої обмеження. Вони можуть недостатньо враховувати контекст і семантику слів, а також не враховувати рідкісні та специфічні лексичні одиниці. Тому важливо використовувати частотні словники в комплексі з іншими методами та засобами вивчення мови.

Отже, частотні словники є цінним інструментом у прикладній лінгвістиці та підтверджують, що при вивченні мови важливо зважати не тільки на якісні характеристики, але й на кількісні параметри, які можуть бути дуже важливим та інформативними для дослідника.

Висновки до розділу 1

Отже, статистична лінгвістика є галуззю лінгвістики, яка використовує методи статистики для аналізу мови. Вона базується на зборі та аналізі великих масивів текстових даних з метою виявлення закономірностей та розуміння мовного явища.

Частотний словник є інструментом, що використовується в прикладній лінгвістиці для вивчення вживання слів в мовному корпусі. Він містить інформацію про частоту вживання слів у певному тексті чи колекції текстів.

Статистична лінгвістика та частотний словник як один із її інструментів відіграють важливу роль у вивченні мови та розв'язанні практичних завдань у сфері мовознавства. Застосування методів статистичної лінгвістики дозволяє зрозуміти особливості мовних одиниць, виявити закономірності у вживанні слів та конструкцій, а також покращити якість автоматичної обробки мови.

Частотний словник, своєю чергою, надає корисні дані для лінгвістичних аналізів, перекладу та стилістичного аналізу текстів.

Отже, використання статистичних методів у прикладній лінгвістиці сприяє розумінню та глибшому аналізу мови, а також допомагає вирішувати практичні завдання, пов'язані з обробкою природної мови. Ці методи можуть бути застосовані в різних галузях, включаючи комп'ютерну лінгвістику, лінгвістичний аналіз та перекладознавство, вдосконалення систем автоматичної обробки мови та інші сфери, пов'язані з вивченням та використанням мови.

РОЗДІЛ II. ДОСЛІДЖЕННЯ КОРПУСУ НОВІТНІХ ПОЕТИЧНИХ ТА ПІСЕННИХ ТЕКСТІВ ВОЄННОГО ЧАСУ

2.1. Специфіка та параметризація корпусу

Корпус новітніх поетичних та пісенних текстів воєнного часу містить 100 віршів, які були обрані за уподобаннями дослідниці. Головним критерієм відбору були власні почуття та емоції, які з'являлися від прочитання того чи іншого вірша.

Сам корпус складається з 5 стовпців: 1 стовпець – вірші, 2 стовпець – тематичний аналіз поезій, 3 стовпець – стилістичний аналіз, 4 стовпець – метрика вірша, 5 стовпець – статистика та аналіз ключових слів, лексики, художніх засобів тощо.

Деяких авторів було знайдено випадково, читаючи новини та різні канали в соціальних мережах. Чиюсь творчість було помічено ще задовго до війни, проте лише в ці моменти людина розкрилася та почала творити емоційну, на нашу думку, поезію. Яскравим прикладом є письменниця Богдана Казанжи, яка лише з початком війни розпочала писати вірші. З деякими людьми ми знайомі з реального життя і вони не зовсім поети, бо мають у своєму творчому арсеналі лише один вірш. Проте він вартий уваги та місця в цьому корпусі текстів воєнного часу. Наприклад, авторка Жанна, з якою ми підтримуємо зв'язок до сьогодні, та її вірш «Я втішала себе до болю». Хтось є медійною людиною та публікує поезії на своїй сторінці в Instagram, а фоловери поширюють їх по всій Україні й таким чином надають розголосу. Одним з таких авторів є Григорій Решетнік, відомий український телеведучий, актор, режисер, диктор, шоумен та громадський діяч. А деякі вірші були знайдені випадково і навіть без авторства.

Також варто відзначити, що більшість авторів (близько 80%) зустрічаються лише раз у цьому корпусі. Це свідчить про різноманітність авторського складу та багатство віршового доробку. Однак, є такі автори, чий

твори представлені більш кількісно, що може свідчити про їхню активну творчу діяльність чи визнання в поетичному середовищі. Такі вірші вирізняються майстерністю, особливістю тематики та довершеністю передавання почуттів, через що й додані до цього корпусу.

У межах проведеного дослідження корпусу статистичного аналізу віршів було виявлено, що вірші Богдани Казанжи та Тоні Гомонюк зустрічаються найчастіше. Це свідчить про їхній визнаний читачами поетичний талант. Їхні твори характеризуються майстерним використанням мовленнєвих засобів, оригінальністю, виразністю та ідейною глибиною. Ці авторки проявляють неабиякий талант у творчому втіленні своїх поетичних ідей, вражаючи читачів своїм вмінням знайти унікальні слова та образи для вираження своїх почуттів та думок.

Корпус містить вірші, написані з 24 лютого 2022 року по 1 травня 2023 року. Використання інтернету як джерела поетичних творів надало широкі можливості для збирання великого обсягу віршів різних авторів і літературних жанрів. Важливо відзначити, що під час збирання віршів з мережі ми дотримувалися принципів етики та авторських прав, забезпечуючи адекватні цитування та покликання на авторів творів.

Тема війни тісно пов'язана з поетичною творчістю, адже війна – одна з найстрашніших і найтрагічніших подій в історії людства, яка має величезний вплив на життя і долю людей. Військові події викликають почуття болю, страждання, втрати та жалоби. Написання віршів про війну може допомогти відтворити й передати глибокі почуття, пов'язані з воєнними подіями, а також висловити своє ставлення до подій. Це може бути важливим не тільки для автора, але й для нагадування людям про жахливі наслідки війни та про те, як вона впливає на життя кожного з нас.

Окрім того, воєнна тематика може бути засобом протесту та вираження невдоволення політикою, що призводить до воєнних конфліктів, згуртовуючи людей навколо спільної ідеї миру та злагоди. До того ж такі вірші можуть слугувати способом виховання молоді в дусі миру та толерантності.

Усі ці вірші об'єднує спільна тема — війна та її наслідки. Вони змальовують біль, трагедію, бажання миру та надії на краще майбутнє після війни. У кожному з віршів автор використовує різні прийоми та мовні засоби, щоб передати свої почуття та співчуття до жертв війни та їхніх рідних, але основна тема залишається незмінною. Зазвичай ці вірші написані у трагічному тоні, щоб викликати емоції та спонукати до роздумів про те, наскільки чудове життя без війни. Один з них розповідає про тугу за Батьківщиною, інший змальовує стан паніки від військової загрози, третій передає біль та страждання, спричинені війною. У четвертому вірші йдеться про те, як війна руйнує звичне життя, створює непередбачувані ситуації, які призводять до незворотних змін та ранять душу. Наступний вірш описує біль та страждання, які викликає війна, а також її наслідки для людей та їхніх родин. Більшість із зібраних поезій зосереджені на трагедії від втрати близьких людей та земляків, передають біль, який переживають родичі загиблих. Також вірші показують, як багато страждань зазнає Україна, які руйнівні наслідки війни і як це впливає на життя звичайних людей. Водночас більшість авторів закликає до відповідальності та засудження катів і злочинців війни.

Обравши 100 текстів для корпусу, було розпочато їхнє опрацюванням. Кожен твір було спочатку проаналізовано за тематикою, розкрито його зміст, ідею. Аналіз вірша на визначення його теми дозволяє зрозуміти головну думку, яку хотів висловити автор у своїх віршах. Тема може бути сформульована явно, але часто вона прихована поміж рядків і вимагає глибшого аналізу, щоб розкрити її. Визначення теми твору допомагає зрозуміти його цілісність, ідейність. Окрім того, аналіз теми може допомогти читачеві осягнути світ, суспільство та людську природу крізь призму творчості автора. Та загалом, це було одним зі способів глибше зрозуміти корпус новітніх поетичних текстів та його значення.

Наступним етапом було проведення статистичного та лінгвістичного аналізу віршів. Вони дали змогу проаналізувати:

1. Метрику вірша: тип метру, кількість складів у рядках, кількість рядків.

