

DOI: <https://doi.org/10.32839/2304-5809/2019-11-75-206>

УДК 811.161.2'322.2'324'38'42:82-31Р.Іваничук

Кульчицький І.М., Гнатовська К.І.
Національний університет «Львівська політехніка»**ДОСЛІДЖЕННЯ ІДІОЛЕКТУ ПИСЬМЕННИКА: КОРПУСНО-СТАТИСТИЧНИЙ ПІДХІД**

Анотація. У статті розглянуто поняття ідіолекту та індивідуального стилю письменника на прикладі Р. Іваничука у сучасних мовознавчих дослідженнях. Розкрито суть термінів «ідіостиль» та «ідіолект», описано методи та підходи до аналізу індивідуального стилю автора у статистичній та корпусній лінгвістиці, порівняно дослідження вітчизняних та зарубіжних науковців. Зазначено, що корпусна лінгвістика та статистичний підхід до вивчення ідіолекту письменника є передовими засобами технологій у лінгвістичних дослідженнях сучасних мовознавчих структур. Визначено низку основних підходів до вивчення ідіостилу автора: метод градуальних опозицій, лінгвостатистичний підхід та квантитивний методи, який і є ключовим у корпусній лінгвістиці.

Ключові слова: ідіостиль, ідіолект, корпусна лінгвістика, лінгвостатистика.

Kulchytskyi Igor, Hnatovska Kateryna
Lviv Polytechnic National University**INVESTIGATION OF THE WRITER'S IDIOLECT: A CORPUS-STATISTICAL APPROACH**

Summary. This research deals with the notion of idiostyle in modern linguistic investigations on the example of Roman Ivanychuk's writings. Firstly, we reveal the essence of the terms "idiostyle" and "idiolect". Secondly, there are described methods and approaches to the analysis of the author's individual style in statistical and corpus linguistics. Also in the study there are compared domestic and foreign scholars investigations on this field. The study speaks about importance of corpus linguistic in modern linguistic studies. There are several basic approaches to the study of the author's idiostyle: the method of gradual oppositions, the linguistic statistical approach, and the quantitative method, which is key in corpus linguistics. Roman Ivanychuk's novels are a thorough description of the history of Ukraine. However, statistical studies of Roman Ivanychuk's works have been carried out at an insufficient level, so creation of a corpus of texts based on author's works is an important contribution to the development of the corpus of the Ukrainian language, as well as a basis for further linguistic statistics. In fact, quantitative research methods, which are considered to be the latest in author's style analysis, are becoming increasingly relevant with the development of computer technologies and the availability of related software. As a result, corpus linguistics has been developing significantly in recent times, as it has been able to cover various aspects of the study of linguistic phenomena in a relatively short time. The study of idiostyle using quantitative methods of linguistic statistics was first used in artistic texts and its syntax, compositional, graphic-phonetic and lexico-semantic compositional features were conducted. Among all methods of research of features of author's style we distinguish cognitive, linguopoietic and linguocultural approaches.

Keywords: idiostyle, idiolect, corpus text, structural model, R. Ivanychuk.

Постановка проблеми. На сьогоднішній день лінгвісти зацікавлені дослідженням індивідуального стилю, ідіостилу, ідіолекту письменників [4]. Незважаючи на те, що ідіолект як феномен досі є предметом дискусій, від абсолютного заперечення до вивчення мови як сукупності окремих мов індивідів, творчість письменників активно досліджують мовознавці [3]. Також такі поняття як ідіостиль, стиль автора, індивідуальний стиль, ідіолект, описуючи мову письменника, часто ототожнюють та не розрізняють. Посилаючись на Л. Ставицьку [4], будемо розглядати ідіостиль та стиль автора як тотожні поняття, синоніми, що означають комплекс індивідуальних особливостей стилю у творах конкретного автора як протягом всієї його творчості, та і в певний період. Відповідно ідіолект – це сукупність притаманних особливостей мовлення окремого індивіда [5].

