

УДК 811.161.2'1'324'38

DOI: [https://doi.org/10.17721/um/54\(2024\).222-254](https://doi.org/10.17721/um/54(2024).222-254)

Оксана ЗУБАНЬ, канд. філол. наук, доц.

ORCID ID: 0000-0002-2644-3892

e-mail: oxana.zuban@knu.ua

Київський національний університет імені Тараса Шевченка,

Київ, Україна

Владислав КРИВЕНОК, студент

ORCID ID: 0009-0006-1070-7378

e-mail: vladyslavkryvenok@knu.ua

Київський національний університет імені Тараса Шевченка,

Київ, Україна

ВІД СЛОВА ДО ЦИФРИ: ЛІНГВОСТАТИСТИЧНИЙ “ПОРТРЕТ” ІДІОСТИЛЮ СТЕПАНА БАНДЕРИ

Вступ. Особистість С. Бандери неодноразово ставала об'єктом дослідження вчених на предмет позамовних категорій, тоді як одна з найвиразніших сторін діяча – мовна – досі не була системно й глибинно вивчена. Не стала публіцистична спадщина С. Бандери й об'єктом системного лінгвостатистичного дослідження, що і зумовлює актуальність дослідження. Метою статті є опис статистичної параметризації лексичної та частиномовної структури текстів С. Бандери на матеріалі текстової вибірки 10-ти публіцистичних статей (першодруків).

Методи: метод ретроспективного редагування тексту повторних друкованих видань, статистичні методи (обчислення статистичних параметрів, зіставлення емпіричних даних із довірчим інтервалом української публіцистики, метод статистичного моделювання), методика графічного унаочнення статистичних даних.

Результати: 1) проведено ретроспективне редагування статей редакції 1978 р. і систематизовано текстовий матеріал дослідження у цифровому форматі, придатному для автоматичного оброблення; 2) укладено текстовий корпус усіх оцифрованих першодруків статей; 3) порівняно два варіанти тексту статей – першого видання та 1978 р.; 4) за реконструйованими текстами укладено частотні словники у Корпусі української мови; 5) сформовано стилеметричну

модель статистичного дослідження і обчислено багато параметрів статистичної структури (лексичної та частиномовної) тексту С. Бандери; б) порівняно емпіричні значення лінгвостатистичних параметрів, обчислених вручну, із довірчими інтервалами медійного стилю та автоматично обчисленими емпіричними даними вебзастосунку TextAttributor 1.0.

Висновки. У ході дослідження було укладено першу в українській комп'ютерній лексикографії лексикографічну систему частотних словників на матеріалі текстів першодруків статей С. Бандери. Проведене лінгвостатистичне моделювання текстів дозволяє говорити про систематизацію статистичних даних і формування лінгвостатистичного “портрета” ідіостилу С. Бандери – це сукупність усіх емпіричних даних про статистичну структуру його текстів за дев'ятьма статистичними параметрами, зіставленими з довірчими інтервалами обчислених статистичних параметрів публіцистичного стилю української мови загалом.

Ключові слова: українська мова, ідіостиль, стилеметрична модель, статистичний параметр, корпус текстів, частотний словник.

Вступ

Особистість С. Бандери неодноразово ставала об'єктом дослідження вчених на предмет позамовних категорій, тоді як одна з найвиразніших сторін діяча – мовна – досі не була системно й глибинно вивчена. Поодинокі наукові статті, присвячені вивченню мовної особистості С. Бандери трапляються у всемережі, однак їхню кількість можна назвати справді мізерною. О. Ципердюк у статті “Публіцистичне мовлення Степана Бандери з погляду сучасних мовних норм” розглядає мовистиль С. Бандери, окремі його ознаки, що не притаманні сучасній українській літературній мові, і робить важливий висновок, що “<...> тексти публіцистики, створені автором у діаспорі, зберігають самотутні мовні риси” [Ципердюк, 2013, с. 118]. Автор статті виокремлює лексичні, орфографічні, словотвірні, морфологічні та синтаксичні риси публіцистичного авторського стилю – що їх майже неможливо віднайти в сучасному типовому літературному зразкові. Таке

дослідження тексту С. Бандери, хоч і є спорадичним, проте дає загальне уявлення про індивідуальні ознаки стилю автора.

Глибинна довершеність думок С. Бандери, що він їх найвиразніше і найточніше “одягнув” в українські слова, стали спонукую до системного аналізу його стилю, який відображає індивідуальне авторське світобачення не тільки за якісними особливостями тексту, а й за кількісними характеристиками текстової структури.

Попри вагомій здобутки української лінгвостатистики, зокрема стилеметрії¹, публіцистична спадщина С. Бандери не стала об’єктом системного лінгвостатистичного дослідження, що й зумовлює *актуальність* наукової роботи, адже, на наше переконання, національний характер публіцистики С. Бандери може стати потужним складником державної розбудови, а стилістика його текстів, що не втратили питомих рис української мови під пресом більшовицької окупаційної політики, є шляхом до мови, “не вбраної в чужий одяг”, до переосмислення сучасних правописних норм і найпевнішим визначником напряду майбутньої їхньої зміни й кодифікації.

Метою статті є опис статистичної параметризації лексичної та частиномовної структури текстів С. Бандери на матеріалі текстової вибірки 10-ти публіцистичних статей (першодруків), що виходили друком за його життя впродовж 1946–1950 рр. Для досягнення мети було здійснено такі *завдання*:

1) зібрано текстовий корпус усіх оцифрованих першодруків статей, що їх вважаємо найточнішим джерелом авторового стилю в зредагованому вигляді з-поміж усіх прижиттєвих видань;

2) шляхом зіставлення двох варіантів статей – першодруку (за відцифрованими архівами) й мюнхенської редакції 1978 р. – проведено ретроспективне редагування статей редакції 1978 р. і систематизовано текстовий матеріал дослідження в цифровому форматі, придатному для автоматичного оброблення;

¹ Див.: [Перебийніс, 1967; Перебийніс, Муравицька, & Дарчук, 1985; Бук, 2021; Дарчук, 2021; Зубань, 2020].

3) порівняно два варіанти тексту статей – першого видання та 1978 р.;

4) з метою автоматичного укладання частотних словників за текстами статей уведено реконструйовані статті до Корпусу української мови [Корпус української мови];

5) формовано стилеметричну модель статистичного дослідження і обчислено низку квантитативних параметрів статистичної структури (лексичної та частиномовної) тексту С. Бандери, порівняно отримані емпіричні значення лінгвостатистичних параметрів, обчислених вручну, із довірчими інтервалами медійного стилю та автоматично обчисленими емпіричними даними вебзастосунку TextAttributor 1.0 [TextAttributor 1.0, 2024].

Під час дослідження використано такі *методи*: метод ретроспективного редагування тексту повторних друкованих видань, статистичні методи (обчислення статистичних параметрів, зіставлення емпіричних даних із довірчим інтервалом української публіцистики, метод статистичного моделювання), методіку графічного унаочнення статистичних даних.

Результати

Формування матеріалу дослідження: ретроспективне редагування текстів. Збірка публіцистичних творів С. Бандери “Перспективи української революції”, останню редакцію й упорядкування до друку якої здійснив С. Ленкавський, вийшла друком у Мюнхені 1978 р. В Україні цю працю вперше перевидано 1998 р. репринтним способом (С. Бандера, 1998). Проте збереглися в архівах і прижиттєві публікації автора. Початковий машинопис автобіографії С. Бандери “Мої життєписні дані” (С. Бандера, 1959) наштовхнув нас на думку про те, що потрібно аналізувати саме автентичний текст автора – першодрук, бо це найточніше джерело його ідіостилу. Уже після загибелі С. Бандери 15 жовтня 1959 р. у жовтні 1961 р. машинопис відредаговано (рис. 1) і частину його вперше надруковано за кордоном у газеті “Томін України”.