2. Ритм: наявність ритму, його тип, місце його розташування в рядку.
3. Частотність слів: які слова найчастіше використовуються у вірші, чи є якісь слова-домінанти, які можуть вказувати на тематику чи настрій вірша.
4. Повторення фраз: які фрази повторюються у вірші, як часто вони зустрічаються, які фрази можуть мати особливе значення або вплив на настрій вірша.
5. Стилiстичні особливості: наявність метафор, алегорій, епітетів, порівнянь, які могли б вказувати на мистецьку цінність вірша та на його авторський стиль. Також було проаналізовано стилістичні прийоми, такі як повторення, анафора, епіфора, паралелізм та інші, за наявності їх у поезії.

Для прикладу розглянемо декілька творів. Вірш Тоні Гомонюк «І ці емоції вже не стримати» описує внутрішній конфлікт та переживання людини, яка живе в умовах війни:

*і ці емоції більше не стримати,
бо хочеться бити, щоб скиглили.
цієї люті більше не втримати,
бо хочеться вбити, щоб вимерли.*

Розчарована дівчина, втомлена від нескінченного болю та страждань, які її оточують, хоче висловити свої емоції та ненависть до ворога. Автор виражає жагу помсти і бажання помститися за все, що вороги зробили з рідною країною. У другому рядку авторка використовує пряме значення “скиглили”, щоб передати своє бажання перемогти ворога до того, що він буде кричати від болю. У четвертому рядку гіперболізацію “вимерли” описує її бажання вбити своїх ворогів до останнього, щоб вони більше не становили загрози для миру та безпеки та щоб їх більше не існувало в цьому світі. У цілому вірш передає душевну напругу і трагізм війни, а також відчуття внутрішньої боротьби, яка відбувається в серці людини, котра живе на війні. Авторка висловлює

важливість небайдужості та відповідальності кожного за свої вчинки під час бойових дій.

Здійснивши лінгвістичний та статистичний аналізи цієї поезії, маємо такі результати:

1. Щодо метрики можна вважати, що це вільний вірш, оскільки немає чіткої ритмічної схеми або використання певного метричного шаблону, такого як ямб або трохей. Замість цього авторка використовує пряму мову та паузи, що підкреслюють силу його емоційної виразності. Відсутність рим та розбіжності в довжині строф також засвідчують його вільний стиль.

2. Найчастіше вживані слова у вірші: *більше, не, щоб, цієї, ці*.

3. У вірші є повторення слів та фраз (*більше не, хочеться, бити, вбити*), що створює ефект посилення емоційного навантаження.

Стилістичні особливості вірша полягають у використанні гострих, агресивних образів (*скиглити, хочеться вбити, щоб вимерли*) та наявності недомовлених речень, що передають емоційний стан ліричного героя.

4. У вірші використовуються слова з високим ступенем емоційності та зарядженості. Наприклад, *лють, вбивство*.

Наступний вірш написаний невідомим автором. Він викликає дуже сильні почуття:

*Ти сидиш у коридорі,
А в чаті пищить “Летить!!!”
І ти щільно, як можеш,
Руками закривши вуха,
Швидко крутиш думки,
Що ж ти ще у житті не встиг?
І куди ця ракета?
В твій дім чи до твого друга?*

Цей вірш щемливо відображає страх і хвилювання за себе та рідних людей в умовах війни. Своє почуття страху та безпорадності автор показує

візуальним образом людини, яка сидить у коридорі, заклавши вуха руками. У чаті лунає повідомлення про ракету, і це посилює страх та тривогу ліричного героя.

У другій частині поеми ліричний герой розмірковує про можливі напрямки польоту ракети, що може призвести до жахливих наслідків. Він намагається зрозуміти, куди летить ракета і що буде з його життям і життям тих, кого він кохає. Це підтверджує найчастотніша лексика та постійні риторичні запитання, які відносяться всі до одного слова — *ракета*.

Ця пісня дуже емоційно описує труднощі, з якими стикаються люди під час війни, і показує, як війна може вплинути на життя людей. Завдяки лексемам та сполукам (*щільно як можеш, пищить, в житті не встиг*) ми можемо відчутти всі переживання героя та автора, які відображають те, що важко підкоритися небезпеці й прийняти її, а також те, що ніколи не знаєш, що трапиться за хвилину з твоїм життям.

Аналіз вірша показав, що:

1. Цей вірш складається з двох строф, кожна з яких містить по чотири рядки. Метрика вірша є вільною, тобто розмір кожного рядка та кількість складів в ньому можуть різнитися. Ритм вірша також вільний, але його утворюють деякі повторювані фрази та візуальні образи, що допомагають змалювати картину.

2. Найчастіше повторювані слова у вірші: *ти, в, і, ракета, життя*. Вони допомагають зосередитися на головному образі вірша — людині, яка переживає небезпеку ракетного обстрілу в своїй країні, та її думках про своє майбутнє, що може змінитися через секунду через клятого сусіда.

3. Фраза з вірша *швидко крутиш думки* підкреслює стан тривоги та паніки, яку переживає головний герой. Також у тексті є візуальний образ ракети, який повторюється у вірші та підкреслює загрозу, яку вона несе.

Частотність вживання певних лексичних одиниць у цьому тексті не є високою, що дозволяє зосередитися на емоційному стані ліричного героя та на загальному настрої вірша — страху перед війною та бажанні мирного життя.

Ще одним яскравим прикладом є вірш «Все це є ти!»:

Все це є ти!

І посмішки , і сльози,

І байрактар, й Азов, і ЗСУ

Бо незалежність – то не здатися загрозі,

І незалежність теж – молитися в тилу.

Все це є ти!

Незламна, неповторна

Як вишиванка з сенсом для душі

Як прапор сили, як тризубу доля

І кожен день в тобі – найкращі то вірші!

Цей вірш має яскраво виражений патріотичний зміст і відображає глибоке почуття до України. Автор зазначає, що все, що стосується збереження незалежності та свободи, варте боротьби та жертв.

У першій строфі відображені кардинально різні емоції — посмішки і сльози. Вона може символізувати не тільки радість перемог, але й гіркоту поразок та втрат. Байрактар, Азов і ЗСУ є найпомітнішими символами української армії та військової могутності, які відображають готовність захищати нашу національну незалежність. Автор зазначає, що незалежність — це не те, від чого можна просто так відмовитися, а те, за що треба боротися і бути готовим померти та пожертвувати найціннішим.

У другій строфі автор порівнює Україну з вишиванкою, яка є унікальною і має глибоке значення для українців. Вона символізує українську культуру та традиції, які необхідно зберегти та передати майбутнім поколінням. Окрім того, автор використовує образ прапора, який символізує силу та єдність. Тризуб — історичний символ України, який нагадує про боротьбу за незалежність.

Остання строфа відображає те, що кожен день українців – найкращий вірш. Це може символізувати те, що кожен день – це нова можливість боротися

за свободу та незалежність. Автор зазначає, що цінності, за які стоїть українська армія, варті жертв і боротьби, адже вони визначають національну ідентичність і майбутнє України.

Загалом вірш насичений українськими традиційними та сучасними визначними символами. Автор використовує художні засоби, які яскраво відображають вільний патріотичний дух. Лінгвістичний аналіз вірша засвідчує, що:

1. Метрика вірша — це суміш різних розмірів. Автор частково використовує ямб, але іноді він підмінюється іншими. Ритм не є регулярним, але в основному плавний та приємний для сприймання.

2. Найчастотніші слова: незалежність, *ти, все, і*.