Аналіз останніх досліджень та публікацій. Індивідуальний стиль Р. Іваничука досліджувало багато науковців, як-от: М. Беліченко, О. Гандзій, Г. Городиловської, Ю. Колядич, С. Підопригори. Творчість Р. Іваничука завжди була в полі зору науковців, тому на цю тему написано досить багато статей, публікацій та дисертацій, в яких дослідники аналізують не лише стиль письменника, а й порівнюють праці з працями інших авторів. Творчість письменника цікава тим,

що він працював протягом 58 років, навіть серед авторів усієї світової літератури знайдеться небагато з таким великим творчим шляхом. Однак серед **невирішених раніше завдань** можемо виокремити проблему про статистичні дослідження творів Романа Іваничука, які були проведені на недостатньому рівні, тому створення корпусу текстів цього автора є важливим внеском у розробку корпусу української мови, а також основою для подальших лінгвостатистичних досліджень.

Мета статті – аналіз понять ідіостилу, ідіолекту та обґрунтування вибору статистичної лінгвістики та корпусів текстів для дослідження авторського стилю.

Завдання статті – розібратися з поняттями термінів ідіостиль, стиль автора, ідіолект, визначити чому саме корпусно-статистичний підхід до вивчення авторського стилю автора є основним, який було обрано для дослідження роману Романа Іваничука.

Виклад основного матеріалу дослідження. Одними з перших, хто розпочав свої дослідження на ниві дослідження мовного стилю письменника, хто закликав до вивчення побутової мови та використання її різними людьми були молодограматисти, серед котрих В. Гумбольдт. Американський лінгвіст Б. Блох (Bernard Bloch) був першим, хто ввів поняття ідіолекту, як сукупності мовних засобів, використовуваних автором, у певний момент

мовлення для взаємодії з іншим мовцем [10]. З появою соціолінгвістики концепцію індивідуальності в мові було відкинуто на другий план, адже багато робіт присвячували саме дослідженню соціолектів. Та ще пізніше з'явилася лінгвістична експертиза, котра дала друге дихання дослідженню ідіолекту. Лінгвістична експертиза продовжила дослідження методів авторської атрибуції текстів на практиці, виокремлюючи лінгвістичні особливості письменників, подальший аналіз чого допомагає розрізнити авторів. Мовне вираження можна розглядати не лише з аналізу самого автора, а й інших чинників, до них належить тип тексту, обставини та адресат.

К. Хейзен (K. Hazen) у статті «Ідіолект» зазначає, що питома значення терміна ідіолект належить до двох категорій: 1) всі можливі висловлювання або загальна сукупність усієї мови однієї особи та 2) мовний випуск певної людини (мається на увазі лише те, що говорить людина, а не її розумові знання та думки) [12].

Отож, поняття ідіолекту відоме у лінгвістиці вже давно. Посилаючись на праці Б. Блоха, М. Халідея й Д. Аберкромбі 50–60-х років ХХ ст., М. Коулсард (M. Coulthard) виокремлює основні тези визначення поняття ідіолект: кожному носієві мови притаманний відмінний та індивідуальний варіант мови, який він використовує у мовленні та письмі і цей ідіолект проявляється через призму відмінних та особливих засобів мови, які використав автор у тексті чи мовленні. Не зважаючи на те, що кожна людина-носій мови має великий словниковий запас, який вона набула протягом років і може використовувати різноманітну кількість слів для опису однієї ситуації, насправді кожен, зазвичай, обирає однакові, типові слова. Саме ця, своєрідна мова, притаманна певній особі і є її візитною карткою, завдяки якій автора можна виокремити та ідентифікувати серед інших мовців [10].

Серед усіх методів дослідження особливостей авторського стилю О. Костецька виокремлює когнітивний, лінгвопоетичний і лінгвокультурологічний підходи [2].

