

РОЗДІЛ IV. ПРИКЛАДНА ЛІНГВІСТИКА: НАПРЯМИ Й АСПЕКТИ ДОСЛІДЖЕННЯ

Соломія Бук

ORCID: orcid.org/0000-0001-8026-3289

УДК 81'(32+33+373+374)=161.2Франко

DOI: 10.31558/1815-3070.2022.43.11

БАГАТСТВО МОВИ ІВАНА ФРАНКА: ПРИКЛАД ВЕЛИКОЇ ПРОЗИ

Статтю присвячено поняттю багатства мови (мовлення, тексту) – це відношення між кількістю різних слів у тексті до довжини тексту (кількості слововживань). Багатство словника вперше обчислено на прикладі електронного корпусу текстів великої прози Івана Франка. Його зіставлено з аналогічним показником частотного словника сучасної української художньої прози (укладеного на матеріалі 24 творів 22 різних письменників) і виявлено передбачувано нижчий показник багатства словника у корпусі великої прози І. Франка. Аналіз показав, що індекс різноманітності залежить не тільки від обсягу тексту, а й від інших параметрів, як-от редакція твору, пряме чи авторське мовлення.

Ключові слова: багатство словника (індекс різноманітності), частотний словник, лінгвостатистична параметризація, обсяг словника і тексту, корпус текстів, лінгвостатистика, прикладна лінгвістика, українська мова.

Постановка наукової проблеми та її актуальність. Контекст сучасного наукового плюралізму та його множинні дослідницькі підходи актуалізують статус статистичної лінгвістики як міждисциплінарної науки на перетині математики (статистики) і мовознавства, яка займається виявленням та описом кількісних характеристик та закономірностей мови та мовлення, а також особливостей авторського ідіолекту. І. Франко належить до ключових постатей національної культури, і багатовимірне вивчення його творчості – пріоритетне завдання франкознавства, а інвентаризація мовлення письменника на лексичному рівні – важливе і потрібне завдання сучасної гуманітаристики. Постановка проблеми у цьому ракурсі дає змогу виявити на новому матеріалі досі не досліджені особливості його ідіолекту. Багата мова письменника віддавна привертала увагу дослідників зі сфери теорії й історії літератури, психології та когнітивістики, семантики, семіотики, стилістики тощо. У процесі вивчення мови письменника науковцям, звичайно, складно точно визначити, у чому полягає багатство лексики певного автора, хоча інтуїтивно цю якість відчують. Багатий ідіолект, тобто різноманітний, розвинений лексика особистості, дає змогу точно й тонко передати і свій власний внутрішній, і сприйняття навколишнього світу, глибоко в нього проникнути. Лінгвостатистика дає новий погляд на цей феномен, пропонує оригінальне його трактування за посередництвом конкретного обчислення та кількісного аналізу. У термінології цього напрямку лінгвістики багатство словника (БС), багатство мови, лексичне багатство, індекс різноманітності – це індекс, що базується на відношенні між кількістю різних слів у тексті (лем, лексем), тобто словником тексту (V), і довжиною тексту, виміряною у слововживаннях (N), а саме: чим більше різних слів у тексті, тим багатша його лексика.

Аналіз досліджень проблеми. В англomовній літературі поняття багатства словника називають TTR (Type – лексична одиниця, Token – її вживання у тексті, Ratio – відношення), тобто те саме відношення кількості слів до слововживань (обсягу тексту). Поняття БС має аксіологічне забарвлення (пор. багатий versus бідний словник, багата versus бідна лексика). А. Павловський індекс лексичного багатства вказує серед необхідних характеристик авторської атрибуції тексту, називаючи його інваріантом стилю письменника (Pawłowski 2003: 174–177). Загалом польське мовознавство вирізняється суттєвим досвідом у визначенні багатства словника своїх письменників (праці М. Бальовського, Є. Ворончака, В. Курашкевича, Я. Лукасевича, Е. Рудницької-Фіри, Е. Стахурського, Я. Самбор та ін.).