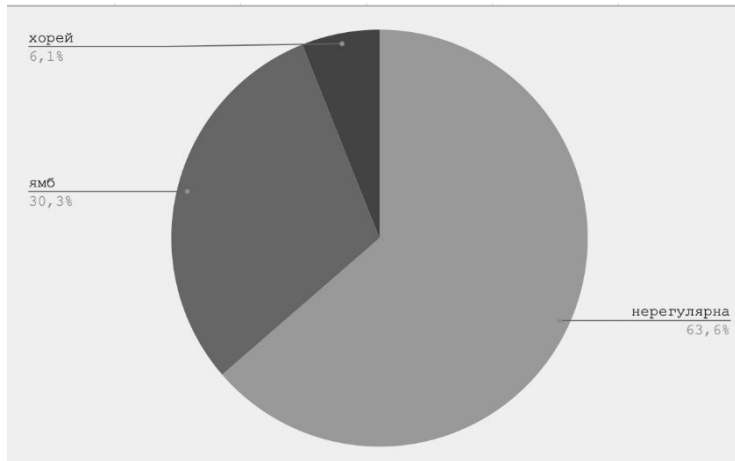
3. Наявні повторення: використання двічі фрази *Все це є ти!*

4. Спостерігаємо такі стилістичні особливості: використання символів української національної культури: *байрактар, Азов, ЗСУ, вишиванка, прапор сили, тризуб*. Якщо зосередитись на частотності вживання певних лексичних одиниць, то найбільш часто вживаються слова, що стосуються людини, ніби наявне звернення до українця, який зображує національну ідентичність та незалежність. Використовуються порівняння, що робить образи більш живими та виразними, повторення, що підкреслює важливість образу “ти”. Також автор застосовує паралелізм, трактуючи різні прояви незалежності (*не здатися загрози і молитися в тилу*.)

Проаналізувавши таким чином усі тексти корпусу, ми побачили, що більшість поетів, пишучи про війну, використовують нерегулярну метрику.

Це можна пояснити різними причинами. Одна з найпоширеніших причин — використання поетом різних розмірів в одному вірші. Наприклад, один вірш може бути написаний ямбом і трохеєм з домішками анапеста.

Зробивши кругову діаграму стосовно метрики, ми отримали не лише цифрову інформацію про корпус, а й візуальне зображення використання системи, за якою розподіляються наголоси в рядках вірша (мал. 2.1).



Малюнок 2.1.1. Метрика віршів

Як бачимо, використання вільного стилю більш поширене в корпусі, бо письменники не дотримуються жорстких правил метрики, а розташовують слова у віршах відповідно до свого бачення і власної майстерності. Тому 63,6% віршів мають нерегулярну метрику й різну кількість складів у рядках, але водночас вони є дуже емоційними та виразними. 30,3% поезій написані ямбом та лише 6,1% — хоресом.

У віршах сформованого корпусу виявлено виразний патріотичний зміст, який відображає глибоке почуття авторів до України та сильний незалежний дух. Лінгвістичний аналіз також розкрив особливості використання різних стилістичних засобів та символів, які відображають українську національну ідентичність та культуру, найчастотніших лексем тощо. Окрім корпусу (додаток 1), ми також створили сайт, де розмістили всі вірші, для того, аби надати до них доступ іншим користувачам. Поки що ця робота не завершена і сайт потребує доопрацювання (додаток 5).

2.2. Корпусно-статистичний аналіз лексики поетичної мови

Статистичний аналіз може бути застосований для дослідження різних мовних рівнів. На нашу думку, у контексті заявленої теми доречно докладніше зупинитися на аналізі лексики. Лексико-семантичний рівень — один із рівнів

мови, який описує словниковий склад і значення слів. Він відображає зв'язок між лексичними одиницями, їх значеннями та сферами використання.

На лексико-семантичному рівні вивчають значення слів, групи лексики за різними параметрами, відтінки значення, виражальну силу слів. На цьому рівні також з'ясовують структуру лексичних одиниць, особливості словотворення та морфологічні ознаки.

Лексико-семантичний рівень важливий для розуміння мови, оскільки він визначає, що означають слова та як вони співвідносяться з іншими словами в мові. Наприклад, щоб правильно зрозуміти текст, потрібно знати значення кожного слова і вміти відрізнити синоніми від антонімів.

На лексико-семантичному рівні також розглядаються можливі лексичні помилки в мові. Наприклад, дехто може вживати слова помилково, обирати невірно синоніми, не точно розрізняти тонкі відтінки значень слів тощо.

Окрім того, на лексико-семантичному рівні досліджують словниковий склад та фразеологічну спадщину мови, що покращує словниковий запас та підвищує ефективність комунікації.

У практичному застосуванні знання лексико-семантичного рівня допомагає лінгвістам досліджувати мову, знаходити нові закономірності та інтерпретувати їх, а письменникам та перекладачам та пересічним мовцям — вільно та точно виражати свої думки та ідеї в мові.

Саме через це ми акцентували увагу на лексико-семантичному рівні. У процесі роботи було проаналізовано сотню віршів, які своєю чергою дали змогу більш чітко зрозуміти сучасну лексику, відношення між словами та те, які лексеми репрезентують тему війни.

Під час дослідження було пораховано ключові слова, питому та запозичену лексику, метафори, епітети, порівняння, художні засоби загалом, нейтральні та забарвлені лексеми, загальноживану лексику та авторські неологізми.

Ключовими словами, що повторюються в багатьох віршах стали іменники *війна* (57 разів), *Україна* (36 разів), *Бог* (37 разів), *світ* і *небо* (34 рази), *життя*

(30 разів), *люди* (22 рази), *біль* (15 разів), *перемога* і *ракета* (14 разів). Ці слова трапляються майже в кожному творі. Звісно, що вони найчастіше зустрічаються, бо ми живемо в світі війни, горя та надії на перемогу.

Звертаючи увагу на питому лексику, яка повторюється в усіх віршах, маємо такі результати: *війна, рідна, душа, біль, сльози, кров, мати, батько*. Використання цих слів засвідчує, що навіть у найгірших ситуаціях ми насамперед думаємо про свою родину.

Запозичена лексика також використовується, проте рідше та не в кожного автора. Наприклад, Ірина Віримчак використовує у своєму творі слово *сирена*, яке наголошує на постійній небезпеці та тривожному стані кожного українця.

Беручи до уваги метафори (*болить душа*), персоніфікації (*скільки сліз пролила Україна*), що повторюються в багатьох віршах, зауважимо, що майже всі зображують біль як фізичну реальність. Зосередивши увагу на епітетах, можемо стверджувати, що повторюється майже в усіх віршах означення *рідний*, наприклад, *українська рідна душа*.

Порівняння в текстах часто стосуються України, наприклад, *Україна як мати, як вишиванка*, або ж взагалі не використовувались. Проте хочеться проілюструвати дуже сильне та болюче, на нашу думку, порівняння. Так, Валентина Матвіїв у своїй поезії пише: *“смерть, як снаряд, вривається уночі”*. Дуже боляче читати ці рядки всім, хто пережив окупацію або обстріл. Стосовно емоційно забарвлених лексем, що повторюються в усіх віршах, то до них належать слова *сльози, біль, батько, мати, душа, дитина, Україна*.

Загальноновживана лексика, що повторюється в усіх віршах, — це слова, які описують емоції, стани та страждання людей, а також назви родинних стосунків та країн (*Україна, Польща, Німеччина*), займенники, дієслова зі значенням повсякденних дій та процесів тощо.

Кожен вірш корпусу торкає найглибші струни душі, адже в кожному рядку бачимо ту неймовірну силу, з якою український народ постає проти зла.

2.3. Створення частотного словника сучасної поетичної мови

Дослідження було спрямоване на аналіз ста віршів, обраних для корпусу новітніх поетичних текстів воєнного часу. Перш за все, ці тексти були оброблені з метою проведення токенізації та підрахунку кількості слів і унікальних слів, а також створення таблиці частотності лексем (додаток 4). Ці кроки сприяли більш детальному аналізу текстів і виявленню основних закономірностей.

На першому етапі дослідження вірші були піддані обробці, яка охоплювала очищення текстів від знаків пунктуації та спеціальних символів, а також розбиття текстів на окремі слова (токени). Цей процес є необхідним для подальшого аналізу текстів на основі окремих лексичних одиниць.