Власне квантитативні методи дослідження, які вважають новітніми у аналізі авторського стилю, стають все актуальнішими з розвитком комп'ютерних технологій та доступністю відповідного програмного забезпечення. Завдяки цьому останнім часом помітно розвивається корпусна лінгвістика, адже за відносно невеликий час їй вдалося охопити різні аспекти та сторони вивчення мовних явищ [14]. З огляду на історичні обставини західним вченим даний метод був доступний раніше, проте прихильникам комп'ютерної лінгвістики довелося відстоювати та аргументувати перспективу методу, над яким вони працюють. Серед них М. Стабз (M. Stubbs), який наголошує на факті того, що методи корпусної лінгвістики дають змогу ширше охопити художній текст для знаходження певної закономірності у вживанні мови. Адже саме корпусна лінгвістика дозволяє опрацювати велику кількість різних текстів, одного жанру, автора чи періоду. До моменту розвитку корпусів не було можливості охопити весь текст, а не лише його фрагмент. М. Стабз виокремлює списки частотності слів, фразеологізмів, лем і словосполучень (N-грам) серед основних статистичних методів.

До українських дослідників квантитативної і корпусної лінгвістики належать В. Широкова,

Н. Дарчука, В. Левицького, Є Карпіловська, А. Загнітко. У своїй монографії В. Левицький зазначає, що процес запровадження квантитативних методів у вітчизняну лінгвістику пройшов нелегкий шлях. На сьогоднішній день квантитативна лінгвістика не втрачає популярності, існують численні журнали, енциклопедія, що свідчить про охоплення цим напрямом все більше ділянок вивчення мови [6].

Дослідження ідіостилю за допомогою кількісних методів лінгвістичної статистики застосовували спочатку в художніх текстах та з його допомогою проводився синтаксичних, композиційних, графіко-фонетичних та лексико-семантичних композиційних особливостей. Лінгвостатистичні методи, які застосовують сучасні лінгвісти (О. Павличко, О. Ахманова, Г. Маєр, О. Падучева та ін.), дають змогу одержати повну кількісну характеристику тексту для вивчення авторського стилю, що дозволяє пізнати внутрішню структуру твору [9]. Н. Єсіпенко, М. Мухін на основі семантичної розмітки корпусу лексики, порівняльного та статистико-стилістичного аналізу на основі вибірки розроблено модель ідіостилю автора [1]. Розроблений алгоритм аналізу ідіостилю автора можна застосовувати для подальших досліджень, а саме: для розподілу лексики за частотністю конкретного автора та подальшого порівняння з іншими – порівняльно-статистичний аналіз; а також компонентний аналіз лексики для групування за темою, порівняння результатів серед праць одного автора та інших. З допомогою аналізу можна дослідити частотні лінгвістичні ознаки стилю автора, виокремити авторську картину світу та виявити спільне та відмінне між авторами та епохами.

Існує також альтернативний метод аналізу – метод градуальних опозицій. Основним принципом цього методу є порівняння оригінального твору автора перекладам, зробленими ним і оригіналу на іноземній мові. Тож перевагою методу градуальних опозицій є застосування підходу порівняння та зіставлення.

Традиційному філологічному моделюванню ідіостилю М. Мухін протиставляє сучасніший, базований на статистичному аналізі лексем твору письменника. Дослідник розробив модель аналізу, яка складається з декількох етапів. Спочатку укладають порівняльну вибірку частотної лексики конкретного автора, здійснюють семантичну розмітку контекстів та будують концептуальний профіль автора. Це все дозволяє виявити певні ознаки авторської концептуальної системи. Кінцевим етапом є порівняльна вибірка лексичних пар або біграм конкретного автора, виявлення особливої лексичної синтагматики в його творі та побудова синтагматичного профіля автора [7].

На думку М. Мухіна, з допомогою традиційного інтуїтивного філологічного аналізу ідіостилю неможливо досягти формалізованих висновків, які можна було б оцінити кількісно. Статистична методика дає змогу доповнити, скорегувати та підтвердити читацькі та наукові бачення ідіостилю різних авторів. Також, зважаючи на об'єм матеріалу, який може бути оброблений, а саме декілька творів кожного окремого письменника, та двоаспектність підходу статистичний метод можна вважати цілісним та точним. Проте не можна не звернути уваги і на недоліки цього підходу. Статистична методика більше орієнтована

на порівняння та аналіз декількох авторів між собою, так як одним з критеріїв виокремлення індивідуальної частотної лексики є неповторюваність лексики одного автора у інших. Цей факт дещо ускладнює використання цього методу у випадку вивчення ідіостилю одного письменника.

