

путем определения ядра, основной и периферической систем акцентных типов и структур. Подавляющее количество заимствований, которые участвуют в акцентном варьировании, составляют франкоязычные слова. В британской произносительной норме наблюдается более полная адаптация франкоязычных слов к акцентным закономерностям английского языка в сравнении с американским произносительным вариантом, но в поздних французских заимствованиях сохраняется тенденция приближения ударения к произношению языка-источника. Значительная часть исследуемых иностранных слов имеет частичную ассимиляцию акцентных структур. Произношение таких заимствований приближено к акцентным закономерностям языка-источника. Стремление носителей языка сохранить оригинальное произношение заимствованных лексем свидетельствует о повышенном интересе к другим лингвокультурам, владении иностранными языками и расширении межъязыковых контактов.

Перспективой дальнейших исследований фонологического нормирования иноязычных лексических заимствований является установление специфики процесса ассимиляции акцентных структур в аспекте других произносительных вариантов английского языка в сопоставительном плане с британской и американской произносительными нормами.

Ключевые слова: ассимиляция, акцентные типы и структуры, свободное акцентное варьирование, британская произносительная норма, американская произносительная норма, лексические заимствования.

Komar Liudmyla. Assimilation of Accentual Structures of Borrowings in the Aspect of British and American Pronouncing Norms. The paper views the assimilation of accentual structures of lexical borrowings in the aspect of British and American codified pronunciation norms. In the process of analysis, the genetic characteristics of the foreign words under study is described, accentual types that are involved in the formation of the accentual structures of lexical borrowings are identified, the peculiarities of the accentual structures of borrowed lexemes in the investigated variants are revealed, the accentual types, structures and varieties that are involved in the accentual variation of foreign words in the main variants of English in a comparative aspect are determined.

According to the syllabic structure of the word, two-syllable borrowings are the most numerous ones in the investigated body of lexical units. Nouns perform the function of nomination and respectively prevail according to the lexical-grammatical characteristics.

The results of the study indicate that coexistence within the norm of the accentual variants of words, which are formed by means of variation of place and degree of stress, is a special feature of the normalization of accentual structures of borrowings. The system of accentual structures with variants is established by determining the core, the main and peripheral systems of accentual types and structures. The vast majority of borrowings that take part in the accentual variation are French words. In British pronunciation norm French words are more adapted to the accentual laws of the English language compared to American pronunciation variant, but late French borrowings show the tendency of approaching the stress to the pronunciation of the source language. A significant part of the foreign words under study has a partial assimilation of accentual structures. The pronunciation of such borrowings is close to the accentual laws of the source language. The desire of native speakers to retain the original pronunciation of borrowed lexemes indicates an increased interest in other linguistic cultures, a good command of foreign languages and the expansion of interlanguage contacts.

The prospect for further research of the phonological assimilation of foreign lexical borrowings is the establishment of the specificity of the process of assimilation of accentual structures in the aspect of other pronouncing variants of the English language in contrastive way with British and American pronunciation norms.

Key words: assimilation, accentual types and structures, free accentual variation, British pronouncing norm, American pronouncing norm, lexical borrowings.

УДК 81'32/33'42:[821.161.2-3"20"]

Ігор Кульчицький, Олена Левченко, Наталія Лотоцька

ДОСЛІДЖЕННЯ ЧАСТОТНОСТІ БІГРАМ НА МАТЕРІАЛІ ТЕКСТІВ РОМАНА ІВАНІЧУКА, ВАСИЛЯ ШКЛЯРА ТА ЛІНИ КОСТЕНКО

Статтю присвячено перспективам дослідження ідіолекту письменника в сучасних мовознавчих студіях. Розглянуто термін *ідіолект* та подано його трактування вітчизняними й зарубіжними лінгвістами. Зауважено, що для багатоаспектного вивчення авторського мовлення варто здійснити кількісний опис творчого доробку письменника. Подано визначення *корпусу текстів* (КТ) та його значення для проведення лінгвістичних досліджень. Зазначено, що окремий текст є системою, об'єднаною комунікативною цілісністю, логічними, граматичними та стилістичними зв'язками. Особливу увагу приділено значенню статистичного дослідження, що застосовується для виявлення особливостей функціональних стилів мови та особливостей стилю окремих авторів, для автоматичного опрацювання тексту та визначення його авторської і часової атрибуції. Визначено, що важливими характеристиками мови є повторюваність букв (монограм), пар букв (біграм), n-грам. Для

проведення дослідження частотності біграм у тексті було взято твори Р. Іваничука, В. Шкляра та Л. Костенко та підраховано ентропію біграм для відповідних масивів текстів. У результаті проведеного дослідження проаналізовано частотність біграм символів, а результати подані у вигляді таблиць та діаграм. Згодом ці дані можна буде використати для порівняння текстів інших авторів.