Після обробки віршів проводився підрахунок кількості слів та унікальних слів. Кількість слів визначалася шляхом підрахунку загальної кількості токенів у всьому корпусі. Унікальні слова, які не повторюються, відокремлювалися від загального числа слів.

Далі на основі отриманих даних була створена таблиця частотності лексем. У цій таблиці ми зафіксували кожне унікальне слово, а також вказали його кількість вживань у текстах. Це дозволяло визначити, які слова використовуються найчастіше, а які зустрічаються рідше.

Аналіз таблиці частотності лексем може дати багато цікавої інформації про лексичний склад віршів. Найчастіше вживані слова можуть вказувати на загальну тематику або стилістику текстів. Водночас, рідкісні слова можуть виділятися своєю унікальністю або специфічним контекстуальним значенням. Ці дані та їх аналіз можуть допомогти розуміти характерні особливості віршів, їхню тематику та лексичні особливості. Окрім того, це дослідження може послужити основою для подальших лінгвістичних аналізів і вивчення стилістичних тенденцій у віршах воєнного часу.

Наукове дослідження, що передбачало складання частотного словника зі 100 віршів, написаних під час воєнного часу, виявило цікаві результати.

Дослідження охопило 75 авторів та весь корпус містив 10 010 словоформ, з яких 4291 були унікальними й не повторювалися.

Словоформа	Кількість
і	284
не	276
в	185
я	138
у	128
ти	116
на	114
що	112
з	92
а	85
як	84
за	83
й	74
до	73
щоб	48
та	43
від	43
вже	41
це	38
ми	37
про	37

із	36
все	35
тебе	35
буде	35

Табл. 2.3.1. Фрагмент таблиці частотності слів

Найчастіше зустрічався сполучник *і*, який використовувався у віршах 276 разів. Це закономірно, адже він єднальний та сполучає різні слова й речення, підкреслюючи взаємозв'язок ідей та понять. Також сполучник для створення звукової гармонії поезій.

Заперечна частка *не* трапилася в корпусі 276 разів. Це може свідчити про наявність заперечень, сумнівів або негативних емоцій, пов'язаних із тематикою війни. Автори, імовірно, намагалися виразити протиставлення та підкреслити трагічний аспект війни, її наслідки та страждання людей.

Займенник *я* в початковій формі був використаний 138 разів, а в інших граматичних формах ще 35 разів (усього – 173). Це пов'язано з відображенням особистих переживань та думок авторів щодо війни, намаганням передати особисту долю та почуття в контексті конфлікту. Для поезій взагалі властиве вживання різноманітних займенників.

Скажімо, займенник *ти* в початковій формі зустрівся в корпусі 116 разів, а в інших граматичних формах ще 77 разів (усього – 193). Його вживання свідчить про адресатність поетичних текстів — автори звертаються до конкретних осіб чи уявного читача, якому транслюють свої думки та переживання. Це забезпечує контакт із реципієнтами та залучає їх до співчитання твору.

Лексема *що* в початковій формі зустрічається 112 разів, а в інших граматичних формах ще 22 разів (усього – 134). Вона використовується для введення роздумів, обґрунтувань або пояснення причинових зв'язків. Уживаність цього слова може вказувати на бажання авторів виразити свою

думку, аналізувати події або розмірковувати про воєнні часи. У дальших дослідженнях необхідно буде на основі аналізу контексту та синтаксичних конструкцій розмежувати сполучник *що* й однотипний займенник.

Займенник *ми* в початковій формі був використаний 37 разів, а в інших граматичних формах ще 57 разів (усього – 94). Його вживання свідчить про почуття спільноти, єднання та солідарності ліричного героя/автора з читачами. Він може вказувати на існування групи або колективного досвіду, спільної відповідальності та боротьби.

Іменник *війна* в різних граматичних формах зустрічається в корпусі 57 разів. Очевидно, що використання цього слова, було неминуче для багатьох авторів, щоб акцентувати увагу на стражданнях, втраті та жорстокості, пов'язаних із конфліктом.

Іменник *життя* трапляється в корпусі 30 разів. Це може свідчити про прагнення авторів знайти сенс, цінність та надію в контексті воєнного періоду. Вони намагаються підкреслити важливість життя та його краси навіть у найскладніші часи.

Слова *світ* та *небо* повторювалися по 34 рази. Це свідчить про важливість цих понять для вираження думок та емоцій ліричних героїв. Іменник *світ*, окрім прямого значення, може символізувати надію на мир, спокій та гармонію, а *небо* — безмежність, свободу та духовність.

Ці результати засвідчують глибокі емоційні переживання авторів під час воєнного періоду. Поети намагаються передати свої думки, почуття та досвід через поетичну мову, використовуючи різноматичну лексику для створення змістовних повідомлень та емоційного зв'язку з читачами.

За обсягом частотний словник лексики новітніх поетичних та пісенних текстів воєнного часу — повний, оскільки охоплює не тільки все різноманіття загальноживаної лексики, а й власні назви, цифри тощо. Він складається із двох частин: 1) частотний список словоформ та 2) алфавітно-частотний список словоформ (додаток 4).

Отже, у результаті здійсненої роботи отримано основні статистичні характеристики тексту за зразком методики комплексного квантитативного опису творів С. Бук [8].

1. Обсяг корпусу (N) — 10 010 слововживань.
2. Кількість різних слів (V) — 4291.
3. Багатство словника (індекс різноманітності), тобто відношення обсягу словника лексем до обсягу тексту (V/N), становить 0,42.

4. Середня повторюваність слова в тексті, тобто відношення обсягу тексту до обсягу словника лексем (N/V) — 2,3. Іншими словами, кожне слово в середньому вжито в досліджуваному тексті приблизно 2,3 разів.

5. Кількість *h*арак *l*егомена — слів із частотою 1 — ($V_1 = 3165$) складає 31,61% тексту.

6. Індекс винятковості у тексті, тобто відношення V_1 до обсягу тексту (V_1/N) становить 0,31.

7. Індекс винятковості у словнику, тобто відношення V_1 до обсягу словника (V_1/V), становить 0,73. Два останні числа — показники варіативності лексики.

8. Протилежним до індексу винятковості є індекс концентрації — відношення кількості слів з високою частотою (10 і вище, їх у тексті корпусу 3624) до обсягу тексту ($V_{10}/N=0,36$).

Узагальнюючи, зауважимо, що ці розрахунки надають інформацію про різноманітність та концентрацію слів у текстовому корпусі. Текст має помірний рівень різноманітності слів з великою кількістю унікальних та рідкісних слів. Також виявляється певна повторюваність слів і наявність високочастотних слів у тексті.

2.4. Порівняльний аналіз довоєнних віршів та поезії, написаної під час війни

Інший частотний словник при виконанні кваліфікаційної роботи був укладений на матеріалі чотирьох віршів одного автора з метою порівняння, як

змінився стиль поезії автора та чи відрізняється мова довоєнних віршів та віршів, написаних під час війни. Для аналізу ми обрали вірші Романа Коляди (додаток 2). Кожен вірш налічує приблизно 100 слововживань. Було проведено попередню обробку тексту перед його токенізацією: видалено вручну всі власні назви та дати/цифри. Це було зроблено для того, щоб зменшити кількість слів, що не буде розпізнана токенізатором.

Опис віршів за параметрами лінгвістичної однорідності.

Перші два вірші було обрано про війну. Вони вражають своєю експресивністю та сильним емоційним зарядом. Перший вірш «Нервів більше нема» відображає кризовий стан, коли ліричний герой звертається до Бога та висловлює свої думки та почуття стосовно поточної воєнної ситуації. Вірш «Лютий без березня, квітня і травня» відображає складну та непросту реальність війни, де ліричний герой описує весну, позбавлену миру та спокою. Він звертається до ідеї, що весна, яка зазвичай символізує оновлення та пробудження природи, втрачає свою силу та красу через війну.