Цей показник міцно ввійшов у світову та українську статистичну лінгвістику попри існування протилежних думок (наприклад, Ю. Шевельов, оцінюючи віртуозність Т. Шевченка, критично відгукнувся про кількісне обчислення БС письменників загалом: «не можна підрахунками кількості слів стверджувати, або доводити багатство чи бідність мови того чи того письменника. Всі такі підрахунки поперше завжди надто неточні, а подруге взагалі нічого не характеризують, просто непотрібні» (Шевельов 1944 ч. 7: 6). Так, його обчислено для творів багатьох українських та закордонних письменників: А. Міцкевича, Б. Пруса (Kuraszkiewicz 1973), К. Чапека (SKC 2007), С. Яблонської (Занько 2015), М. Яцкова (Цьох 2021) та інших. Зауважено різні закономірності стосовно БС. Наприклад, група українських дослідників частотних словників дійшла висновків, що багатство словника обернено пропорційне довжині тексту, тобто, чим довший текст, тим потенційно менше з'являється у ньому нових слів (Перебийніс та ін. 1985: 143). В. Курашкевич визначив цінні квантитативні залежності: 1) віршовані твори мають звичайно багатшу лексику, ніж прозові твори того самого автора; 2) багатший словник мають оригінальні тексти порівняно з перекладними; 3) твори, складені з менших частин різної тематики, мають більше різних слів, ніж суцільні тексти тої самої довжини; 4) багатший словник мають тексти реалістичної тематики, ніж ліричні, релігійні чи з абстрактною лексикою; 5) багатший словник мають високолітературні твори (Kuraszkiewicz 1973: 48).

Наукова новизна статті полягає у тому, що БС творів І. Франка досі не було об'єктом окремого дослідження, і його зіставлення з частотним словником сучасної української художньої прози (ЧС СУХП) виконано вперше. Великий інтерес також становить потенційне виявлення нових параметрів, з якими може корелювати ця характеристика, як-от різні редакції одного твору або дихотомія авторського наративу і мовлення персонажів. Отже, **мета статті** – виявити особливості проявлення БС у різних творах великої прози І. Франка (ВПФ), а також у ВПФ загалом; **завдання** – обчислити БС усіх дев'яти українськомовних творів ВПФ та зіставити їх між собою, а також з БС сучасної української художньої прози (ЧС СУХП); диференціювати статистичну інформацію БС щодо прямого й авторського мовлення ВПФ; порівняти БС різних редакцій одного твору («Воа constrictor» 1884 та 1905–1907 pp.). **Об'єкт дослідження** – тексти великої прози І. Франка, **предмет дослідження** – виявлення особливостей проявлення БС у ВПФ та параметрів тексту, від яких воно може залежати.

Фактичний матеріал дослідження становить електронний корпус текстів І. Франка, що складається з останніх прижиттєвих редакцій усіх дев'яти українськомовних творів його великої прози, оскільки їх вважаємо останньою волею автора: «Борислав сміється» (1880–1881), «Захар Беркут» (1883), «Не спитавши броду» (1885–1886), «Для домашнього огнища» (1892), «Основи суспільності» (1894–1895), «Перехресні стежки» (1900), «Воа constrictor» (1905–1907), «Великий шум» (1907), «Петрії й Довбушуки» (1909–1912). Завдяки морфологічному маркуванню корпусу для кожного з текстів укладено частотний словник (ЧС), з якого отримано найважливіші лінгвостатистичні характеристики для обчислення БС: обсяг тексту і обсяг словника (реєстру, тобто кількість лематизованих слів). Працю виконано у межах повної лінгвостатистичної параметризації творчості І. Франка, її мета – виявити й проаналізувати кількісно визначальні характеристики цієї групи текстів письменника, які виформовують своєрідну лінгвостатистичну модель ідіолекту автора (Бук 2021: 50–58), БС відіграє у ній незамінну роль.

Виклад основного матеріалу й обґрунтування результатів дослідження. Багатство мови творів великої прози І. Франка містить *табл. 1*.

Таблиця 1 – Багатство мови творів великої прози І. Франка

	Романи І. Франка	Багатство словника	Обсяг тексту, слововживання
1	Борислав сміється	0,11	77 455
2	Захар Беркут	0,13	50 223
3	Не спитавши броду	0,15	49 170
4	Для домашнього огнища	0,15	44 841
5	Основи суспільності	0,13	67 172
6	Перехресні стежки	0,11	93 890
7	Воа constrictor, 1905–1907	0,181	34 215
8	Великий шум	0,175	37 005
9	Петрії й Довбушуки, 1909–1912	0,14	52 751
	Загалом	0,052 1,26/9=0,14	506 722

БС ВПФ – це не сума БС усіх творів, розділена на 9, оскільки у них є велика кількість спільної лексики. Також з тої самої причини не можна обчислювати БС ВПФ як просте відношення обсягу словника ВПФ 26 888 лем до обсягу текстів ВПФ 506 722 слововживань (0,13). В окремій таблиці зводимо словники усіх дев'яти романів разом (виходить 26 369 різних лем) і ділимо на обсяг слововживань (506 722) = **0,052**.