Отже, для дослідження ідіостилю романів Романа Іваничука було обрано лексико-статистичний підхід. Перш за все оригінальний текст було перетворено в електронну форму, пізніше текст вичитувався, усувалися неточності: орфографічні помилки, зведення до єдиної системи кодування тощо. Текст, за попередньо розробленими позначками, було роззначено. Цим текстом було обрано роман

було «Вода з каменю». Текст пройшов роззначення структурного рівня: виділяємо розділи та підрозділи, абзаци та речення, мову автора та мову персонажів. Для проведення статистичних розрахунків було розроблено власні програми на мові програмування Python. Після усіх цих кроків твір готовий до проведення статистичного профілю.

Висновки та перспективи подальшого дослідження. На основі аналізу структури романів Романа Іваничука, створення та укладання нових корпусів, представлений в електронному вигляді масив мовних даних, структурований та розмічений, буде слугувати для подальших лінгвістичних досліджень ідіолекту письменника.

Список літератури:

1. Єсипенко Н.Г. Лексико-семантичні компоненти авторського стилю і мовна картина світу (на матеріалі англійської прози воєнної та мирної тематики) : автореф. дис. ... канд. філол. наук : спец. 10.02.04 «Германські мови» Нац. ун-т ім. Ю. Федьковича. Чернівці, 2007. 20 с.
2. Костецька О.П. Індивідуальне мовлення автора як об'єкт лінгвістики та підходи до його дослідження. *Наукові записки НУ «Острозька академія»*. 2014. № 49. С. 196–199.
3. Кульчицький І.М. Дані текстових корпусів у лінгвістичних дослідженнях : монографія / за ред. О.П. Левченко. Львів : Вид-во Львівської політехніки, 2015. С. 29–45.
4. Кульчицький І.М. Деякі аспекти створення корпусу художніх творів В.С. Стефаника. Київ, 2013. С. 143–148.
5. Кульчицький І.М. Корпуси текстів як лінгвотехнологічне підґрунтя виявлення змін в українській / за ред. А. Архангельська (голов. ред.), О. Левченко, О. Тищенко. Оломоуць : Ун-т ім. Ф. Палацького, 2016. С. 269–298.
6. Левицький В.В. Квантитативные методы в лингвистике. Винница, 2007. 264 с.
7. Мухин Ю.М. Лексическая статистика и идиостиль автора: корпусное идеографическое исследование (на материале произведений В. Набокова, А. Платонова и М. Шолохова) : автореф. дис. ... канд. філол. наук : спец. 10.02.19 «Теория языка». Екатеринбург, 2011. 22 с.
8. Мухин М.Ю. Лексическая статистика и идиостиль автора: корпусное идеографическое исследование (на материале произведений М. Булгакова, В. Набокова, А. Платонова, М. Шолохова) : автореф. дис. ... канд. філол. наук : 10.02.19. Екатеринбург, 2011. 43 с.
9. Павличко О.О. Щодо статистичних параметрів авторського стилю (на матеріалі творів Е.М. Ремарка). *Мовні і концептуальні картини світу : зб. наук. пр.* Київ : ВПЦ «Київський університет», 2010. Вип. 29. С. 186–191.
10. Bloch B. A set of postulates for phonemic analysis. *Language*. 1948, pp. 3–46.
11. Coulthard M. Author Identification, idiolect, and linguistic uniqueness. 2004, pp. 431–447.
12. Hazen K. Idiolect. / Keith Brown. (Editor-in-Chief). *Encyclopedia of Language & Linguistics*, Second Edition. Vol. 5. 2006, pp. 512–513.
13. Stubbs M. *Quantitative Methods in Literary Linguistics*. Cambridge, 2014, pp. 46–62.
14. Vaughan E. *Corpus Analysis*. The International Encyclopedia of Language and Social Interaction. 2015, pp. 1–17.