Наступні два вірші були написані до війни. Перший вірш «Ти пахнеш яблуком» має романтичний та ліричний характер. Він описує запах яблук та кориці, який має символічне значення в контексті почуттів та таємниці, що переповнюють серце ліричного героя. Вірш «Любов» передає почуття та емоції закоханого чоловіка. Автор зображує кохану особу як недосяжну, таємничу та чарівну. Він використовує природні образи і метафори, щоб описати присутність коханої та її вплив на світ довкола.

Перша вибірка — воєнні вірші:

- 1) метод організації вибірки: зональний. Ця вибірка є однорідною, її укладено на основі тексту одного автора, що належить до одного стилю мовлення. Було взято 215 слововживань.

Друга вибірка — довоєнні вірші:

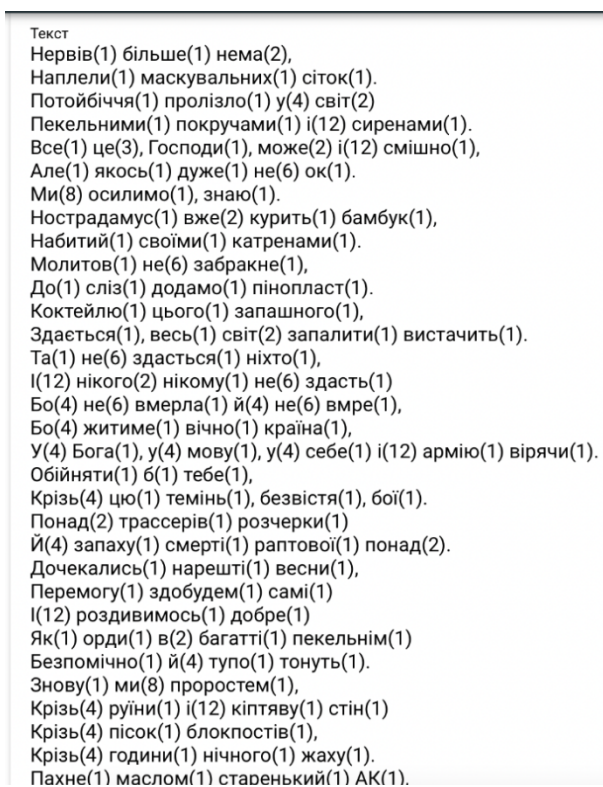
- 1) метод організації вибірки: зональний. Ця вибірка є однорідною, її укладено на основі одного тексту одного стилю. Було взято 227 слововживань.

На жаль, не маючи змоги використовувати мову програмування Python, було зроблено частотний словник інакше. Але цей процес та його результати також були цікаві та інформативні. За допомогою інтернету вийшло зробити частотний словник, а також створити таблицю, використовуючи сайт planetcalc.com.

Перша вибірка (вірші, написані під час війни):

У цій вибірці наявно 162 унікальних слова.

Сам підрахунок слів у тексті виглядає ось так:



Текст

Нервів(1) більше(1) нема(2),
Наптели(1) маскувальних(1) сіток(1).
Потойбіччя(1) пролізло(1) у(4) світ(2)
Пекельними(1) покручами(1) і(12) сиренами(1).
Все(1) це(3), Господи(1), може(2) і(12) смішно(1),
Але(1) якось(1) дуже(1) не(6) ок(1).
Ми(8) осилимо(1), знаю(1).
Нострадамус(1) вже(2) курить(1) бамбук(1),
Набитий(1) своїми(1) катренами(1).
Молитов(1) не(6) забракне(1),
До(1) сліз(1) додамо(1) пінопласт(1).
Коктейлю(1) цього(1) запашного(1),
Здається(1), весь(1) світ(2) запалити(1) вистачить(1).
Та(1) не(6) здасться(1) ніхто(1),
І(12) нікого(2) нікому(1) не(6) здасть(1)
Бо(4) не(6) вмерла(1) й(4) не(6) вмере(1),
Бо(4) житиме(1) вічно(1) країна(1),
У(4) Бога(1), у(4) мову(1), у(4) себе(1) і(12) армію(1) вірячи(1).
Обійняти(1) б(1) тебе(1),
Крізь(4) цю(1) темінь(1), безвістя(1), бої(1).
Понад(2) трассерів(1) розчерки(1)
Й(4) запаху(1) смерті(1) раптової(1) понад(2).
Дочекались(1) нарешті(1) весни(1),
Перемогу(1) здобудем(1) самі(1)
І(12) роздивимось(1) добре(1)
Як(1) орди(1) в(2) багатті(1) пекельнім(1)
Безпомічно(1) й(4) тупо(1) тонуть(1).
Знову(1) ми(8) проростем(1),
Крізь(4) руїни(1) і(12) кіптяву(1) стін(1)
Крізь(4) пісок(1) блокпостів(1),
Крізь(4) години(1) нічного(1) жаху(1).
Пахне(1) маслом(1) ставенький(1) АК(1).

Мал.2.4.1. Скриншот результату аналізу віршів

Сайт, на якому було зроблено дослідження, аналізує вибірку та відразу біля кожного слова надає інформацію про те, скільки разів воно використовується в тексті. Це Такий формат даних був не дуже зручним для роботи, адже за таких умов не було зрозуміло, яке слово використовується

найбільше разів. Тому ми створили таблицю, де чітко показано ці дані.

i	12
ми	8
не	6
у	4
бо	4

Мал. 2.4.2. Скриншот результатів таблиці

Таблицю можна завантажити на телефон або на ноутбук для перегляду.

word	quantity
i	12
ми	8
не	6
бо	4
й	4
у	4
крізь	4
без	3
а	3
стіна	3
це	3
що	3
весна	2
може	2
є	2
нема	2
нікого	2
в	2
світ	2
вже	2
понад	2
лютий	1

Мал. 2.4.3. Скриншот результатів таблиці

Друга вибірка (довоєнні вірші):

У другій вибірці наявно 148 унікальних слова. Також надаємо таблицю підрахунку найбільш частотних слів:

i	12
ти	11
в	6
з	6
не	5

Мал. 2.4.4. Скриншот результатів таблиці

Але табличний формат набагато зручніший:

word	quantity
i	12
ти	11
в	6
з	6
не	5
-	5
яблуком	4
йдеш	4
як	4
кожен	4
крок	4
неба	3
чи	3
що	3
про	3
руки	3
пахнеш	2
корицею	2
серці	2
нас	2
то	2
у	2
це	2
назустріч	2
снам	2
сон	2

Мал. 2.4.5. Скриншот результатів таблиці

Якщо порівнювати вибірки, то у першій вибірці найбільш частотний сполучник *i* (вживається 12 разів), *i* в другій також *i* (вживається 12 разів). Якщо взяти перші 10 слів з першої вибірки, та перші 10 слів з другої вибірки, то можна зробити висновок, що у воєнній поезії здебільшого використовується займенник *ми*, ніби об'єднуючи всіх українців, та частка *не* . Тоді як у віршах, що були написані до війни, більше використовується займенник *ти*, адже

ліричний герой звертається до конкретного адресата, а також іменники *небо*, *дні*, дієслова *пахнеш*, *бриниш*, які не зосереджують увагу на собі. Для прикладу порівняємо два вірші. Перший вірш:

*Нервів більше нема,
Наплели маскувальних сіток.
Потойбіччя пролізло у світ
Пекельними покручами і сиренами.
Все це, Господи, може і смішно,
Але якось дуже не ок.
Ми осилимо, знаю.
Нострадамус вже курить бамбук,
Набитий своїми катренами.
Молитов не забракне,
До сліз додамо пінопласт.
Коктейлю цього запашиного,
Здається, весь світ запалити вистачить.
Та не здасться ніхто,
І нікого нікому не здасть
Бо не вмерла й не вмере,
Бо житиме вічно країна,
У Бога, у мову, у себе і армію вірячи.
Обійняти б тебе,
Крізь цю темінь, безвістя, бої.
Понад трассерів розчерки
Й запаху смерті раптової понад.
Дочекались нарешті весни,
Перемогу здобудем самі
І роздивимось добре
Як орди в багатті пекельнім
Безпомічно й тупо тонуть.*