Ключові слова: ідіолект, корпус текстів, статистичні методи, статистичний параметр, частотний аналіз, ентропія, біграма.

Постановка наукової проблеми та її значення. Мова диктує свої закони кожному, хто нею користується, і ступінь підсвідомого чи свідомого засвоєння цих законів мовцем позначається на якості добору й розташування мовного матеріалу в мовленні [9, с. 188].

З розвитком комп'ютерних технологій з'явилися нові сучасні принципи та методи вивчення й опрацювання тексту, зокрема мови письменників, серед яких – статистичний та корпусний, що привертають особливу увагу мовознавців.

Для багатоаспектного дослідження ідіолектів лінгвісти вдаються до кількісного опису доробку письменників, що дає точну інформацію про особливості функціонування лексики в їхній творчості [2, с. 86].

Аналіз дослідження проблеми. Серед вчених, що досліджували різні аспекти індивідуального мовлення, варто згадати британського фонетиста Д. Аберкромбі, американського структураліста Ч. Хоккета, британського лінгвіста М. Халлідея, французького лінгвіста А. Мартіне і філософів М. Даммета й Д. Девідсона.

Першими мовознавцями, які підходили до описання мови і мовних елементів як системних зв'язків, були Ф. де Соссюр та І. Бодуен де Куртене. Англійський учений Дж. Юла, німецький науковець Е. Форстеманн, російські математики В. Буняковський, А. Марков та ін. почали першими використовувати математичні, зокрема й статистичні, методи для лінгвістичного опису.

Іншої думки був А. Росс [16], який вважав, що теорія імовірностей і математична статистика можуть бути інструментом для перевіряння й підтвердження тих лінгвістичних висновків, що допускають числове трактування. Варто також згадати праці з математичної лінгвістики Р. Піотровського [11], В. Фукса [18], Дж. Ципфа [23], дослідження про відношення між рангом і частотністю слова, здійснені американцями К. Шенноном та У. Уівером [22], дослідження теорії інформації для лінгвістичного опису Г. Хердана [19].

В Україні застосуванням квантитативних та корпусних методів у лінгвістиці активно займаються В. Широков, В. Перебийніс, Н. Дарчук, Є. Карпіловська, А. Загнітко, В. Левицький та ін.

Мета цієї статті – визначити роль статистичного аналізу текстів у мовознавстві. **Завдання** статті – дослідити частотність біграм символів у текстах Р. Іваничука, В. Шкляра та Л. Костенко, що дозволить виявити особливості сполучення символів в індивідуальному мовленні письменників.

Виклад основного матеріалу й обґрунтування отриманих результатів дослідження. Вивчення ідіолекту автора стало одним із провідних напрямків сучасного мовознавства. Деякі характеристики автентичності мови можна виявити лише із застосуванням корпусу текстів, як-от: багатство словника, індекси різноманітності та винятковості тощо.

Загалом термін *ідіолект* трактують лише як сукупність індивідуальних способів функціонування мови. Зарубіжні вчені, серед яких Ж. Дюбуа, Л. Геспен, Т. Левандовські, трактують ідіолект як сукупність особливостей функціонування конкретної мови окремим індивідом у певний момент часу, а також як лінгвістичні константи, що слугують основою для вираження думок мовця, концентруючи свою увагу на протиставленні загальнонародних особливостей мови, властивих окремим мовним спільнотам (діалект, говір, соціолект), безпосередньо ідіолекту [17, с. 338–339].

Варто зауважити, що в західній лінгвістиці досі не існує чітко розробленого поняття *ідіолект*, а терміна *ідіостиль* немає в поняттєвому апараті. У найпростішому тлумаченні *ідіолект* часто виступає синонімом *стилю*. На думку М. Арріве, ідіолект – це сукупність особливостей використання мови індивідом, тоді як Ч. Хоккет зазначає, що це сукупність звичок, які характеризують мовлення індивіда, що належить до певної лінгвістичної групи [20].

Вивчення мовних явищ із використанням корпусу текстів (далі – КТ) дає змогу не тільки підтвердити чи спростувати окремі аспекти теорії, а й уможливило дослідження мови письменника, його стилю, а також – виявлення закономірностей функціонування мовних одиниць на різних рівнях.