Мої життєписні дані

Народився 1 січня 1909 року в селі Угринів Старий, повіт Калущ в Галичині, яка в той час належала до Австро-Угорської імперії разом з двома іншими західньохристиянськими країнами: Буковиною і Закарпаттям.

Мій батько, Андрій Бандера, греко-католицький священник був з той час парохом у селі Угринівці (до парохиї належало ще сусіднє село Бережниця Шляхецька). Батько походив з роду змислов Стриб, а мати, Яна була дочкою Білоуса. Її батько, Мирослав Бандера, походив зі старої священичої родини, була донькою греко-католицького пароха в тому ж селі Угринівці Старшор - Володимир Глодзінський і Катерина з дому Кушлак. Я був другою дитиною моїх батьків. Старшою за мене була сестра Марта. Молодші: брат Олександр, сестра Володимира, брат Василь, сестра Оксана, брат Богдан і наймолодша сестра Мирослава, що померла немовлям.

Рис. 1. Фрагмент початкового машинопису автобіографії С. Бандери
“Мої життєписні дані” з редакторськими правками
(С. Бандера. 1961. № 42(648) за 14 жовтня 1961 р.)

У такому варіанті з деякими правками цей текст пізніше передруковували всі інші видання. За нашими підрахунками, стаття “Мої життєписні дані”, надрукована в “Перспективах української революції” 1978 р., містить загалом близько 400 різноманітних текстових правок. Зокрема для тижневика “Гомін України” редактор зробив близько 230 правок, тобто іншу частину виправлень зроблено в наступних виданнях, зокрема і мюнхенському. Не зваживши на цей факт, можна було б натрапити на серйозну перешкоду на шляху отримання вірогідних даних про статистичну структуру ідіостилю С. Бандери (така реконструкція матиме особливу значущість для вивчення ідіостилистичних ознак автора на морфемному рівні організації тексту в майбутніх дослідженнях).

Проблема зіставлення статистичних параметрів ідіостилю текстів різних редакцій одного твору була розглянута С. Бук [Бук, 2012] у порівнянні текстів авторських редакцій 1884 і 1907 рр. повісті Івана Франка “*Voа constrictor*”. Це дослідження переконливо доводить, що твори двох редакцій характеризуються різними емпіричними значеннями статистичних параметрів ідіостилю І. Франка. І хоча наше дослідження базується на текстах С. Бандери, редагованих не автором, а різними видавництвами відповідно до правопису часу публікації, було вирішено здійснити ретроспективне редагування текстів мюнхенського видання: відтворити первинні тексти без жодних перелицювань і за цими текстами провести лінгвостатистичний експеримент.

Завдяки машинописові статті, що її С. Бандера написав в останній рік життя, ми отримали єдиний зразок автентичного письма автора, що не зазнало жодних редагувань. Інших таких машинописів у відкритому доступі нема, тому статті мюнхенського видання було реконструйовано до вигляду перших публікацій. Наприклад, стаття “Значення широких мас та їх охоплення” (С. Бандера, 1946с) вперше була опублікована в часописі-місячнику “Визвольна політика” (Мюнхен), тому вважатимемо, що ця перша публікація – найточніший варіант ідіостилю С. Бандери. Нині 28 статей набули свого первинного вигляду. Зіставлення двох варіантів статей (рис. 2) відбувалося повністю в ручному режимі і становило надзвичайно трудомістке завдання: у середньому редагування 5-х сторінок (формату А 4) тексту вимагало приблизно 2-х годин напруженої роботи.

Дві версії тієї самої статті С. Бандери – це тексти, що відрізняються правописно, стилістично, лексично і граматично. З огляду на це, аналіз якості змін може засвідчувати оригінальність і неповторність певного слова чи словосполучки в тексті. Умовно можемо виокремити такі причини змін: 1) ненормативність мовної одиниці з погляду тогочасної мовної норми; 2) стилістична маркованість мовної одиниці, зокрема діалектизми; 3) авторський новотвір (оказіоналізм); 4) уподобання редактора. Хоча часто певну зміну складно обґрунтувати й зрозуміти її доцільність у тексті.

В тридцятих роках, від 1-го Конгресу ОУН до 1933 р., в революційно-визвольному підпіллі на Західних Українських Землях ішов внутрішній змаг за зміст, форми і організацію дальшої визвольної боротьби. Після майже десятилітнього періоду бойової діяльності УВО (Української Військової Організації) вступила в дію Організація Українських Націоналістів. Почався змаг між двома тенденціями та суперечки, не раз в гострих формах, між їхніми речниками й сторонниками. Серед членства аж загально утвердилась бода свідомість, що далі зводити революційну дію лише до бойових форм недоцільно, як недоцільно залишати політичну діянку боротьби існуючим легальним партіям опортуністичного напрямку, — натомість необхідно розгорнути широку революційно-політичну акцію і творити підпілля ідеологічно-політичну організацію з націоналістичною ідеологією та революційно-визвольною політичною концепцією. В тому відношенні, щодо потреби творення на ЗУЗ Організації Українських Націоналістів, не було ще ані сумнівів, ані суперечок.

У тридцять роки, від 1-го Конгресу ОУН до 1933 року, в революційно-визвольному підпіллі на західноукраїнських землях ішла внутрішня боротьба за зміст, форму й організацію дальшої визвольної боротьби. Після майже десятирічного періоду бойової діяльності УВО (Української Військової Організації) вступила в дію Організація Українських Націоналістів. Почалися змагання між двома тенденціями і суперечки, не раз гострі, між їх речниками і прихильниками. Серед членства бода вже утвердилась загальна свідомість, що обмежувати далі революційну дію лише бойовими формами недоцільно, як недоцільно залишати політичну діянку боротьби існуючим легальним партіям опортуністичного напрямку, — натомість необхідно розгорнути широку революційно-політичну акцію і творити підпілля ідеологічно-політичну організацію з націоналістичною ідеологією і революційно-визвольною політичною концепцією. У тому відношенні щодо потреби творити на західноукраїнських землях Організацію Українських Націоналістів не було ні сумнівів, ані суперечок.

Рис. 2. Порівняння двох редакційних варіантів статті “Командир – провідник (сл. пам. Романа Шухевича)” в газеті “Шлях перемоги” (С. Бандера, 1954) (ліворуч) і оцифрованій версії “Перспектив української революції” (С. Бандера, 1998) (праворуч)

Для ефективного й зручного пошуку лексики, що засвідчує певне мовне явище в текстах С. Бандери (напр., іншомовних слів із м'яким [л'], із буквосполуками *ія, ію, іє, ав, ев, ов*) було написано програмовий код в інтернет-середовищі програмування Colaboratory. Комп'ютерна програма може виконувати такі функції:

1) за вибраним модулем пошуку слів – пошук слів за зазначеною буквосполукою (чи буквосполуками) або пошук певних словоформ – сформувати список слів, із яких користувач програми вручну вибере ті, що задовольняють його пошуковий запит;

2) порахувати абсолютну частоту обраних слововживань у тексті, яким попередньо присвоюється число-номер тексту, у якому вжито слово.