Як показує *табл. 1*, БС «Воа constrictor» = 0,181, воно найбільше серед усіх інших романів та повістей І. Франка. Це найкоротший за обсягом твір великої прози (34 215 слововживань). Далі зі збільшенням обсягу творів зменшується БС: «Великий...» (37 005) – 0,175, «Для Домашнього...» (44 841) і «Не спитавши...» (49 170) – 0,15, «Петрії...» (52 751) – 0,14, «Основи...» (67 172) і «Захар...» (50 223) – 0,13, і, нарешті, найдовші твори «Борислав...» (77 455) і «Перехресні...» (93 890) мають очікувано найменше БС = 0,11. Отже, чим коротший твір, тим різноманітніша у ньому лексика. Так матеріал ВПФ підтверджує згадані вище результати аналізу інших текстів (Перебийніс та ін. 1985: 143). Привертає увагу факт, що за однакового БС (0,11), твори «Борислав...» і «Пере-

хресні...» суттєво відрізняються за обсягом (77 455 проти 93 890 відповідно), тобто на 16 435 слововживань, а це становить близько 20 % тексту. Отже, можна узагальнити, що лексика «Перехресних...» значно багатша за лексику «Борислава...».

Описуючи, що показник багатства лексики зменшується зі збільшенням обсягу тексту, укладачі частотного словника сучасної української художньої прози припускають, що «різниця у збільшенні кількості слів і словоформ просто є свідченням того, що пряма мова менш різноманітна лексично, ніж авторська» (ЧС СУХП 1: 15). Дані з *табл. 2* засвідчують таку саму тенденцію. Це можна зауважити і на творах з майже однаковим співвідношенням прямого й авторського мовлення («Перехресні...», «Великий шум»), і на різних творах. Зокрема, за приблизно однакового обсягу прямого мовлення в «Бориславі...» (33 330) та авторського в «Захарі...» (32 994), індекс різноманітності вищий в авторському: 0,12 та 0,17, відповідно. Ще більше відрізняється багатство словника на користь наративного тексту у «Великому шумі» (0,25) порівняно з прямим мовленням «Для домашнього...» (0,17), водночас їх обсяг майже однаковий: 18 822 та 18 914 відповідно.

Таблиця 2 – Багатство словника у прямому й авторському мовленні творів великої прози І. Франка

Твори	Пряме мовлення		Авторське мовлення		Разом	
	БС	слововж.	БС	слововж.	БС	слововж.
«Борислав сміється»	0,12	33 330	0,16	44 125	0,11	77 455
«Захар Беркут»	0,15	17 229	0,17	32 994	0,13	50 223
«Не спитавши броду»	0,16	21 590	0,21	27 580	0,15	49 170
«Для домашнього огнища»	0,17	18 914	0,20	25 927	0,15	44 841
«Основи суспільності»	0,14	28 278	0,17	38 894	0,13	67 172
«Перехресні стежки»	0,13	45 513	0,15	48 377	0,07	93 890
«Воа constrictor», 1907	0,23	7 435	0,21	26 780	0,18	34 215
«Великий шум»	0,18	18 183	0,25	18 822	0,18	37 005
«Петрії й Довбушуки»	0,18	16 219	0,17	36 532	0,14	52 751
ВПФ загалом	0,067	206 691	0,072	300 031	0,052	506 722

БС для різних редакцій одного твору виміряно на прикладі повісті І. Франка «Воа constrictor». Вона належить до творів бориславського циклу. Вперше її опубліковано 1878 р. у журналі «Громадський друг» (№ 2), який видавали М. Павлик та І. Франко (тому він був конфіскований цензурою), а також у збірниках «Дзвін» та «Молот», що були продовженням цього журналу. Пізніше, 1884 р., твір вийшов окремим виданням, накладом редакції журналу «Зоря», з яким тоді співпрацював автор (Франко 1984). 1905–1907 р. Франко повернувся до роботи над повістю, внаслідок чого з'явилася нова редакція твору, яка має істотні відмінності від попередніх публікацій (Франко 1907).