*Знову ми проростем,
Крізь руїни і кіптяву стін
Крізь пісок блокпостів,
Крізь години нічного жаху.
Пахне маслом старенький АК,
А Господь вже гойдає дзвін
Поховальний над тими,
Що йшли із мечем
І що йдуть легіонами на@уй*

Другий вірш:

*Ти пахнеш яблуком. Яблуком і корицею.
В серці бриниш нерозгаданою таємницею.
Янголи з неба на нас нетутешніми птицями
Дивляться й далі летять із дитячими лицами.*

*Звідки ти, щастя? Неначе Всевишній десницею
Взяв і віддав (ти сказала тоді: «це я!»)
Нині ж очей Своїх всевидючими зіницями
Дивиться з неба, і, то благословить громовицями,
Тихо рахуючи дні з безконечників сплетеною вервіцею,
То заскрипить вітряком чи старою криницею.*

*В лісі галявини вкрилися густо суницями,
Хмарами небо рясніє всуціль білолицими,
Серце дзвенить у любові гартованою крицею...
А ти пахнеш яблуком. Яблуком і трохи корицею.*

Для першого вірша характерний загальний тон песимізму, невпевненості і напруженості. Він описує світ, переповнений негативними емоціями,

боротьбою та хаосом. Автор звертається до Господа і висловлює свої страхи та надії. Вірш містить військові образи, посилення на Нострадамуса і апокаліптичний настрій. Наприкінці автор висловлює свою віру в перемогу і відродження після руйнувань.

Другий вірш має протилежний настрій, він є романтичним і сповненим кохання. Автор описує почуття щастя, таємниці, яскраві природні образи і благодатність навколишнього світу. Він говорить про надію і благословення від Всевишнього. Вірш має більш ліричний і спокійний характер.

Статистично аналізуючи обидва вірші, можна розглянути кілька аспектів, таких як використання частин мови, ритміка і використання художніх засобів.

У першому вірші спостерігається використання більшої кількості прикметників та дієслів, що передають напруженість і хаос (наприклад, *наплести, покручі, безпомічно, тупо тонуть*). Також автор використовує повторення слів (*не здасться, житиме, кризь*) для підкреслення ідей. Ритміка вірша є вільною, без чіткого метру або рими.

У другому вірші значно більше уваги приділено опису природи і почуттів. Використовуються прикметники та іменники, що створюють образи природи (наприклад, *суніці, небо рясніє, птиці*). Вірш римований і ритмічний.

Такі лексичні та ритмічні відмінності сприяють створенню різних художніх образів та інакшого настрою в кожному вірші.

Візьмемо інші два вірші. Перший вірш:

Любов

Ти йдеш, не пригинаючи трави,

Легка і недосяжна. Таємниця

У кожному повороті голови.

Ти сарна, одночасно, і левиця.

Ти йдеш, не залишаючи слідів,

Окрім, хіба що, в серці. Неможлива,

З якихось понадмислимих світів

Ти пролилася, як липнева злива.

*І кожен крок – це крок назустріч снам
І кожен сон – про перший поцілунок.
З усіх зірок миліша та, що нам
Упала в руки, як небес дарунок.*

*Ти йдеш, не розтискаючи руки,
В ній пташенятко – мрія. Чи про мене?
Чи приймуть води вічної ріки
Нас нерозлучних. Як весна священна*

*Ти йдеш наосліп, ти ідеш навспак,
Не так і не туди, де всі чекають.
Ти невблаганна, наче неба знак.
І я його, кохана, прочитаю.*

*І кожен крок – це крок назустріч снам
І кожен сон – про перший поцілунок.
З усіх зірок миліша та, що нам
Упала в руки, як небес дарунок.*

Другий вірш:

*Лютий без березня, квітня і травня:
Такою була весна.
Весна без боїв, без хвилини мовчання
Ще може наснитися нам.
А поки, стіна поміж пеклом і раєм –
Це ми. Ми і є – стіна.
Стіна непорушна, бо зло - безіменне*

*Число, а ми – імена.
Нема поза цим більш нікого й нічого,
Усе, що в нас є – війна.
Бо ми – це свобода, і ми – перемога.
І ми ж – за неї ціна.*

Статистично аналізуючи обидва вірші, можна звернути увагу на кілька аспектів. У першому вірші маємо такі особливості:

- Використання прикметників та іменників, що створюють образи (наприклад, *легка, недосяжна, сарна, левиця*).
- Використання повторень слів і фраз для підкреслення ідей (наприклад, *Ти йдеш, кожен крок - це крок назустріч снам, кожен сон - про перший поцілунок*).
- Використання художніх засобів, таких як порівняння (наприклад, *Ти сарна, одночасно, і левиця*) і метафори (наприклад, *з усіх зірок миліша та, що нам упала в руки, як небес дарунок*).

У другому вірші:

- Використання ритмічних повторень звуків і фраз (*Лютий без березня, квітня і травня, Ми і є стіна, Ми це свобода*).
- Використання асоціативних зв'язків та образів (наприклад, *стіна поміж пеклом і раєм*).
- Використання метонімії (наприклад, *ми імена, ми ж за неї ціна*) для передачі ідей та творення образів.

Отже, у першому вірші автор використовує багато образних слів і фраз, а також художніх засобів. Другий вірш, натомість, складається з простіших фраз, побудований на ритмічних повторах і асоціативних зв'язках. Ці стилістичні відмінності сприяють створенню різних настроїв та вражень у кожному вірші. Перший вірш, який описує кохання, має романтичний характер. Він зосереджується на образі любові, яка є недосяжною, таємничою та пристрасною. Автор використовує образи природи, порівняння та метафори,

щоб передати красу та потужність почуттів. Другий вірш, натомість, за тематикою належить до громадянської лірики. Він описує весну як час безкомпромісної боротьби і перемоги, ліричний герой ідентифікує себе зі стіною, що відокремлює пекло і рай. Поет використовує ритмічні повторення, метонімію, щоб передати силу духу та стійкість у боротьбі за свободу.

Звичайно, для чіткого розмежування та релевантних висновків необхідно проаналізувати більшу кількість поезій обох періодів, проте навіть за такими результатами на основі порівняння малих вибірок можна зробити певні висновки про специфіку використання одним поетом слів та творення образів у різні періоди та для опису різних ситуацій і передавання різних емоцій.

Висновки до розділу 2

У межах дослідження було укладено корпус новітніх поетичних та пісенних текстів воєнного часу, що містить 100 віршів, написаних 75 авторами, і присвячений тематиці війни. Завдяки статистичному аналізу були виявлено закономірності та особливості вживання слів та сполук, а також розкрито специфіку змісту цих віршів.

Статистичний аналіз також допоміг встановити частоту вживання певних слів, висловів та стилістичних засобів, що може свідчити про певні тенденції в мовленні сучасних авторів.

Корпус текстів уможливив створення частотного словника, за допомогою якого було прораховано обсяг корпусу, багатство словника, середню повторюваність слів в тексті, індекс винятковості у тексті та винятковості у словнику та індекс концентрації. Ці розрахунки надають інформацію про різноманітність та концентрацію слів у текстовому корпусі.

За допомогою порівняння віршів Романа Коляди, написаних до війни та під час війни, було виявлено певні зміни в його поетичному стилі та тематичному спрямуванні. Аналіз текстів віршів дозволив виявити, які мовні особливості та тематичні акценти з'явилися в його творчості під впливом воєнних подій.

ВИСНОВКИ

Статистичний аналіз є важливим інструментом лінгвістичних досліджень, оскільки він розкриває мовні тенденції та залежності, дозволяє виявити та проаналізувати глибше лексичну специфіку тексту, виражену частотою вживання слів та сполук, унікальністю, варіантністю, звернути увагу на ключові слова, стилістичні особливості, тропи тощо. Це дає можливість описати реальний стан мови аналізованої сфери, стилю чи автора, а також розкрити зміни у вживанні лексем, дослідити нові контексти. Застосовуючи статистичні методи, лінгвісти отримують об'єктивні та кількісно обґрунтовані результати досліджень та можуть опрацьовувати значні обсяги текстів. Статистичний аналіз віршів употужнює літературознавчий та дозволяє не лише комплексно описати мову, а й краще зрозуміти образність та ідею поетичних творів, індивідуальний стиль автора.