Для розуміння суттєвості редакційних змін наведемо кілька прикладів зіставного аналізу трьох статей: “Значення широких мас та їх охоплення” (С. Бандера, 1946с), “До проблеми політичної консолідації” (С. Бандера, 1946b), “Мої життєписні дані” (С. Бандера, 1959). Так, на стилістичному рівні аналізу було виявлено:

- заступлення фразеологічної словосполуки *в великій мірі*, що її часто словники з культури української мови подають як кальковану з московської мови, сполукою *великою мірою*;

- заміну прикметника *важний*, що в Словнику української мови [Словник української мови, 1970–1980] має ремарку *розм.*, на більш нейтральний варіант *важливий* (аналогічно *найважливіший* → *найважливіший*). Окрім того, стилістичну зміну можна вважати результатом зміни на рівні морфем – тобто словотвірної зміни (*-н-* → *-лив-*);

- застарілу лексику *умовини* (що існує лише в множині) замінено в усіх випадках на *умови*;

- С. Бандера уживає прислівник вищого ступеня порівняння *дальше* у вторинному значенні (те саме, що *далі*), що можна почути в розмовному стилі, тому редактор заміняє його на *далі*;

- транспонований займенник *все* замінено на *завжди* (*все покликаються кола УНР* → *завжди покликаються кола УНР*); прислівник часу *завсіді*, що маркований розмовним функційним стилем [Словник української мови, 1970–1980], заступає інший прислівник – *завжди*;

- цікавою є дієслівно-іменникова словосполучка *зуживати здобутків*, де залежний компонент стоїть у нетиповому для нього родовому відмінку, на відміну від стилістично блідої фрази *користуватися здобутками*.

Особливої уваги потребують вказівні займенникові прикметники *цей / той*, що зазвичай позначають ступінь віддаленості предмета. Редактори статей С. Бандери надто часто виправляють займенникове слово *той* на *цей* у тих випадках, коли про ступінь віддаленості не йдеться зовсім (напр., *та цілість*, *тим вимогам*, *ті рямці*, *тій формації* тощо). Коли йдеться про абстраговані ситуації, де предмет розглядають безвідносно до його положення (якого він може взагалі не мати) у просторі, уживають здебільшого займенник *той* [Фаріон, 2013, с. 240].

Неабияк цікаво натрапляти у статтях С. Бандери на діалектизми, як фонетичні, так і лексичні (серед яких є чужомовні одиниці), що їх замінювано на нейтральні словесні форми: *вистарчати* → *вистачати*, *мент* → *момент*, *рямці* → *рамці*, *атут* → *козир*, *вальор* → *значення*, *вичерпливий* → *вичерпний*, *вакації* → *ферії*, *живець* → *місцевість*, *жиючи* → *живучи*, *знова* → *знов*, *кошівка* → *кошиківка*, *парох* → *священик*, *парохія* → *парафія*.

Аналіз редакційних змін у трьох статтях показав, що першодруки містять великий пласт лексики, правопис якої або відрізняється від сучасної мовної норми, або навіть вступає в конфлікт із мовними приписами, що були чинними до 1933 р.² Окрім того, зіставний аналіз переконав нас у важливості реконструкції первинної форми статей, адже одиниці, що їх уживав автор, відкривають шлях для заглиблення в його мовомислення, у справжню, виформовану впродовж життя, свідомість.

Лінгвостатистичний експеримент. Десять статей первинної редакції було внесено до Корпусу української мови [Корпус української мови, 2024], що забезпечило автоматичне опрацювання тексту та укладання частотних словників (ЧС) словоформ і лексем за текстовими вибірками окремих статей, враховуючи рангові списки за частинами мови.

Квантитативний опис структури тексту С. Бандери зроблено на базі зональної вибірки, структурованої на 10 підвибірок за статтями автора, написаними в часовому проміжку січень 1946 – жовтень 1950 рр.³ Обсяг текстової вибірки становить 75071 слововжитків, із них 4607 слововжитків (або 6,1 %) є не розпізнаними системою

² Див.: Кримський, А. Ю. *Найголовніші правила українського правопису*. Видання Волинського Кооперативного Видавництва, 1921; Шумлянський, Ф. (1927). *Найголовніші правила українського правопису Української академії наук видання 1921 р. (Розвідка)*. Відбиток. *Записки Херсонського інституту народньої освіти ім. Н. К. Крупської*, ч. 2; *Український правопис (проект) (1926)*. Держ. вид-во України; *Український правопис (1928)*. Держ. вид-во України; *Український правопис (1933)*. Рад. школа.

³ Див. за Списком іпостративного матеріалу: (С. Бандера, 1946а); (С. Бандера, 1946б); (С. Бандера, 1946с); (С. Бандера, 1948а); (С. Бандера, 1948б); (С. Бандера, 1950б); (С. Бандера, 1950в); (С. Бандера, 1950с); (С. Бандера, 1950д); (С. Бандера, 1950е).

автоматичного морфологічного аналізу. Зважаючи на обсяг, хронологічну визначеність, жанрову обмеженість, тематичну та стильову однорідність, вважаємо сформовану текстову вибірку статистично та лінгвістично репрезентативною.

Оскільки система автоматичного морфологічного аналізу в Корпусі української мови працює на резидентному словнику, укладеному за нормативними словниками української літературної мови, то значний пласт лексики досліджуваних текстів залишається нерозпізнаним і становить надважливе джерело для майбутнього дослідження – унікальну лексичну ознаку авторського стилю. Тому нерозпізнані слововживання було пролематизовано вручну і додано до частотних словників. Після аналізу 2747 елементів статистичного ряду нелематизованої лексики було визначено 1385 одиниць, що сукупно можна вважати вирізняльним компонентом ідіостилю С. Бандери. Розмах їхніх частот: від 1 до 248. Найбільша кількість одиниць із частотою 1–912 (або 66 % від усього списку). Найчастотнішими трьома лексемами в цьому частотному спискові є *большевицький* (248), *большевизм* (224), *імперіялізм* (114). Показово, що С. Бандера жодного разу у своїх статтях не вживає слова *більшовик* або його дериватів.

Отриманий список лексем, що їх не знають словники сучасної української літературної мови, може стати потужним джерелом для словників, а найперше – для укладачів наступних правописів української мови.

З Корпусу української мови було вивантажено 10 частотних словників лексем (окремо для кожної статті) для подальшої ефективної роботи з ними в табличному форматі Excel. Кожній одиниці лексичного реєстру за емпіричним значенням абсолютної частоти було приписано ранг – важливу характеристику, що вказує на позицію лексеми за частотою щодо інших лексем (табл. 1).

Лінгвостатистичний експеримент передбачав розроблення стилеметричної моделі аналізу. У нашому дослідженні було використано стилеметричну модель, розроблену колективом авторів автоматичної системи лінгвостатистичного аналізу TextAttributor 1.0 [TextAttributor 1.0. Методика, 2024]. Стилеметрична модель спрямована на (1) обчислення багатьох статистичних параметрів (СП) і (2) зіставлення емпіричних даних цих СП із типовими

еталонними числовими значеннями СП у медійному (публіцистичному) стилі української мови. Використана в дослідженні стилеметрична модель об'єднує такі статистичні параметри: 1) обсяг тексту, 2) обсяг словника лексем, 3) кількість нарах legomena, 4) індекс багатства тексту, 5) індекс винятковості словника й тексту, 6) індекс іменних означень (індекс епітізації), 7) індекс дієслівних означень, 8) індекс номінальності, 9) ступінь прономіналізації, 10) індекс модальності, 11) індекс субстантивності. Крім перерахованих СП, стилеметрична модель нашого дослідження включає ще один морфолого-статистичний параметр – 12) відносна частота (%) частини мови (окрім вигука).