References:

1. Yesypenko, N.H. (2007). *Leksyko-Semantychni komponenty avtorskoho styliu i movna kartyna svitu (na materialii anhlovnoyi prozy voiennoi ta myrnoi tematyk)* [Lexical-semantic components of the author's style and the linguistic picture of the world (based on English-language prose of military and peace topics)] (PhD Thesis). Chernivtsi: National university of Yurii Fedkovych.
2. Kostetska, O.P. (2014). *Indyvidualne movlennia avtora yak ob'iekt lnhvistyky ta pidkhody do yoho doslidzhennia* [Author's Individual Speech as a Linguistic Object And Approaches To His Research]. *Scientific Notes of Ostroh Academy*, no. 49, pp. 196–199.
3. Kulchytskyi, I.M. (2015). *Dani Tekstovyykh korpusiv u lnhvistychnykh doslidzhenniakh* [Text enclosure data in linguistic studies] (eds. O.P. Levchenko). Lviv: Lviv Polytechnic, pp. 29–45.
4. Kulchytskyi, I.M. (2013). *Deiaki aspekty stvorennia korpusu khudozhnikh tvoriv V.S. Stefanyka* [Some aspects of creation of the corpus of artistic works of V.S. Stefanyk]. Kyiv, pp. 143–148.
5. Kulchytskyi, I.M. (2016). *Korpusy tekstiv yak lnhvotekhnolohichne pidgruntia vyavlennia zmin v ukrainskii* [Text corpora as a linguistic technological basis for detecting changes in Ukrainian] (eds. A. Arkhahelska, O. Levchenko, O. Tyshchenko). Olomouts: University of F. Palatskyi, pp. 269–298.
6. Levytskyi, V.V. (2007). *Kvanytatyvnye metody v lnhvystyke* [Quantitative methods in linguistics]. Vynnytsa, 264 p.
7. Mukhin, M.Y. (2011). *Leksicheskaya statistika i idiostil' avtora: korpusnoye ideograficheskoye issledovaniye (na materiale proizvedeniy V. Nabokova, A. Platonova i M. Sholokhova)* [The author's lexical statistics and idyll: a corpus ideographic study (based on works by V. Nabokov, A. Platonov, and M. Sholokhov)]. (PhD Thesis), Ekaterinburg.
8. Mukhin, M.Y. (2011). *Leksicheskaya statistika i idiostil' avtora: korpusnoye ideograficheskoye issledovaniye (na materiale proizvedeniy M. Bulgakova, V. Nabokova, A. Platonova, M. Sholokhova)* [The lexical statistics and idiom of the author: corpus ideographic studies (based on the works of M. Bulgakov, V. Nabokov, A. Platonov, M. Sholokhov)]. (PhD Thesis), Ekaterinburg.
9. Pavlychko, O.O. (2010). *Shchodo statystychnykh parametriv avtorskoho styliu (na materialii tvoriv E.M. Remarka). Movni i kontseptualni kartyny svitu* [Regarding the statistical parameters of the author's style (on the material of works by E.M. Remark). *Language and Conceptual Pictures of the World*]. Kyiv: Kiev University, no. 29, pp. 186–191.
10. Bloch, B. (1948). A set of postulates for phonemic analysis. *Language*, pp. 3–46.
11. Coulthard, M. (2004). Author Identification, idiolect, and linguistic uniqueness, pp. 431–447.
12. Hazen, K. (2006). Idiolect. (eds. Keith Brown). *Encyclopedia of Language and Linguistics*, Second Edition, vol. 5, pp. 512–513.
13. Stubbs, M. (2014). *Quantitative Methods in Literary Linguistics*. Cambridge, pp. 46–62.
14. Vaughan, E. (2015). *Corpus Analysis*. The International Encyclopedia of Language and Social Interaction, pp. 1–17.