Є усі підстави для коректного порівняння БС обох редакцій, оскільки для тексту кожної з них укладено окремий ЧС за однаковими принципами. ЧС першої редакції повісті видано окремою книгою (Бук 2013).

Таблиця 3 – Багатство словника повсіті І. Франка «*Boa constrictor*» 1884 та 1905–1907 рр.

	Твори І. Франка	Багатство словника	Обсяг тексту, слововживання
1	<i>Boa constrictor</i> , 1884	0,197	25 427
2	<i>Boa constrictor</i> , 1905–1907	0,182	34 214

Отже, БС першої редакції твору становить 0,197, а другої – 0,182. Як показують дані *табл. 3*, редакції суттєво відрізняються за обсягом тексту: друга редакція (34 214) майже на третину більша від першої (25 427 слововживань), проте виявлено, що БС відрізняється несуттєво. Можна стверджувати, що автор, суттєво переробивши твір, особливу увагу звертав на різноманітність лексики, у підсумку її збагативши. Про це ж свідчить детальне текстологічне порівняння цих видань, наведене у коментарях до 50-томного зібрання творів: «І. Франко провів велику роботу по редагуванню журнального тексту: крім стилістичних виправлень та вилучень натуралістичних описів лютування холери в Галичині, а також довгих діалогів, він дописав закінчення, яке має принципове значення для розуміння образу Германа Гольдкремера» (Франко 1976–1986 14: 453).

Методику кількісного зіставлення лексики різних жанрів та творів між собою, а також із порівняльним корпусом сучасної прози послідовно використано у статистичних дослідженнях творчості багатьох письменників, зокрема А. Міцкевича, Ю. Словацького, К. Чапека, французьких символістів, польських романтиків, натуралістів та ін. Так, французький дослідник П. Гіро порівняв лексику символістів П. Корнеля, Ж. Расіна, Ш. Бодлера, С. Малларме, Л. Валері й Г. Аполлінера із загальним ЧС французької мови (Guiraud 1954: 78, 83, 100, 102, 104). Е. Стахурський порівняв лексику поеми А. Міцкевича «Пан Тадеуш» з іншими романтичними поемами (Stachurski 2005), Т. Скубалянка зробила подібне дослідження зі Ю. Словацьким на тлі традиції (Skubalanka 1966), Ф. Чермак порівняв ЧС К. Чапека із корпусом сучасних чеських текстів (SKC 2007).

БС ВПФ у зіставленні з ЧС сучасної української художньої прози продемонстровано у *табл. 4*. Верифікативність зіставлення БС ВПФ з частотним словником української художньої прози загалом ґрунтується на фактично однаковому їх обсязі та спільних принципах і методах їх побудови.

Таблиця 4 – Порівняння багатства словника ВПФ зі сучасною українською великою прозою (ЧС СУХП)

Квантитативні параметри	Корпус текстів великої прози І. Франка	Корпус текстів Частотного словника сучасної української художньої прози
<i>N</i> , обсяг тексту, слововживання	506 722	500 000
<i>V</i> , обсяг словника, лексеми	26 368	33 391
<i>V/N</i> , багатство словника	0,052	0,067

Обсяг словника сучасної української художньої прози становить 33 391 лексему – це на 7 023 лексем (21 %) більше, ніж обсяг лексики для І. Франка (26 368). Природно, що й індекс різноманітності (БС) української прози (0,067) вищий, ніж в І. Франка (0,054). Такий результат можна вважати очікуваним, оскільки

ки корпус текстів сучасної української художньої прози становлять 24 твори 22 різних письменників.

Висновки та перспективи дослідження. Лінгвостатистичний параметр БС для творів І. Франка обчислено вперше. У ширшому контексті сучасних методів дослідження мови письменника, зокрема у славістичних студіях, – це важлива лінгвостатистична характеристика ідіостилю. БС є незамінним у створенні квантитативного портрету ВПФ та ідіолекту письменника загалом.