Таблиця 1

**Найчастотніші лексеми у статті
“Слово до українських націоналістів-революціонерів за кордоном”
(С. Бандера, 1948b)**

Ранг	Лексема	Клас	Частота
1	і	С	612
2	в	П	383
3	на	П	372
4	не	Ь	297
5	до	П	259
6	політичний	А	255
7	український	А	232
8	бути	Г	215
9	з	П	185
10	що	С	176
11	й	С	169
12	наш	О	163
13	за	П	157
14	вона	М	147
15	такий	О	143
16	боротьба	К	142
17	Україна	к	142
18	а	С	132
19	у	П	130
20	для	П	129

Для реалізації другого завдання – зіставлення емпіричних даних СП із типовими еталонними числовими значеннями СП у медійному (публіцистичному) стилі української мови – було використано числові значення довірчого інтервалу вебзастосунку TextAttributor 1.0 [TextAttributor 1.0, 2024] для довгих і коротких текстів медійного стилю (зокрема публіцистичного підстилю). “Довірчий інтервал кожного статистичного параметра визначався за середнім квадратичним відхиленням середнього числового значення індексу (СЧЗІ) у текстовій вибірці: $СЧЗІ \pm \sigma$. Довірчий інтервал за σ був обраний за результатами лінгвостатистичного експерименту, тому що статистичні індекси можуть мати малі числові значення (від 0 до 1), особливо в коротких текстах, а використання 95 % довірчого інтервалу при обчисленні 2σ визначало від’ємні нижні порогові значення інтервалу. За результатами зіставлення емпіричних числових значень із пороговими значеннями довірчого інтервалу система визначає 3 варіанти висновків <...>: (1) числове значення індексу входить до довірчого інтервалу медійного стилю української мови; (2) числове значення індексу не входить до довірчого інтервалу медійного стилю української мови і нижче, ніж нижнє порогове значення цього інтервалу; (3) числове значення індексу не входить до довірчого інтервалу медійного стилю української мови і вище, ніж верхнє порогове значення цього інтервалу. Якщо за результатами автоматичного зіставлення емпіричних даних із межами довірчого інтервалу реалізуються відповіді (2) і (3), то в такому випадку констатується лінгвостатистична ознака ідіостилу автора” [TextAttributor 1.0. Методика, 2024].

Комплексний квантитативний опис за параметрами стилеметричної моделі можемо вважати лінгвостатистичним “портретом” ідіостилу автора. Систематизацію статистичних параметрів стилеметричної моделі для кожної статті демонструє табл. 2.

Таблиця 2

Стилеметрична модель статті
“Слово до українських націоналістів-революціонерів за кордоном”

“СЛОВО ДО УКРАЇНСЬКИХ НАЦІОНАЛІСТІВ-РЕВОЛЮЦІОНЕРІВ ЗА КОРДОНОМ”				
Слововживань	18598	Лематизованих	17785	95,63
		Нелематизованих	813	4,37
Лексем	3270	Лематизованих	2808	85,87
		Нелематизованих	462	14,13
Нарах Legomena	1805	Лематизованих	1454	80,55
		Нелематизованих	351	19,45
СП кількісного співвідношення слів реєстру словника й обсягу текстових слововживань				
індекс багатства		0,16		
індекс винятковості словника		0,55		
індекс винятковості тексту		0,10		
СП кількісного співвідношення лексико-граматичних класів слів у тексті				
індекс іменних означень		1,80		
індекс дієслівних означень		0,40		
ступінь номінальності		2,33		
індекс прономіналізації		0,06		
індекс модальності		0,03		
індекс субстантивності		0,29		

Закінчення табл. 2

Частини мови			
	абсолютна частота		відносна частота (%)
Іменник	5076		28,89
Прикметник	2818		16,04
Займенник	іменниковий	1107	6,30
	прикметниковий	1005	5,72
Прислівник	власне прислівник	866	4,93
	форми на -но/-то	20	0,11
	предикатив	105	0,60
Дієслово	2177		12,39
Числівник	97		0,55
Прийменник	2019		11,49
Сполучник	1703		9,69
Частка	580		3,30
Сума	17573		

Стилеметричний аналіз ідіостилю С. Бандери. Лінгвостатистичний аналіз ідіостилю С. Бандери здійснено на базі стилеметричної моделі, що систематизує емпіричні числові значення статистичних параметрів, обчислених за ЧС кожної аналізованої статті (табл. 2). Оскільки автоматичний морфологічний аналізатор корпусу текстів розпізнає не всі текстові слововживання, то для трьох базових атрибутів тексту – кількість слововживань, обсяг словника, кількість *hapaх legomena* – у таблиці наведено два емпіричні числові значення, друге з яких – кількість оброблених системою. Для обчислення статистичних параметрів тексту взято саме дані про автоматично опрацьовану (лематизовану й морфологізовану) лексику.

Лише для тих статистичних параметрів, що обчислюються за кількістю *hapaх legomena* (одиниць, що трапились у тексті один раз), було взято повні кількісні дані з урахуванням обробленої та необробленої лексики. За емпіричними даними кожного статистичного параметра в підвибірках статей побудовано діаграми, що хронологічно демонструють зміну емпіричного значення і дають можливість робити висновки про закономірності розвитку статистичної структури тексту С. Бандери.

Текстові підвибірки за статтями систематизовано в хронологічному порядку із присвоєнням кожній статті порядкового номера:

- “Значення широких мас та їх охоплення” (С. Бандера, 1946с) – 1;
- “До проблеми політичної консолідації” (С. Бандера, 1946b) – 2;
- “До засад нашої визвольної політики” (С. Бандера, 1946a) – 3;
- “Плянність революційної боротьби в краю” (С. Бандера, 1948a) – 4;
- “Слово до українських націоналістів-революціонерів за кордоном” (С. Бандера, 1948b) – 5;
- “Українська національна революція, а не тільки проти-режимний резистанс” (С. Бандера, 1950е) – 6;
- “В десятю річницю створення Революційного Проводу ОУН” (10.2.1940) (С. Бандера, 1950a) – 7;
- “Війна в Кореї і національно-визвольна політика” (С. Бандера, 1950b) – 8;

• “Третя світова війна і визвольна боротьба” (С. Бандера, 1950d) – 9;

• “Друзі – українські націоналісти!” (С. Бандера, 1950с) – 10.

Окрім того, зіставлення отриманих емпіричних даних із довірчими інтервалами СП медійного стилю української мови за системою TextAttributor 1.0 виокремлює індивідуальні стилістичні ознаки тексту з-поміж типових.

СП: обсяг тексту. Під обсягом тексту маємо на увазі кількість слововживань автоматично обробленої лексики. Графічний розподіл цього СП за роками першого друкування статті (рис. 3) демонструє, що впродовж 1946–1950-х рр. обсяг Бандериної статті то зростає, то спадає, тобто не бачимо нахилу до збільшення обсягу тексту чи зменшення. Хоч зигзагоподібна форма графіка і приблизна його симетричність можуть свідчити про певні психологічні коливання автора, проте для підтвердження цієї гіпотези необхідно дослідити всю генеральну сукупність текстів С. Бандери.



Рис. 3. Графік зміни кількості слововживань у статтях С. Бандери впродовж 1946–1950 рр.

СП: обсяг словника. Під обсягом словника маємо на увазі обсяг лексем, лематизованих автоматично. Ламана на графіку (рис. 4) майже повторює ламану попереднього графіка (рис. 3): тобто обсяг словника залежить від обсягу тексту, що характерно

для коротких текстів, але не для довгих, адже відповідно до закону Дж. Ципфа: що більше слововживань має текст, тим потенційно рідше в ньому можна натрапити на нові лексеми.



Рис. 4. Графік зміни обсягу словника в статтях С. Бандери впродовж 1946–1950 рр.

СП: індекс багатства. Індекс багатства обчислюють за формулою $ib = \frac{V}{N}$, де V – це обсяг словника, а N – це обсяг тексту.