За І. Кульчицьким, КТ – це призначена для різноманітних мовознавчих досліджень лінгвістично описана сукупність мовних чи мовленнєвих фактів, поданих в електронній формі та споряджених відповідним спеціалізованим програмним забезпеченням [7, с. 29].

Останнім часом прикладні науки та взаємодія різних наукових дисциплін відіграють важливу роль для отримання спільних результатів у різних сферах науки та життєдіяльності людини і суспільства. Значну увагу приділяють використанню статистичних методів для потреб лінгвістики.

Методи статистичної лінгвістики використовують для виявлення особливостей функціональних стилів мови та особливостей стилю окремих авторів, для створення методики навчання іноземної мови й оптимізації навчання, для розв'язання завдань зі стенографії тощо. Особливо важливі статистичні методи для розвитку автоматичного опрацювання тексту – автоматичного зняття полісемії, реферування та індексування, а також для визначення авторської та часової атрибуції твору.

Статистичні методи в сучасному мовознавстві стали важливим та надійним інструментом аналізу мовних даних. Окрім забезпечення вірогідності результатів, статистичні методи дають змогу розкрити властивості одиниць мови та будови тексту, виявити, які без цих досліджень було б неможливо. Ефективність статистичних методів зумовлена такими чинниками:

- результати дослідження базовані на точних числових даних, які завжди можна перевірити, замість приблизних якісних на кшталт «часто», «незначно» та ін.;
- точно визначають істотність чи випадковість коливань зіставляваних показників;
- дають змогу визначити таку кількість матеріалу для досліджень, яка є необхідною та достатньою для вірогідних висновків;
- вимагають формалізації під час визначення одиниць підрахунків, що мінімізує чинник суб'єктивності, за необхідності дає змогу перевірити результати та за умови дотримання правил підрахунків забезпечити високу вірогідність результатів;
- допомагають мовознавцям позбутися хибних уявлень та тверджень [12, с. 29].

В. Перебийніс вважає, що всі складні системи підпорядковуються статистичним законам, наголошуючи на необхідності використовувати статистичні методи в мовознавстві, адже мова є складною системою, а відтак підпорядковується статистичним законам [13].

До системних властивостей мови належать ті, що характеризують конкретну її одиницю як компонента мовної системи. Зокрема, це – будова одиниці, її належність до граматичного класу, синтаксична та лексична сполучуваність, місце в мовній системі та підсистемах, накладені мовною системою обмеження (наприклад, відсутність деяких словозмінних форм іменників, відсутність ступенів порівняння прикметників і прислівників, неповна дієслівна парадигма), семантичні особливості тощо [10, с. 28].

Представники теорії тексту вважають, що кожний окремий текст є системою, об'єднаною комунікативною цілісністю, логічними, граматичними та стилістичними зв'язками [1]. Особливості вживання конкретної одиниці в певному тексті визначають її функційні властивості: частота, позиція, сполучуваність, яка залежить від характеру тексту, від функціонального чи авторського стилю і змінюється від тексту до тексту [10].

Найпростішим різновидом статистичного підходу до вивчення мови, зокрема мови письменників або окремих творів, є підрахунок уживаності слів, оскільки багатство словника в певний спосіб характеризує їхню мову. Для кожного письменника, як і будь-якого мовця, характерна своя специфічна частотність мовних елементів, іншими словами, кожному авторові притаманні свої улюблені, а тому й частотні слова, словосполучення, фрази, синтаксичні конструкції тощо.

У монографії «Статистичні параметри стилів» [13] зазначено, що для текстів статистичними параметрами стилів можна вважати частоти голосних, приголосних, пропуски між словами, а також м'яких і сонорних груп приголосних. Кожна мова має власні статистичні

параметри, і знання частоти появ літер та їхніх сполучень (біграм, триграм, чотириграм) певної мови дає змогу автоматично її ідентифікувати.

Важливими характеристиками мови є повторюваність букв (монограм), пар букв (біграм), n-грам, сполучуваність букв одна з одною, чергування голосних і приголосних тощо. Послідовність n сусідніх літер тексту називається n-грамою. Із тексту довжиною N можна утворити різних n-грам. Наприклад, слово *стаття* (текст довжиною 6) розбивається на 5 біграм: *стаття* = *ст та ат тт тя* [24].

Частотний аналіз передбачає, що частота появи заданої літери алфавіту в досить довгих текстах та сама для різних текстів однієї мови. Частотні характеристики коливаються залежно від тематики твору й автора. Отож, вони можуть бути також використані для підтвердження або спростування авторства.