У межах дослідження було укладено корпус новітніх поетичних та пісенних текстів воєнного часу, що містить 100 віршів, написаних 75 авторами, і присвячений тематиці війни. Кожен вірш у цьому корпусі відображає погляд автора на воєнні події. За допомогою статистичного і стилістичного аналізу було виявлено закономірності та особливості вживання слів та сполук, а також розкрито специфіку тематики віршів.

Статистичний аналіз також допоміг встановити частоту вживання певних слів, висловів та стилістичних засобів, що може свідчити про певні тенденції в мовленні сучасних авторів. Корпус текстів уможливив створення частотного словника словоформ та алфавітного списку словоформ, за допомогою яких було прораховано обсяг корпусу, багатство словника, середню повторюваність слів в тексті, індекс винятковості у тексті та винятковості у словнику та індекс концентрації. Ці розрахунки надають інформацію про різноманітність та повторюваність слів у текстовому корпусі.

Зокрема, аналіз показав, що найчастотнішими в корпусі текстів є сполучник *і*, який використовувався у віршах 276 разів, заперечна частка *не*

(276 разів), займенники *я* (138 разів), *ти* (116 разів), *ми* (37 разів), лексема *що* (112 разів), іменники *війна* (57 разів), *Україна* (36 разів), *Бог* (37 разів), *світ* і *небо* (34 рази), *життя* (30 разів), *люди* (22 рази), *біль* (15 разів), *перемога* і *ракета* (14 разів).

У результаті здійсненої роботи отримано такі основні статистичні характеристики тексту:

1. Обсяг корпусу (N) — 10 010 слововживань.
2. Кількість різних словоформ (V) — 4291.
3. Багатство словника (індекс різноманітності), тобто відношення обсягу словника лексем до обсягу тексту (V/N), становить 0,42.
4. Середня повторюваність слова в тексті, тобто відношення обсягу тексту до обсягу словника лексем (N/V) — 2,3. Іншими словами, кожне слово в середньому вжито в досліджуваному тексті приблизно 2,3 разів.
5. Кількість *harax legomena* — слів із частотою 1 — ($V_1 = 3165$) складає 31,61% тексту.
6. Індекс винятковості у тексті, тобто відношення V_1 до обсягу тексту (V_1/N) становить 0,31.
7. Індекс винятковості у словнику, тобто відношення V_1 до обсягу словника (V_1/V), становить 0,73.
8. Протилежним до індексу винятковості є індекс концентрації — відношення кількості слів з високою частотою (10 і вище, їх у тексті корпусу 3624) до обсягу тексту ($V_{10}/N=0,36$).

У поезіях про війну статистичний аналіз разом із контекстним аналізом допоміг виявити вживання емоційно забарвленої лексики, образів, художніх засобів, характерних для цього жанру. Він також розкрив зміни у вживанні мовних засобів під час війни та виявив певні тенденції в лінгвістичному відображенні конфлікту.

Також аби порівняти стиль та мову віршів, написаних до війни та під час воєнних дій, ми проаналізували чотири вірші Романа Коляди. Порівняння віршів письменника, написаних до війни та під час війни, виявило певні зміни в

його поетичному стилі та тематичному спрямуванні. Аналіз текстів віршів дозволив виявити, які мовні особливості та тематичні акценти з'явилися в його творчості під впливом воєнних подій.

Різниця між віршами, написаними до війни та під час війни, полягає у контексті, настрої та способі вираження. Ось декілька ключових відмінностей:

1. Контекст: вірші, написані до війни, відображають мирний час, радість, надію та прагнення до гармонії. Вони проникнуті почуттями любові, краси природи, роздумами про життя. Натомість вірші, написані під час війни, відображають страждання, біль, жах та руйнування. Вони проникнуті тривогою, втратою, відчуттям безпорадності та прагненням миру.
2. Настрій: вірші, написані до війни, мають більш легкий, грайливий, таємничий та романтичний настрій. Вони можуть передавати почуття спокою, любові або натхнення. З іншого боку, вірші, написані під час війни, сповнені глибокою тривогою, гнівом, розпачем та болем. Вони можуть викликати почуття обурення та виступати проти насильства й війни.
3. Спосіб вираження: вірші, написані до війни, більш символічні, метафоричні. Вони можуть використовувати образи природи та художні засоби. З іншого боку, вірші, написані під час війни, більш реалістичні, у них слова вживаються здебільшого в прямих значеннях

Отже, статистичний аналіз є важливим інструментом для вивчення лінгвістичних явищ, глибшого осмислення поетичної мови та чіткішого розуміння поезії загалом. У контексті віршів про війну він допомагає виявити специфічні особливості вираження та передачі емоцій, а також розкрити динаміку сучасного поетичного мовлення, зокрема й пов'язаного з російською агресією.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Арешенков Ю.О. Лінгвістичний аналіз художнього тексту : навч. посіб. для студ. філол. спец. вищ. навч. закл. Кривий Ріг : Видавничий дім, 2007 35 с.
2. Бобкова Т.В. Корпус текстів: Основні аспекти визначення URL: http://www.mova.info/corpus_papers/bobkova-corpus.pdf
3. Богдан С. Методи й методика лінгвостилістичних досліджень : методичні рекомендації для студентів-філологів. Луцьк, 2011. 28 с.
4. Борецька Х. Корпусно-статистичне дослідження лексичних особливостей публіцистичного тексту (на матеріалі газети «день»): Бакалаврська робота: 035 «Філологія». Український католицький університет. Львів: УКУ, 2021. 84 с.
5. Бук С.Н. Корпусно-лексикографічні та лінгвостатистичні виміри великої прози Івана Франка.: автореф. дис. ... д-ра філол. наук : 10.02.01; Київ. нац. ун-т ім. Тараса Шевченка. Київ, 2021. 33 с.
6. Бук С. Квантитативна параметризація текстів Івана Франка: проект та його реалізація. *Вісник Львівського університету*. Серія : Філологічна. 2013. Вип. 58. С. 290-307. URL: http://nbuv.gov.ua/UJRN/Vlnu_fil_2013_58_30.
7. Бук С. Основи статистичної лінгвістики : Навчально-методичний посібник / Відп. ред. проф. Ф.С. Бацевич. Львів : Видавничий центр ЛНУ імені Івана Франка, 2008. 124 с. URL: <https://victana.lviv.ua/biblioteka/98-matematychna-linhvistyka/462-buk-s-n-osnovy-statystychnoi-linhvistyky>
8. Бук С. Статистичні характеристики роману Івана Франка «Основи суспільності» (на основі частотного словника твору). *Вісник Нац. ун-ту «Львівська політехніка»*. Серія «Проблеми української термінології». 2010. № 676. С. 90–93.
9. Білик К.М. Корпусне дослідження текстів рубрики «надзвичайні новини» у французькій, англійській та українській мовах. *Науковий вісник міжнародного гуманітарного університету*. 2019. №43 С.151–155.

10. Брусенцев В.О. конспект лекцій навч. дисципліни, освіт. програма «Інформ. та документац. діяльність» першого (бакалавр.) рівня освіти, спец. 029 «Інформ., бібл. та арх. справа», спеціалізація «Інформ. та документац. діяльність», ф-т соц. комунікацій / М-во культури України, Харків. держ. акад. культури, Каф. інформ. технологій. Харків : ХДАК, 2018. 63 с.

11. Дарчук Н.П., Васильєва І.В., Васильєв О.М. Векторна модель аналізу стилістики текстів досліджень української мови. *Український фізичний журнал*. 2021. Т. 66, № 5. С. 373–378. URL: http://nbuv.gov.ua/UJRN/UPhJ_2021_66_5_4

12. Дарчук Н.П, Сорокін В. Корпус текстів як джерело для мовознавчих і літературознавчих досліджень. *Лексикографічний бюлетень* : Зб. наук. пр. Київ : Ін-т української мови НАН України, 2006. Вип. 13. С. 34–37.