Наприклад, стаття “Значення широких мас та їх охоплення” (С. Бандера, 1946с) має такий індекс багатства:

$$ib = \frac{473}{1132} = 0,42.$$

Ламана зміни індексу багатства на графіку рис. 5 дає підстави вважати, що зі збільшенням обсягу тексту обсяг словника зменшується, тобто потенційно нових слів з’являється все менше, що підтверджує дію закону Дж. Ципфа. Однак дія закону порушується в порівнянні двох статей: стаття “Друзі – українські націоналісти” має більший обсяг тексту (1652), ніж стаття “Значення широких мас та їх охоплення” (1123), проте індекси багатства в них однакові (0,42). Це вказує на більшу різноманітність лексики у другій статті. Така закономірність характеризує і статті “Третя світова війна <...>” (10450) і “Українська національна революція <...>” (13429), що мають

індекс багатства 0,2, отже, перша стаття – відносно багатша за лексичним складом ніж друга. Порівнявши дві статті “До проблеми політичної консолідації” (8763) і “Третя світова війна <...>” (10450), бачимо, що більша з них має більший індекс багатства (0,2 проти 0,18), що засвічує більшу різноманітність лексики в статті “Третя світова війна <...>” (тобто в ній менше повторюваних слів, ніж у першій).



Рис. 5. Графік зміни індексу багатства
у статтях С. Бандери впродовж 1946–1950 рр.

Типові ознаки медійного стилю матимуть тексти, індекс багатства яких не виходить за межі довірчого інтервалу [0,26–0,5]. Отже, статті під номерами 2, 4, 5, 6, 7, 9 мають ознаки ідіостилу за індексом багатства: ступінь різноманітності тексту в них нижчий за нижнє порогове значення довірчого інтервалу. Середнє значення багатства тексту для 10-х статей (0,21) вказує на загальну тенденцію частіше повторювати слова, тобто використовувати обмежений словник лексем досить часто.

СП: індекс винятковості словника. Індекс винятковості словника обчислюється за формулою $ivl = \frac{N_1}{V}$, де N_1 – кількість слів-гапаксів (таких, що мають абсолютну частоту в тексті 1), V – обсяг словника. Наприклад, перша стаття “Значення

широких мас та їх охоплення” має індекс винятковості словника $ivl = \frac{343}{514} = 0,67$.

Середнє значення індексу винятковості словника складає 0,59, що є типовою ознакою для медійного стилю (довірчий інтервал – $[0,5-0,8]$). Усі статті перебувають у цьому інтервалі (рис. 6).



Рис. 6. Графік зміни індексу винятковості словника в статтях С. Бандери впродовж 1946–1950 рр.

Для першої й останньої статті (що розміщені хронологічно) значення цього індексу однакове, так само, як і індекс багатства, що наводить нас на думку про те, що ці два індекси певним чином пов'язані між собою. Понад то, форма їхніх графіків є подібною.

Отже, у руслі дії закону Дж. Ципфа можна констатувати: що вищий індекс багатства тексту, тим вищий індекс винятковості словника. Перевіримо цю закономірність через графічне зіставлення двох ламаних (рис. 7): нижня ламана – індекс багатства, верхня ламана – індекс винятковості словника.

Базовим принципом систематизації емпіричних даних на графіку є впорядкування за ростом числових значень індексу багатства, тому номери текстів за віссю Х-ів також впорядковані за цим принципом.



Рис. 7. Графік зіставлення двох СП – багатства тексту й індексу винятковості – у статтях С. Бандери

Бачимо, що загалом зі зростанням багатства тексту зростає індекс винятковості словника, однак числове збільшення емпіричних даних індексу винятковості нижче ніж підвищення емпіричних даних індексу багатства тексту. Для деяких статей ця закономірність дещо порушена, що може бути спричинено часткою нелематизованих слів у досліджуваних вибірках.

СП: індекс винятковості тексту. Для обчислення індексу винятковості тексту необхідно кількість слів-гапаксів поділити

на обсяг тексту: $ivt = \frac{N_1}{N}$. Середнє значення індексу винятковості

тексту складає 0,14, що є ознакою ідіостилу С. Бандери. (довірчий інтервал системи TextAttributor 1.0 – [0,2–0,7]).

Як засвідчує графічне представлення на рис. 8, середнє значення є нижчим за норму медіастилу (змодельовано прямою на графіку): слова із частотою 1 становлять менше ніж 20 % тексту, що визначає низький ступінь лексичної винятковості тексту – а тому є ознакою ідіостилу. Хоча в чотирьох статтях цей показник є типовим для текстів публіцистичного стилю,

але наближений до нижньої межі довірчого інтервалу. Важливо зазначити те, що типові значення цього параметра мають статті відносно невеликого обсягу: найбільша стаття має обсяг 3050 слововжитків.



Рис. 8. Графік зміни індексу винятковості в статтях С. Бандери впродовж 1946–1950 рр.

СП: індекс іменних означень (іменник & прикметник). Прикметник як носій ‘статичної ознаки предмета’ завжди характеризує іменник як носія категорійного значення ‘предметність’. Отже, кількісне співвідношення між цими частинами мови є важливим параметром для визначення епітизації тексту. Для знаходження індексу іменних означень необхідно кількість іменників поділити на кількість прикметників тексту. Прикметників з узагальнено-вказівною семантикою (займенникових прикметників) цей параметр не враховує.

Довірчий інтервал для цього параметра – [2,1–3,6]. Чим менше в тексті прикметників, тим вище значення індексу епітизації (при значенні 1 кількість іменників і прикметників однакова). Лише три статті мають типові для публіцистики значення (рис. 9). Усі інші статті мають ознаки ідіостилю

С. Бандери: іменників майже удвічі більше за прикметників, тобто в середньому кожен другий іменник має при собі залежний прикметник (про це свідчить середнє значення індексу – 1,9). Отже, С. Бандера використовує у своїх статтях психологічний прийом – прагнення посилити емоційний вплив на читача через прикметник, оскільки прикметник завжди робить наголос на якійсь одній із багатьох ознак предмета.

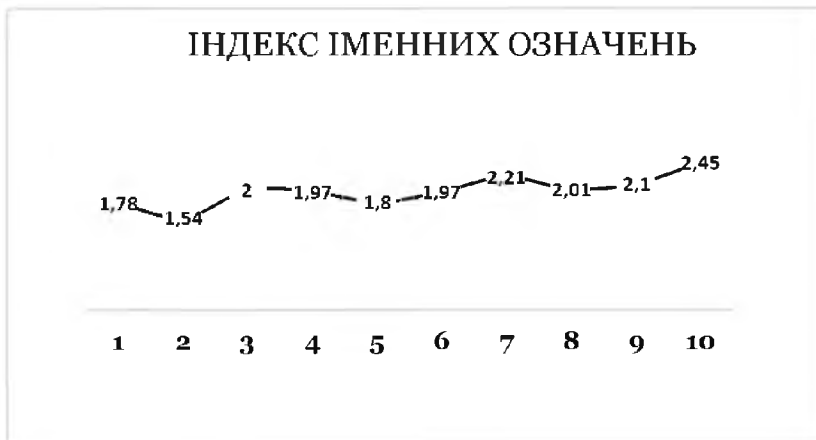


Рис. 9. Графік зміни індексу іменних означень у статтях С. Бандери впродовж 1946–1950 рр.

СП: індекс дієслівних означень (прислівник & дієслово). Цей параметр свідчить про ступінь вияву ознаки дії в тексті. Для його обчислення треба кількість прислівників поділити на кількість дієслів тексту. Довірчий інтервал значень індексу дієслівних означень такий: $[0,4–0,6]$, однакова кількість дієслів і прислівників засвідчується значенням 1. Отже, робимо висновок, що лише одна стаття (“До засад нашої визвольної політики” (С. Бандера. 1946а)) виходить за нижню межу цього проміжку, тобто має ознаку ідіостилію автора (рис. 10). Середнє значення індексу дієслівних означень для 10 статей С. Бандери: 0,46. Отже, для тексту автора не є характерною надмірна прислівниковість.