Матеріалом дослідження для вивчення частотності біграм символів у текстах слугували художні твори Р. Іваничука [4; 5; 6], В. Шкляра [14; 15; 16], Л. Костенко [8]. Перед дослідженням усі тексти за допомогою програм мовою Python перевірено на коректність кодів-символів та здійснено необхідне коректування.

Варто зазначити, що ці постаті знакові для української літератури др. пол. ХХ–поч. ХХІ ст. Р. Іваничук та В. Шкляр відомі своїми історичними романами та повістями, тоді як Л. Костенко – культова українська поетеса-шістдесятниця. Наше завдання полягає в аналізі й порівнянні частоти біграм-символів у романах, повістях та віршах окремих.

Для дослідження обрані тексти розділено на три масиви: твори Р. Іваничука, В. Шкляра та Л. Костенко. У текстах залишено лише букви української абетки, пробіл, апостроф та дефіс (розширений алфавіт української мови). Для кожного масиву тексту обраховано кількість біграм розширеної абетки, їхню частоту та ентропію. Загальну характеристику масивів текстів подано в таб. 1.

Таблиця 1

Загальна характеристика масивів творів

Автор	Кількість рядків	Кількість символів
Роман Іваничук	14484	1564272
Василь Шкляр	8985	970380
Ліна Костенко	2198	237384

У процесі дослідження частоти біграм-символів виконано такі дії:

Дія 1. Для масивів текстів романів, повістей та поезії пораховано кількість кожного члена множини символів, їхню частоту та ентропію.

Дія 2. У масивах вищезгаданих текстів вибрано послідовно десять сегментів, інтервал між якими становив 0,02 довжини тексту. Для кожного з них повторено дію 1.

Дія 3. У відповідних масивах текстів вибрано десять випадкових сегментів, розмір яких порівнюваний із послідовними сегментами. Для кожного сегмента повторено дію 1.

Ентропію обраховано за формулою:

$$H = - \sum_{i=1}^{36} p_i \log_2 p_i ,$$

де p_i – частота переходу букв одна в одну в біграмах.

Під час опрацювання отримано результати, що подано в таблицях 2, 3, 4.

Таблиця 2

Загальна характеристика масивів творів Р. Іваничука

	Кількість		Ентропія
	рядків	біграм	
Весь текст	14484	1578755	114,60
Послідовний сегмент (середнє)	1159	126331	110,97
Випадковий сегмент (середнє)	1448	155229	112,45

Таблиця 3

Загальна характеристика масивів творів В. Шкляра

	Кількість		Ентропія
	рядків	біграм	
Весь текст	8985	979364	113,79
Послідовний сегмент (середнє)	719	78371	109,45
Випадковий сегмент (середнє)	898	96265	110,97

Таблиця 4

Загальна характеристика масивів творів Л. Костенко

	Кількість		Ентропія
	рядків	біграм	
Весь текст	2198	239581	114,84
Послідовний сегмент (середнє)	176	19184	108,46
Випадковий сегмент (середнє)	219	23474	109,43

Як бачимо, значення величини ентропії дещо відрізняються в текстах Р. Іваничука, В. Шкляра, Л. Костенко. Характер коливань значень ентропії біграми представлено на діаграмах (рис. 1-2).