13. Дарчук Н.П. Корпусна лінгвістика : проблеми, методи, перспективи. Київ : навчальна програма. Київський національний університет імені Тараса Шевченка, 2013. 11 с.

14. Демська-Кульчицька О.М. Що таке корпус текстів? *Культура слова*. 2004. № 64. С. 35–38.

15. Демська-Кульчицька О.М. Основи національного корпусу української мови. Київ : Інститут української мови НАНУ, 2005. 219 с.

16. Жуковська В.В. Ресурси корпусної лінгвістики у дослідженні історичної динаміки мови. Слово і речення : синтактика, семантика, прагматика / за ред. В. О. Огнев'юк, Л. Л. Хоружа, О. Є. Бондарева. Київ, 2013. С. 151–156.

17. Зубань О.М. Стилеметричні ознаки морфемних структур слів у поетичному мовленні Т. Шевченка (на матеріалі Корпусу української мови). *Мовні і концептуальні картини світу*, 2014. № 48. С. 165–178.

18. Зубрицький А.Ю. Інтелектуальна система дослідження та аналізу текстів. Магістерська робота зі спеціальності 121 «Інженерія програмного забезпечення». Київ : КПІ, 2019.

19. Занько Д.О. Статистична структура твору Софії Яблонської «Чар Марока». Матеріали III Міжнародної науковопрактичної конференції (м. Одеса,

27-28 листопада 2015 року). Херсон : Видавничий дім «Гельветика», 2015. С. 18–20.

20. Ковальова О. Політичний дискурс : сучасні лінгвістичні інтерпретації. *Актуальні питання гуманітарних наук*. 2020. Вип. 27, Т. 2. С. 101–107.

21. Кондратенко Н.В. Український політичний дискурс : Текстуалізація реальності : Монографія. Одеса: Чорномор'я, 2007. 156 с.

22. Котов Е.В. Дискурсивна природа сучасного політичного процесу: моделювання та прогнозування. Кваліфікаційна наукова праця на правах рукопису. Анотація. URL: <https://abstracts.donnu.edu.ua/article/view/10205>

23. Кочан І.М. Лінгвістичний аналіз тексту : Навч. посіб. 2-ге вид., перероб. і доп. Київ : Знання, 2008. 423 с.

24. 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

25. Кривенко Г.Л. Корпусні дослідження дискурсу: становлення, стан і перспективи. *Вісник Київського національного лінгвістичного університету*. Серія Філологія. 2017. Вип. 20(1). С. 51–63.

26. Кульчицький І.М. Вісник Національного університету «Львівська політехніка». Серія: Інформаційні системи та мережі. Львів : Видавництво Львівської політехніки, 2017. № 872. С. 139–148.

27. Лотоцька Н.Я. Ідіолект Романа Іваничука Корпуснобазований та лінгвокогнітивний підходи. Дис. доктора філософії: 035 Філологія. Львів, 2021. 325 с.

28. Матвеева С.А. Корпус текстів як інструмент обробки мовного матеріалу. Філософія мови та нові тенденції в перекладознавстві й лінгвістиці: збірник матеріалів III Міжнародної науково-практичної конференції. У 2 ч. 2021. Ч. 1. С.150–155.

29. Михайлюк А.Р. Алгоритм укладання корпусу текстів на матеріалі творів Тараса Прохаська. Магістерська робота з прикладної лінгвістики. Вінниця : ДонНУ ім. В.Стуса, 2021. 65 с.
30. Настенко О. Методи дослідження фоніки поетичного тексту. Науковий вісник Східноєвропейського національного університету імені Лесі Українки. Філологічні науки. Літературознавство. 2015. Вип. 9. С. 11–116.
31. Недашківський Є.А. Класифікація типів лінгвістичних моделей Чернівців : ЧНТУ, 2016. 12 с.
32. Никоненко А.О. Огляд комп'ютерно-лінгвістичних методів обробки природномовних текстів. *Штучний інтелект*. 2011. № 3. С. 174–181.
33. Огородник В.В., Лотоцька Н.Я., Кульчицький І.М. Лінгвістичний аналіз повістей Романа Іваничука. *Молодий вчений*. 2019. № 11 (75). С. 979–981.
34. Перебийніс В.І Статистичні методи для лінгвістів : посібник. Вид. 2, випр. і допов. Вінниця : Нова Книга, 2013. 176 с.
35. Перебийніс В.І. (ред.) Статистичні параметри стилів. Київ : Наукова думка, 1967. 260 с.
36. Подшиваленко Б.О. Застосування методів статистичного аналізу для розв'язання задачі ідентифікації текстів : матеріали 25-го Міжнар. молодіж. форуму, 20–22 квітня 2021 р. Харків : ХНУРЕ, 2021. Т. 7, 10. С. 65–66.
37. Ситар Г. Статистичний аналіз цілісних словосполучень: на матеріалі Українського національного лінгвістичного корпусу. *Лінгвістичні студії*. 2019. Вип 36. С. 158–164.
38. Ситар Г. Статистичний аналіз прислів'їв і приказок: показник асоціації mutual information. *Лінгвістичні студії*. 2018. Вип. 35. С. 170–177. URL: http://nbuv.gov.ua/UJRN/lingst_2018_35_29
39. Словник української мови: В 11 томах. Київ: Наук. думка, 1970-1980. Т. 6. URL: <http://sum.in.ua/>
40. Теслюк В.М. Метод статистичного аналізу текстів драматургії Б.Шоу і розмовного стилю. *Вісник Національного університету «Львівська*

політехніка». Комп'ютерні науки та інформаційні технології. 2015. № 826. С. 161–168. URL: http://nbuv.gov.ua/UJRN/VNULPKNIT_2015_826_25

41. Хомицька І.Ю., Теслюк В.М. Статистичний аналіз англійських поетичних текстів. *Науковий вісник НЛТУ України*. 2015. Вип. 25. С. 350–356. URL: http://nbuv.gov.ua/UJRN/nvnltu_2015_25.2_60

42. Частотні словники. Дарчук Н. П. (керівник проєкту), Сорокін В.М. (програміст), Сірук О.Б., Ходаківська Я.В., Чейлитко Н.Г., Лангенбах М.О. Mova.info. Лінгвістичний портал. URL: <http://www.mova.info/page.aspx?l1=61>

43. Щербаков Д.С. Алгоритмічне та програмне забезпечення при визначенні авторства тексту. Магістерська дисертація: 121 Інженерія програмного забезпечення. Київ, 2019. 89 с.

ДОДАТКИ

Додаток 1

Корпус

Корпус поетичних та пісенних текстів воєнного часу. URL: <https://docs.google.com/spreadsheets/d/1v8l5J8DJzuZd6xEYIURWJwzVatyH2Sn2N2AsaRSFRgk/edit#gid=0>

Додаток 2

Вірші Романа Коляди

Роман Коляда. URL: https://docs.google.com/spreadsheets/d/1vSMMmNw5ZJen3B-uM07gbC3COfpsXfHW8_P6sXl-Jfc/edit#gid=0

Додаток 3

Таблиця з тематичним, стилістичним та статистичним аналізом корпусу

Тематичний та статистичний аналіз. URL: <https://docs.google.com/spreadsheets/d/1v8l5J8DJzuZd6xEYIURWJwzVatyH2Sn2N2AsaRSFRgk/edit#gid=0>

Додаток 4

Таблиця частотності слів

Частотність. URL: <https://docs.google.com/spreadsheets/d/1ONdWHA55N-R7AtwT7eEIKEmtIm0mCQMzsUH9cVhWcYo/edit?usp=sharing>

Додаток 5

Сайт

Макет сайту. URL: <https://docs.google.com/document/d/1ANQgIvnrjWlhSLOL2EbgVTxx5yOrtkhDY9qy0rhFkEA/edit>