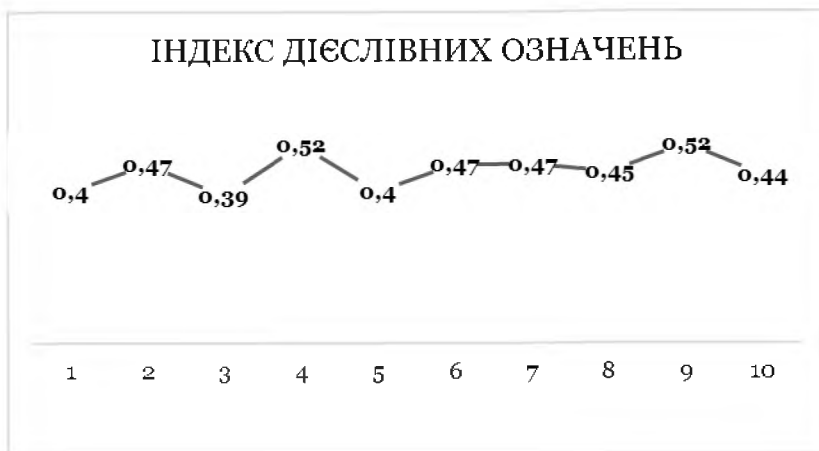


Рис. 10. Графік зміни індексу дієслівних означень у статтях С. Бандери впродовж 1946–1950 рр.

СП: ступінь номінальності (іменник & дієслово). Ступінь номінальності виражає відношення суми вживань іменників до суми вживань дієслів, тобто вказує на співвідношення між статичністю і динамічністю в тексті. Довірчий інтервал для значень ступеня номінальності – $[2-3,1]$, за однакової кількості іменників і дієслів ступінь номінальності має значення 1. Із побудованої діаграми на рис. 11 бачимо, що сім статей виявили типові ознаки публіцистичного стилю, а 3 статті (“Значення широких мас та їх охоплення” (С. Бандера, 1946с), “До засад нашої визвольної політики” (С. Бандера, 1946а), “Війна в Кореї <...>” (С. Бандера, 1950b)) мають значення, що виходять за межі довірчого інтервалу. Для першої із зазначених статей характерною є більша дієслівність, а для двох інших – більша іменниковість. Аномально високі значення ступеня номінальності (4,28) може бути свідченням науковості тексту, його логічної аргументованості.

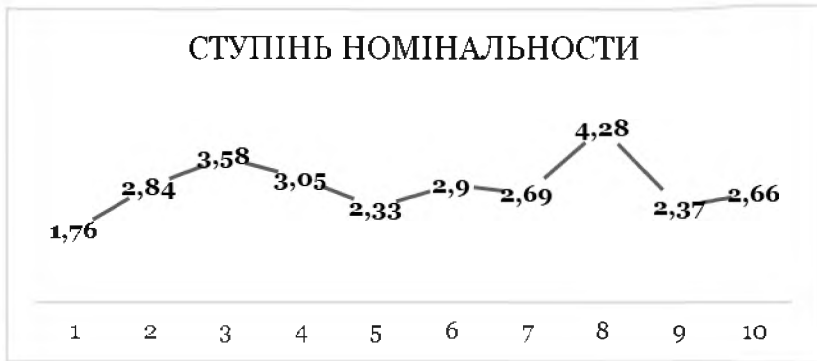


Рис. 11. Графік зміни ступеня номінальності в статтях С. Бандери впродовж 1946–1950 рр.

СП: індекс прономіналізації (займенник). Щоб знайти значення цього індексу, необхідно кількість займенників, що трапились у тексті, поділити на кількість слововжитків.

Типові значення цього індексу лежать у межах $[0,03-0,10]$. Отже, усі 10 статей (рис. 12) мають типові показники індексу прономіналізації, тобто займенниковості тексту, що визначає ступінь його кореферентності.



Рис. 12. Графік зміни індексу прономіналізації у статтях С. Бандери впродовж 1946–1950 рр.

СП: індекс модальности (частка). Індекс модальности виражає відношення кількості часток до кількості слововживань тексту і свідчить про ступінь емотивности тексту. Типові значення цього індексу лежать у межах $[0,02-0,05]$: позначено на графіку прямою (рис. 13). Отже, лише одна стаття (“Війна в Кореї <...>” (С. Бандера, 1950b)) має за цим параметром ознаку ідіостилію С. Бандери: низьке значення вказує на збідненість емотивности статті.

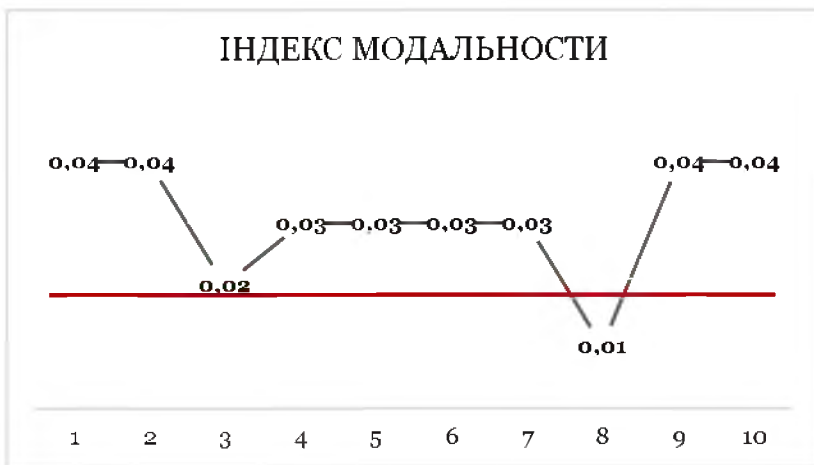


Рис. 13. Графік зміни індексу модальности у статтях С. Бандери впродовж 1946–1950 рр.

СП: індекс субстантивности (іменник). Індекс субстантивности виражає відношення кількості іменників до обсягу слововживань тексту, свідчить про насичення тексту іменниками, статичність стилю. Довірчий інтервал індексу – $[0,26-0,32]$. З графіка на рис. 14 висовуємо, що сім статей показали типові значення параметра, а три (“Значення широких мас <...>” (С. Бандера, 1946с), “До засад нашої визвольної політики” (С. Бандера, 1946а), “Війна в Кореї <...>” (С. Бандера, 1950b)) показали відмінні від цього інтервалу значення індексу субстантивности, однак відхилення не є суттєвим.



Рис. 14. Графік зміни індексу субстантивності у статтях С. Бандери впродовж 1946–1950 рр.

Валідація емпіричних даних лінгвостатистичного аналізу за перевіркою вебзастосунку TextAttributor 1.0. Порівняння емпіричних числових значень проаналізованих статистичних параметрів із результатами атрибуції текстів С. Бандери в роботі вебзастосунку TextAttributor 1.0 [TextAttributor 1.0, 2024] подано в табл. 3, що підтверджує істинність отриманих емпіричних даних, обчислених вручну та за результатами роботи автоматичної системи, попри несуттєву похибку, що, як уже зазначалося, спричинена часткою необробленої лексики в автоматичній системі.

Дискусія і висновки

Створена в дослідженні лексикографічна система частотних словників у Корпусі української мови – перша в українській комп’ютерній лексикографії параметризована система текстоорієнтованих електронних частотних словників мови С. Бандери. Особливого значення набуває той факт, що частотні словники створювались на базі реконструйованого текстового матеріалу на базі джерел-першодруків статей автора. Укладені частотні словники, автоматичне обчислення та систематизація статистичних характеристик текстів автора у формі статистичних параметрів відкривають перед філологами нові можливості та перспективи в дослідженні статистичної структури текстів автора на лексичному й частининомовному рівнях організації.