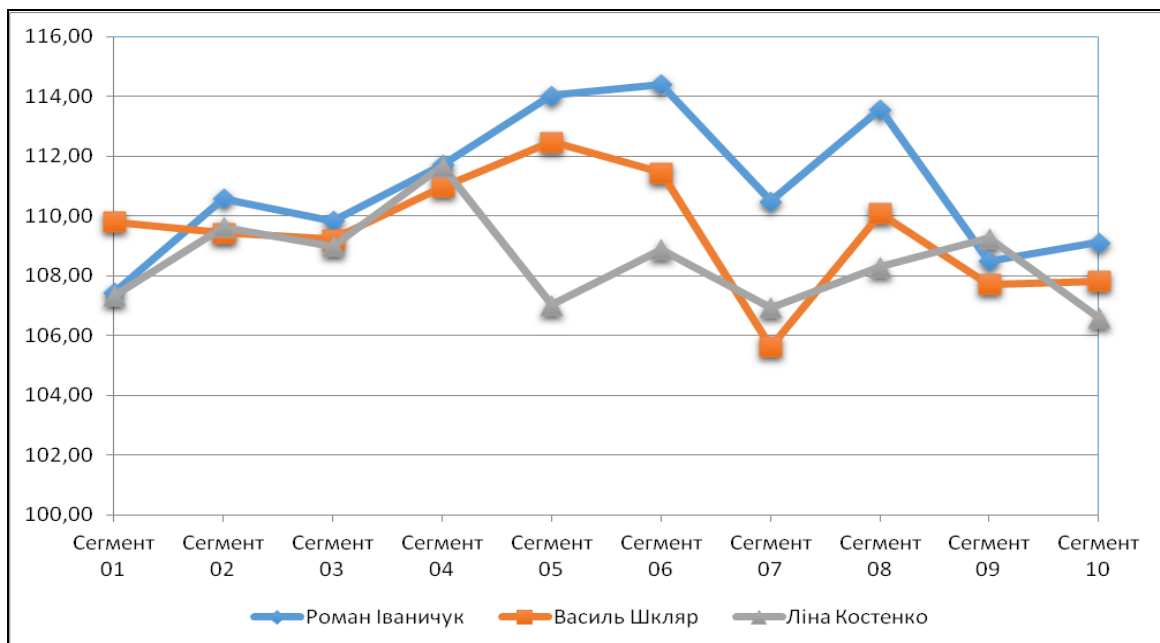


Рис. 1. Ентропія біграм послідовна

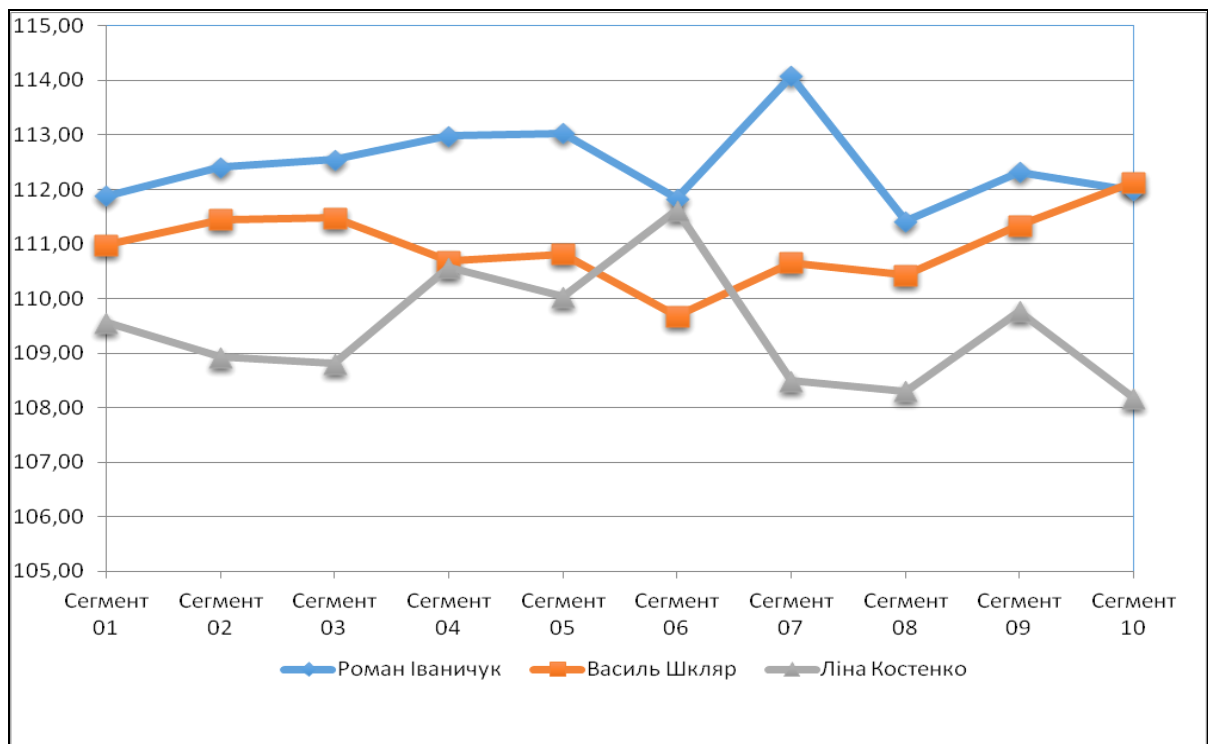


Рис. 2. Ентропія біграм випадкова

Як бачимо, на діаграмах 1 та 2 розкид значень середньої ентропії біграм у послідовних сегментах більший, ніж у випадкових. Крім цього, значення ентропії біграм відрізняються відносно авторства, хоча й простежуються сегменти дотику й перетину.

Під час опрацювання масивів текстів виявлено 10 найчастотніших біграм у текстах письменників, результати дослідження подано в таб. 5.