Зіставлення БС ВПФ з українською великою прозою загалом показало об'єктивні кількісні співвідношення між ними. На матеріалі творів письменника вслід за художньою прозою ХХ ст. також зафіксовано дві тенденції: 1) зі збільшенням обсягу тексту зменшується його лексичне багатство та 2) для текстів з однаковим обсягом індекс різноманітності вищий для авторської розповіді порівняно з мовленнєвою взаємодією персонажів. Виявлено, що на БС впливають не тільки обсяг твору та різних лексем у ньому, а й різні редакції твору.

Перспективи дослідження багатоаспектні. Цікаво було б зіставити БС великої прози автора із аналогічною характеристикою його малої прози, поезії, публіцистики, наукових праць, перекладів (особливо автоперекладів) тощо, а також з іншими українськими чи іншими слов'янськими авторами великого жанру. Нові горизонти окресило би зіставлення БС ВПФ із референтним корпусом західноукраїнської великої прози його сучасників. Водночас такий напрям поставить питання, хто із українських письменників Західної України, які жили в його час, зміг би змагатися з ним за обсягом написаного.

Література

1. Бук С. Велика проза Івана Франка: електронний корпус, частотні словники та інші міждисциплінарні контексти: монографія. Львів: ЛНУ імені Івана Франка, 2021. 421 с.
[Buk S. Velyka proza Ivana Franka: elektronnyi korpus, chastotni slovnyky ta inshi mizhdystsyplinarni konteksty: monohrafiia. Lviv: LNU imeni Ivana Franka, 2021. 421 p.]
2. Бук С. Частотний словник повісті Івана Франка «Boa constrictor» (редакція 1884 р.). *Стежками Франкового тексту (комунікативні, лінгвосеміотичні, когнітивні та лінгвостатистичні виміри прози)*. Львів: ЛНУ імені Івана Франка, 2013. С. 202–501.
[Buk S. Chastotnyi slovnyk povisti Ivana Franka «Boa constrictor» (redaktsiia 1884 r.). *Stezhkami Frankovoho tekstu (komunikatyvni, lnhvosemiotychni, kohnityvni ta lnhvostatystychni vymiry prozy)*. Lviv: LNU imeni Ivana Franka, 2013. S. 202–501]
3. Занько Д. О. Статистична структура твору Софії Яблонської «Чар Марока». *Актуальні проблеми філології*. Одеса. 2015. С. 18–20.
[Zanko D. O. Statystychna struktura tvorv Sofii Yablonskoi «Char Maroka». *Aktualni problemy filolohii*. Odesa. 2015. S. 18–20]
4. Перебийніс В. С., Муравицька М. П., Дарчук Н. П. Частотні словники та їх використання. Київ: Наукова думка, 1985. 204 с.
[Perebyinis V. S., Muravytska M. P., Darchuk N. P. Chastotni slovnyky ta yikh vykorystannia. Kyiv: Naukova dumka, 1985. 204 s.]
5. Цьох Л. Квантитативний профіль ідіостилю М. Яцкова (на матеріалі повісті «В лабетах» та новелістики). V Міжнародна науково-практична конференція «Modern directions of scientific research development» (24–26 жовтня 2021 р.). Чикаго, США. С. 757–765.
[Tsokh L. Kvantytatyvnyi profil idiostylu M. Yatskova (na materialy povisti «V labetakh» ta novelistyky)]. V *International Scientific and Practical Conference «Modern directions of scientific research development» (October 28–30, 2021)*. Chicago, USA. P. 757–765.