Таблиця 3

Емпіричні значення статистичних параметрів
за статтями С. Бандери, автоматично визначених
системою TextAttributor 1.0 (сірі смуги)
та обчислених вручну (білі смуги)¹

	1	2	3	4	5	6	7	8	9	10
ІБ	0,42	0,18	0,3	0,25	0,16	0,2	0,24	0,35	0,2	0,42
ib	0,42	0,19	0,3	0,26	0,16	0,2	0,24	0,35	0,2	0,42
IBC	0,67	0,56	0,63	0,6	0,55	0,59	0,62	0,63	0,58	0,67
ivl	0,65	0,53	0,6	0,58	0,53	0,56	0,57	0,61	0,54	0,65
IBT	0,29	0,11	0,2	0,17	0,1	0,13	0,17	0,24	0,13	0,29
ivt	0,27	0,1	0,18	0,15	0,08	0,11	0,14	0,22	0,11	0,27
ІО	1,78	1,54	2,0	1,97	1,8	1,97	2,21	2,01	2,1	2,45
ііо	1,82	1,44	1,99	1,87	1,72	1,86	2,13	1,93	1,98	2,18
ІДО	0,4	0,47	0,39	0,52	0,4	0,47	0,47	0,45	0,52	0,44
ідо	0,37	0,46	0,36	0,51	0,39	0,47	0,48	0,44	0,51	0,42
СтН	1,76	2,84	3,58	3,05	2,33	2,9	2,69	4,28	2,37	2,66
stn	1,8	2,67	3,55	2,9	2,22	2,72	2,6	4,12	2,25	2,45
Прон	0,08	0,05	0,06	0,05	0,06	0,06	0,06	0,05	0,07	0,08
ipro	0,09	0,06	0,06	0,05	0,07	0,06	0,06	0,05	0,07	0,09
ІМод	0,04	0,04	0,02	0,03	0,03	0,03	0,03	0,01	0,04	0,04
imod	0,05	0,04	0,02	0,03	0,04	0,03	0,03	0,02	0,05	0,04
ІСуб	0,25	0,29	0,34	0,32	0,29	0,31	0,3	0,35	0,29	0,31
isub	0,26	0,28	0,34	0,31	0,28	0,3	0,3	0,34	0,28	0,29

¹ Умовні скорочення статистичних параметрів: ІБ / ib – індекс багатства; IBC / ivl – індекс винятковості словника; IBT / ivt – індекс винятковості тексту; ІО / ііо – індекс іменних означень; ІДО / ідо – індекс дієслівних означень; СтН / stn – ступінь номінальності; Прон / ipro – індекс прономіналізації; ІМод / imod – індекс модальності; ІСуб / isub – індекс субстантивності.

Апробована в дослідженні стилеметрична модель, теоретично обґрунтована і валідована в системі TextAttributor 1.0, може розглядатися як типологічна у вивченні текстів різних стилів та ідіостилів, а в частині частиномовних статистичних параметрів демонструє вищий рівень узагальнення квантитативної моделі тексту, ніж модель статистичної структури лексики, тому що статистичне індексування за частиномовними ознаками, як порівняти з кількістю слів, зменшує порівнювані лінгвістичні явища в тисячі разів.

Лінгвостатистичне моделювання текстів С. Бандери дозволяє говорити про систематизацію статистичних даних і формування лінгвостатистичного “портрета” ідіостилу С. Бандери – це сукупність усіх емпіричних даних про статистичну структуру його текстів за дев'ятьма статистичними параметрами, зіставленими з довірчими інтервалами обчислених статистичних параметрів публіцистичного стилю української мови загалом.

Внесок авторів: Оксана Зубань – методологія, валідація даних, написання (перегляд і редагування); Владислав Кривенок – концептуалізація, формальний аналіз, програмне забезпечення, валідація даних, написання (оригінальна чернетка).

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

Бук, С. Н. (2021). *Велика проза Івана Франка: електронний корпус, частотні словники та інші міждисциплінарні контексти*. ЛНУ імені Івана Франка.

Бук, С. (2012). Кількісне зіставлення текстів (на матеріалі редакцій 1884 та 1907 років повісті Івана Франка “Воа constrictor”). *Українське літературознавство*, 76, 179–192.

Дарчук, Н. (2021). Безполучниковість у поетичних творах Лесі Українки. *Лінгвістичні студії*, 42, 117–138. <https://doi.org/10.31558/1815-3070.2021.42.12>

Зубань, О. (2020). *Комп'ютерне лексикографічне моделювання морфемної системи української мови*. ВПЦ “Київський університет”. <https://doi.org/10.17721/978-966-439-819-7>

Корпус української мови (2024). <http://www.mova.info/corpus.aspx>

Перебийніс, В. С. (Ред.). (1967). *Статистичні параметри стилів*. Наукова думка.

Перебийніс, В. С., Муравицька, М. П., & Дарчук, Н. П. (1985). *Частотні словники та їх використання*. Наукова думка.

Словник української мови: у 11 т. (1970–1980). Наукова думка.
http://ukrlit.org/slovnky/slovnky_ukrainskoi_movy_v_11_tomakh

Фаріон, І. (2013). *Мовна норма: знищення, пошук, віднова (культура мовлення публічних людей)*. Місто НВ.

Ципердюк, О. (2013). Публіцистичне мовлення Степана Бандери з погляду сучасних мовних норм. *Наукові праці [Чорноморського держ. ун-у ім. Петра Могили комплексу Кієво-Могилянська академія]*. Сер: Філологія. Мовознавство, 216, 204, 118–123.

TextAttributor 1.0. (2024). <http://ta.mova.info>

TextAttributor 1.0. Методика (2024). <http://ta.mova.info/methods>

СПИСОК ІЛЮСТРАТИВНОГО МАТЕРІАЛУ²

Бандера, С. (1946а). До засад нашої визвольної політики. *Визвольна політика*, ч. 9–10, 12–19.

Бандера, С. (1946б). До проблеми політичної консолідації. *Визвольна політика*, ч. 4–5, 6–24.

Бандера, С. (1946с). Значення широких мас та їх охоплення. *Визвольна політика*, ч. 1, 15–17.

Бандера, С. (1948а). Пляновість революційної боротьби в краю. *Визвольна політика*, ч. 1 (18), 8–18.

Бандера, С. (1948б). Слово до українських націоналістів-революціонерів за кордоном. *Видання ЗЧ ОУН*.

Бандера, С. (1950а). В десятю річницю створення Революційного Проводу ОУН (10.2.1940). *Сурма*, ч. 18–19, 1–8.

Бандера, С. (1950б). Війна в Кореї і національно-визвольна політика. *Сурма*, ч. 21, 1–3.

Бандера, С. (1950с). Друзі – українські націоналісти! *Сурма*, ч. 24, 4–5.

Бандера, С. (1950d). Третя світова війна і визвольна боротьба. *Сурма*, ч. 22–23, 1–4; ч. 24, 7–10.

Бандера, С. (1950е). Українська національна революція, а не тільки протирежимний резистанс. *Український самостійник*, ч. 3, 2–3; ч. 5, 2–3; ч. 7, 2–3; ч. 8, 2; 10, 2; ч. 11–13, 2–3; ч. 15, 2–3.

Бандера, С. (1954). Командир – провідник (слідами сл. пам. Романа Шухевича). *Шлях перемоги*, ч. 3, 2; ч. 5, 2; ч. 6, 2; ч. 7, 2.

² Узято з <https://ounuis.info/library/> (електронна бібліотека Архіву ОУН), окрім останнього джерела.