Таблиця 5

10 найчастотніших біграм у текстах

Роман Іваничук		Василь Шкляр		Ліна Костенко	
Біграма	Кількість	Біграма	Кількість	Біграма	Кількість
НА	17854	НА	11328	НА	2385
СТ	13477	НЕ	8965	НЕ	2017
РО	13452	ТА	7402	ЛИ	1942
ВІ	12606	ТИ	7316	РО	1839
ЛИ	12328	ВІ	7182	ВІ	1750
НЕ	12226	ЛА	6993	СТ	1689
ТИ	12069	ПО	6934	НІ	1664
ГО	11975	РО	6885	ТА	1600
ПО	11880	АЛ	6884	ТЬ	1553
ОВ	11800	ГО	6819	ТИ	1518

У цій таблиці представлено інформацію про сполучуваність символів, тяжіння символів один до одного. Результати дослідження подано в порядку спадання частоти відповідних біграм. Варто зауважити, що найчастотніша й спільна для всіх письменників є біграма «НА» і згодом можна буде простежити, у яких частинах мови ця біграма трапляється найчастіше.

Висновки та перспективи подальшого дослідження. У результаті дослідження частотності біграм-символів встановлено, що вибір із текстів сегментів для дослідження у

випадковий спосіб дає стабільніші значення показника, ніж вибір послідовних сегментів. Крім того, визначено найчастотнішу й спільну біграму для трьох письменників. Звернімо увагу, що вибрана методика підрахунків не дає змоги зі статистичною достовірністю визначити авторську атрибуцію великих масивів текстів. Зокрема, необхідними є подальші дослідження для визначення показника, що уможливить розрізнення стилів. Зауважимо, що отримані результати опрацювання текстів можна буде використати для порівняння текстів інших авторів, а також вони сприяють підтвердженню раніше виявлених та визначенню нових статистичних параметрів і закономірностей будови текстів, їхніх структурних особливостей.

Джерела та література

1. Арнольд И. В. Основы научных исследований в лингвистике : учеб. пособ. – М. : Высш. шк., 1991. – 140 с.
2. Бук С. Сучасні методи дослідження мови письменника у слов'язнознавстві / С. Бук // Проблеми слов'язнознавства. – 2012. – Вип. 61. – С. 86–95.
3. Демська-Кульчицька О. М. Базові поняття корпусної лінгвістики / О. М. Демська-Кульчицька // Українська мова. – 2003. – №1. – С. 42–47.
4. Іванчук Роман. Євангеліє від Томи: триптих повістей / Роман Іванчук. – Харків : Фоліо, 2011. – 316 с.
5. Іванчук Роман. Манускрипт з вулиці Руської: іст. роман / Роман Іванчук. – Львів : Піраміда, 2011. – 201 с.
6. Іванчук Роман. Хресна проща: романний триптих / Роман Іванчук. – Львів : Піраміда, 2011. – 281 с.
7. Кульчицький І. М. Технологічні аспекти укладання корпусів текстів / І. М. Кульчицький // Дані текстових корпусів у лінгвістичних дослідженнях : монографія / В. А. Широков, І. В. Шевченко, А. П. Загнітко та ін. ; за ред. О. П. Левченко. – Львів : Вид-во «Львівської політехніки», 2015. – С. 29–45.
8. Костенко Л. В. Вибране / Л. В. Костенко. – К. : Дніпро, 1989. – 559 с.
9. Павличко О. О. Щодо статистичних параметрів авторського стилю (на матеріалі творів Е. М. Ремарка) / Оксана Олексіївна Павличко // Мовні і концептуальні картини світу : зб. наук. пр. – К. : ВПЦ «Київський універ-ситет», 2010. – Вип. 29. – С. 186–191.
10. Перебийніс В. І. Що дає статистика лінгвістам? // Вісник Київського лінгвістичного університету. Серія філологія. Т. VI. – 2003. – № 2. – С. 27–32.
11. Пиотровский Р. Г. Математическая лингвистика : учеб. пособие / Р. Г. Пиотровский. – М. : Высш. шк., 1977. – 383 с.
12. Ставицька Л. Про термін ідіолект / Л. О. Ставицька // Українська мова. – 2009. – № 4. – С. 3–15.
13. Статистичні параметри стилів / за ред. В. С. Перебийніс. – Київ : Наукова думка, 1967. – 260 с.
14. Шкляр В. Елементал / В. Шкляр. –Х. : Книжковий клуб «Клуб сімейного дозвілля», 2013. – 222 с.
15. Шкляр В. Кров кажана / В. Шкляр. –Х. : Книжковий клуб «Клуб сімейного дозвілля», 2013. – 272 с.
16. Шкляр В. Тінь сови / В. Шкляр. –Х. : Книжковий клуб «Клуб сімейного дозвілля», 2014. – 304 с.
17. Diccionario de Lingüística / Dubois, J., Giacomo, M., Guespin, L., Marcellesi, C., Marcellesi, J-P., Mével. J-P. – Madrid : Alianza, 1994. – 636 p.
18. Fucks W Mathematische Analyse von Sprachelementen, Sprachstil und Sprachen / W. Fucks. – Koln und Oplade, 1955. – 110 p.
19. Herdan G. Language as Choice and Chance / G. Herdan. – Noordhoff : Groningen, 1956. – 356 p.
20. Neveu F. Dictionnaire des sciences du langage / F. Neveu. – Paris : A. Colin, 2004. – [Electronic resource]. – Mode of access : <http://www.franckneveu.fr/mediapool/76/768102/data/IDIOLLECTE.pdf>
21. Ross A. Philological Probability Problems / A. Ross // «Statist. Soc.». – Vol. XII. – 1950. – P. 19–59.
22. Shannon C. Weaver W. The Mathematical Theory of Communication / C. Shannon, W. Weaver. – Univ of Illinois Press, 1949. – 117 p.
23. Zipf G. K. Human Behavior and the Principle of Least Effort / G. K. Zipf. – Addison-Wesley, 1949. – 573 p.
24. Практическая криптология [Електронний ресурс] / С. А. Сушко. – Режим доступу : <http://bit.nmu.org.ua/ua/student/metod/cryptology/%D0%BB%D0%B5%D0%BA%D1%86%D0%B8%D1%202.pdf>