- Бандера, С. (1959). *Мої життєписні дані* (початковий машинопис автобіографії С. Бандери з редакторськими правками 1961 р. для “Гомону України”). Ф. 21 оп. 1 од. зб. 37.
- Бандера, С. (1961). *Мої життєписні дані. Гомін України*, 42/648, 2: 43/649, 2: 44/650, 2: 45/651, 2.
- Бандера, С. (1998). *Перспективи української революції*. Видавнича фірма “Відродження”. <http://dontsov-nic.com.ua/stepan-bandera-perspektyvy-ukrajinskoji-revoljutsiji/>

REFERENCES

- Buk, S. N. (2021). *Long prose fiction by Ivan Franko: electronic corpus, frequency dictionaries, and other interdisciplinary contexts*. Ivan Franko Lviv National University [in Ukrainian].
- Buk, S. (2012). Quantitative comparison of texts (based on the material of the 1884 and 1907 editions of Ivan Franko's novel “Boa constrictor”). *Ukrainian Literary Studies*, 76, 179–192 [in Ukrainian].
- Corpus of the Ukrainian language (2024) [in Ukrainian]. <http://www.mova.info/corpus.aspx>
- Darchuk, N. (2021). Non-conjunctiveness in the poetic works of Lesya Ukrainka. *Linguistic Studies*, 42, 117–138 [in Ukrainian]. <https://doi.org/10.31558/1815-3070.2021.42.12>
- Dictionary of the Ukrainian language: in 11 vol. (1970-1980). Naukova Dumka [in Ukrainian]. http://ukrlit.org/slovnyk/slovnyk_ukrainskoi_movy_v_11_tomakh
- Farion, I. (2013). *Language norm: destruction, search, renewal (culture of public speech)*. Misto NV [in Ukrainian].
- Perebyinis, V. S., Muravytska, M. P., & Darchuk, N. P (1985). *Frequency dictionaries and their use*. Naukova dumka [in Ukrainian].
- Perebyinis, V. S., (Ed.). (1967). *Statistical parameters of styles*. Naukova dumka [in Ukrainian].
- TextAttributor 1.0 (2024) [in Ukrainian]. <http://ta.mova.info>
- TextAttributor 1.0. Методика (2024) [in Ukrainian]. <http://ta.mova.info/methods>
- Tsyperdiuk, O. (2013). Publicistic speech of Stepan Bandera in terms of modern language norms. *Scientific works. Philology. Linguistics*, 216, № 204, 118–123 [in Ukrainian].
- Zuban, O. (2020) *Computer lexicographic modeling of the morpheme system of the Ukrainian language*. Publishing house “Kyiv University” [in Ukrainian]. <https://doi.org/10.17721/978-966-439-819-7>

SOURCES OF ILLUSTRATIVE MATERIAL³

Bandera, S. (1946a). To the principles of our liberation policy. *Vyzvolna polityka* [Liberation politics], pts. 9–10, 12–19 [in Ukrainian].

Bandera, S. (1946b). To the problem of political consolidation. *Vyzvolna polityka* [Liberation politics], pts. 4–5, 6–24 [in Ukrainian].

Bandera, S. (1946c). The significance of the masses and their outreach. *Vyzvolna polityka* [Liberation politics], pts. 1, 15–17 [in Ukrainian].

Bandera, S. (1948a). The planned nature of the revolutionary struggle in the region. *Vyzvolna polityka* [Liberation politics], pt. 1 (18), 8–18 [in Ukrainian].

Bandera, S. (1948b). A word to Ukrainian nationalist revolutionaries abroad. *Vyd. ZCh OUN* [in Ukrainian].

Bandera, S. (1950a). On the tenth anniversary of the establishment of the OUN Revolutionary Leadership (10.2.1940). *Surma* [Trumpet], pts. 18–19, 1–8 [in Ukrainian].

Bandera, S. (1950b). The Korean War and National Liberation Policy. *Surma* [Trumpet], pt. 21, 1–3 [in Ukrainian].

Bandera, S. (1950c). Friends – Ukrainian nationalists! *Surma* [Trumpet], pt. 24, 4–5 [in Ukrainian].

Bandera, S. (1950d). World War III and the liberation struggle. *Surma* [Trumpet], pts. 22–23, 1–4; pt. 24, 7–10 [in Ukrainian].

Bandera, S. (1950e). Ukrainian national revolution, not just anti-regime resistance. *Ukrainskyi samostiinyk* [Ukrainian independent], pt. 3, 2–3; pt. 5, 2–3; pt. 7, 2–3; pt. 8, 2; pt. 10, 2; pts. 11–13, 2–3; pt. 15, 2–3 [in Ukrainian].

Bandera, S. (1954). Commander – leader (in the footsteps of the glorious memory of Roman Shukhevych). *Shliakh peremohy* [The path of victory], pt. 3, 2; pt. 5, 2; pt. 6, 2; pt. 7, 2 [in Ukrainian].

Bandera, S. (1959). *My life story* (initial typescript of S. Bandera's autobiography with editorial changes in 1961 for the "Voice of Ukraine"). F. 21 op. 1 od. zb. 37 [in Ukrainian].

Bandera, S. (1961). My life story. *Homin Ukrainy* [Voice of Ukraine], 42/648, 2; 43/649, 2; 44/650, 2; 45/651, 2 [in Ukrainian].

Bandera, S. (1998). Prospects for the Ukrainian revolution. Vydavnycha firma "Vidrodzhennia" [in Ukrainian]. <http://dontsov-nic.com.ua/stepan-bandera-perspektyvy-ukrajinskoji-revolutsiji/>

Отримано редакцією збірника / Received: 01.04.24

Прорецензовано / Revised: 24.04.24

Схвалено до друку / Accepted: 29.04.24

³ Retrieved from <https://ounuis.info/library/>, except last source.

Oksana ZUBAN, PhD (Philol.), Assoc. Prof.

ORCID ID: 0000-0002-2644-3892

e-mail: oxana.zuban@knu.ua

Taras Shevchenko National University, Kyiv, Ukraine

Vladyslav KRYVENOK, Student

ORCID ID: 0009-0006-1070-7378

e-mail: vladyslavkryvenok@knu.ua

Taras Shevchenko National University, Kyiv, Ukraine

**FROM WORD TO NUMBER:
LINGUISTIC AND STATISTICAL “PORTRAIT”
OF THE IDIOLECT OF STEPAN BANDERA**

Background. The personality of S. Bandera has repeatedly been the subject of research by scholars in terms of non-linguistic categories. However, one of the most expressive aspects of the figure – the linguistic one – has not yet been the subject of systematic and in-depth study. Furthermore, the journalistic heritage of S. Bandera has not yet been subjected to a systematic linguistic and statistical analysis, which renders this study a valuable contribution to the field. The objective of this article is to provide a statistical parameterisation of the lexical and part-of-speech structure of S. Bandera’s texts, based on a textual sample of 10 journalistic articles (first editions).

Methods: the method of retrospective editing of the text of reprints, statistical methods (calculation of statistical parameters, comparison of empirical data with the confidence interval of Ukrainian journalism, the method of statistical modelling), and the method of graphical presentation of statistical data.

Results: 1) the articles from the 1978 edition were edited retrospectively and the textual material was digitised and categorised in a format suitable for automatic processing; 2) a corpus of all digitised first editions of articles was compiled; 3) the text of the articles – from the first edition and the 1978-edition – were compared; 4) frequency dictionaries in the Ukrainian Language Corpus were compiled based on the reconstructed texts; 5) a stylometric model of statistical research was formed, and several parameters of the statistical structure (lexical and part-of-speech) of S. Bandera’s text were calculated; 6) empirical values of linguistic and statistical parameters calculated manually were compared with confidence intervals of the media style and with the automatically calculated empirical data of the TextAttributor 1.0 web application.

Conclusions. *The study resulted in the creation of the inaugural lexicographic system of frequency dictionaries in Ukrainian computer lexicography, based on the initial editions of S. Bandera's articles. The linguistic and statistical modelling of the texts allows us to speak of the systematisation of statistical data and the formation of a linguistic and statistical "portrait" of S. Bandera's idiolect. This is a set of all empirical data on the statistical structure of his texts, comprising nine statistical parameters, which have been calculated and compared with the confidence intervals of the statistical parameters of the journalistic style of the Ukrainian language in general.*

Key words: *Ukrainian language, idiolect, stylometric model, statistical parameter, corpus of texts, frequency dictionary.*

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; у рішенні про публікацію результатів.