References

1. Arnold, I. V. 1991. *Osnovy nauchnykh issledovaniy v lingvistike*. Moscow: Vyssh. shk.
2. Buk, S. 2012. “Suchasni metody doslidzhennia movy pysmennyka u slovianoznavstvi”. *Problemy slovianoznavstva*, 61: 86–95.
3. Demska-Kulchitska, O. M. 2003. “Bazovi poniatyia korpusnoi linhvistyky”. *Ukrainska mova*, 1: 42–47.
4. Ivanychuk, Roman. 2011. *Yevanhelie vid Tomy: tryptykh poviste*. Kharkiv: Folio.
5. Ivanichuk, Roman. 2011. *Manuskrypt z vulytsi Ruskoj*. Lviv: Piramida.
6. Ivanichuk, Roman. 2011. *Khresna proshcha*. Lviv: Piramida.
7. Kulchitskyi, I. M. 2015. *Tekhnolohichni aspekty ukladannia korpusiv tekstiv*. Lviv: Vydavnytstvo Lvivskoy politekhniky.
8. Kostenko, L.V. 1989. *Vibrane*. K.: Dnipro.
9. Pavlychko, O. O. 2010 “Shchodo statystychnykh parametriv avtorskoho styliu (na materialy tvoriv E. M. Remarka)”. *Movni i kontseptualni kartyny svitu*, 29:186–191.

10. Perebyinis, V. I. 2003. "Shcho daie statystyka linhvistam?". *Visnik Kyyskoho linhvistychnoho universytetu*, 4: 227–32.
11. Piotrovskiy, R. G. 1977. *Matematicheskaia lingvistika*. Moskva: Vyssh. shk.
12. Stavyska, L. 2009. "Pro termin idiolekt". *Ukrainska mova*, 4: 3–15.
13. *Statystichni parametry styliv* edited by V. S. Perebyinis. 1967. Kyiv: Naukova dumka.
14. Shklyar, V. 2013. *Elemental*. Kharkiv: Knyzhkovyi klub «Klub simeinoho dozvillia».
15. Shklyar, V. 2013. *Krov kazhana*. Kharkiv: Knyzhkovyi klub «Klub simeinoho dozvillia».
16. Shklyar, V. 2014. *Tin sovy*. Kharkiv: Knyzhkovyi klub «Klub simeinoho dozvillia».
17. *Diccionario de Lingüística*, edited by Dubois, J., et al. 1994. Madrid: Alianza.
18. Fucks, W. 1995. *Mathematische Analyse von Sprachelementen, Sprachstil und Sprachen*. Koln und Oplade.
19. Herdan, G. 1956. *Language as Choice and Chance*. Noordhoff: Groningen.
20. Neveu, F. 2004. *Dictionnaire des sciences du langage*. Paris: A. Colin.
21. Ross, A. 1950. "Philological Probability Problems". *Statist. Soc.* XII: 19–59.
22. Shannon, C. Weaver W. 1949. *The Mathematical Theory of Communication*. Univ of Illinois Press.
23. Zipf, G. K. 1949. *Human Behavior and the Principle of Least Effort*. Addison-Wesley.
24. *Practicheskaia kriptologija*. <http://bit.nmu.org.ua/ua/student/metod/cryptology/%D0%BB%D0%B5%D0%BA%D1%86%D0%B8%D1202.pdf>

Кульчицкий Игорь, Левченко Елена, Лотоцкая Наталья. Исследование частотности биграмм на материале текстов Романа Иванычука, Василя Шкляра и Лины Костенко. В статье речь идет о перспективах исследований идиолекта писателя в современных языковедческих студиях. Замечено, что для многоаспектного изучения авторской речи следует осуществить количественное описание творчества писателя. Дано определение *корпуса текстов* (КТ) и его значение для проведения лингвистических исследований. Отмечено, что отдельный текст является системой, объединенной коммуникативной целостностью, логическими, грамматическими и стилистическими связями. Особое внимание уделено значению статистического исследования, что применяется для выявления особенностей функциональных стилей языка и особенностей стиля отдельных авторов, для автоматической обработки текста и определения его авторской и временной атрибуции. Определено, что важными характеристиками языка является повторяемость букв (монограмм), пар букв (биграмм), n-грамм. Для проведения исследования частотности биграмм было взято тексты Р. Иванычука, В. Шкляра и Л. Костенко и подсчитано энтропию биграмм для соответствующих массивов текстов. В текстах было оставлено только буквы украинского алфавита, пробел, апостроф и дефис (расширенный алфавит украинского языка). Для каждого массива текста было подсчитано количество биграмм расширенной азбуки, их частоту и энтропию. В процессе обработки текстов было установлено, что выбор в текстах сегментов для исследования случайным образом дает стабильные значения показателя, чем выбор последовательных сегментов. Кроме того, определено самую распространенную и общую биграмм для трех писателей. Замечено, что выбранная методика подсчетов не позволяет со статистической достоверностью определить авторскую атрибуцию больших массивов текстов. В ходе проведенного исследования проанализированы частотность биграмм символов, а результаты представлены в виде таблиц и диаграмм. Впоследствии эти данные можно будет использовать для сравнения текстов других авторов.

Ключевые слова: идиолект, корпус текстов, статистические методы, статистический параметр, частотный анализ, энтропия, биграмма.

Kulchytskyi Igor, Levchenko Olena, Lototska Nataliia. Analysis of the Bigrams Frequency in Roman Ivanychuk's, Vasyl Shklar's and Lina Kostenko's Texts. The study of writer's idiolect has become one of the leading areas of modern linguistics. The article presents the importance of computer technologies opportunities in research and description of language. The research deals with modern techniques and their value for studying writer's language in linguistics. The methods of corpus linguistics concern such lingual phenomena in corpus texts that are difficult to explore without computer technologies. The term idiolect and his interpretation by Ukrainian and foreign linguists are presented. For multidimensional study of writer's speech, it is worth doing a quantitative analysis of writer's text. The definition of corpus text and its importance in linguistic research are given. The text is referred to as the highest level in the language system, along with linguistic units such as phoneme, morpheme, word, phrase, sentence, united by communicative integrity, logical, grammatical and stylistic connections. Statistical research is used to identify the features of the functional language styles and individual author's style, for the automatic processing of the text and determination of its author's and time attribution. The important characteristics in the language is the letter frequency (monograms), pairs of letters (bigrams), n-grams. The material of the research is Roman Ivanychuk's, Vasyl Shklar's and Lina Kostenko's texts. Studying the bigrams frequency the entropy of the bigrams for the corresponding arrays of texts was analysed. In the texts we left the letters of the Ukrainian alphabet, the gap, the apostrophe and the hyphen (extended alphabet of the Ukrainian language). For each array of text, the number of extended alphabets, its frequency and entropy was calculated. In the process of texts analysis we investigated the selection of segment texts for random research yields a more stable value of the index than the choice of successive segments. In addition, the most frequent and common bigram for three writers has been observed. In consequence of the research, the bigrams frequency is analyzed, and the results are presented in the form of lists and diagrams. Subsequently, these data can be used to compare the texts of other authors.

Key words: idiolect, corpus text, statistical methods, statistical parameter, frequency analysis, entropy, bigrams.