6. ЧС СУХП – Частотний словник сучасної української художньої прози / за ред. В. С. Переймис. Київ: Наукова думка, 1981. Т. 1, 2.
[ChS SUKhP – Chastotnyi slovnyk suchasnoi ukrainskoi khudozhnoi prozy / za red. V. S. Perebyinis. Kyiv: Naukova dumka, 1981. T. 1, 2]
7. Шевельов Ю. У кого з українських письменників багатша мова. *Земля (Плауен)*, ч. 6 (22.10.1944, 6), ч. 7 (29.10.1944, 5–6). URL.: <https://zbruc.eu/node/93438> (20.04.22).
[Shevelov Yu. U koho z ukrainskykh pysmennykiv bahatsha mova. *Zemlia (Plauen)*, ch. 6 (22.10.1944, 6), ch. 7 (29.10.1944, 5–6). URL.: <https://zbruc.eu/node/93438> (20.04.22)]
8. Guiraud P. Les caractères statistiques du vocabulaire. Essai de méthodologie. Paris: Presses universitaires de France, 1954. 111 p.
9. Kuraszkiewicz W. Obfitość słownictwa w kilku dużych tekstach polskich («Postylla», «Wizerunek», «Worek Judaszów», «Pan Tadeusz», «Beniowski», «Lalka», «Popioły», «Kwiaty polskie»). *Studia Polonistyczne (Poznań)* IV, 1973. S. 45–63.
10. Pawłowski A. O problemie atrybucji tekstu w ligwistyce kwantytatywnej (na przykładzie tekstu Romaina Gary). J. Linde-Usiekniewicz, R. Huszcza (red.). *Prace językoznawcze dedykowane Profesor Jadwidze Sambor*. Warszawa: nakładem Wydziału Polonistyki Un. Warszawskiego, 2003. S. 169–190.
11. Skubalanka T. Słownictwo poezji miłosnej Juliusza Słowackiego na tle tradycji. Toruń, 1966. 414 s.
12. Stachurski E. Słownictwo «Pana Tadeusza» Adama Mickiewicza na tle tekstów innych poematów romantycznych. Kraków: Wydawnictwo Naukowe Akademii Pedagogicznej, 2005. 262 s.
13. SKC – Slovník Karla Čapka / Hl. ed. Fr. Čermák. Praha: Nakladatelství i Lidové noviny; Ústav Českého národního korpusu, 2007. 715 s.

Використані джерела

1. Франко І. Воа constrictor: повѣсть Ивана Франка / накладомъ редакції «Зорѣ». Львѣвъ: Зъ печатнѣ Ставропигійского Института, 1884. 114 с.
[Franko Y. Voа constrictor: povѣst Yvana Franka / nakladom redaktsii «Zore». Lvov: Z pechatnie Stavropyhiiskoho Ynstitutu, 1884. 114 s.]
2. Франко І. Воа constrictor: нове, перероблене виданє. Львів: Накладом Українсько-руської видавничої спілки, 1907. 150 с.
[Franko I. Voа constrictor: nove, pereroblene vydanie. Lviv: Nakladom Ukrainsko-ruskoі vydavnychoі spilky, 1907. 150 s.]
3. Франко І. Я. Зібрання творів: у 50 т. / АН УРСР, Ін-т л-ри ім. Т. Г. Шевченка. Київ: Наук. думка, 1976–1986. Т. 1–50.
[Franko I. Ya. Zibrannia tvoriv: u 50 t. / AN URSSR, In-t l-ry im. T. H. Shevchenka. Kyiv: Nauk. dumka, 1976–1986. T. 1–50]

Список скорочень

БС – багатство словника.
ВПФ – велика проза І. Франка.
ЧС – частотний словник.
СУХП – сучасна українська художня проза.
TTR – Type Token Ratio.

THE RICHNESS OF IVAN FRANKO'S LANGUAGE: A CASE OF LONG PROSE FICTION

Buk Solomija

Department for General Linguistics Ivan Franko National University of Lviv.

Abstract

Background: Research is a part of the wider project of statistical parametrization of Ivan Franko's long prose fiction (FLPF). Vocabulary richness is an important element of it. It is calculated as the TTR (ratio of lexical items to the total number of words). It has some tradition in both Ukrainian and world linguistics, being measured for a lot of the writers and texts.

Purpose: 1) to identify the original features of vocabulary richness for FLPF, 2) to compare vocabulary richness of FLPF with the frequency dictionary of modern Ukrainian long prose fiction; 3) potentially to find out the other parameters of the text which could influence on the vocabulary richness.

Results: The frequency dictionary were compiled for all the Ukrainian novels by I. Franko: *Boryslav Laughs; Zakhar Berkut; Without Asking a Wage; For the Hearth; Pillars of Society; The Cross-paths; Boa constrictor, 1907; The Great Noise; Petrii y Dovbushchuky*. The highest index of variety was found for «Boa constrictor» (0,181), the lowest was found for «The Cross-Paths» (0,106), which confirms the inverse dependence of the vocabulary richness on the text size. It is expected that there are fewer different tokens in the FLPF dictionary then in general Ukrainian long prose fiction (0,052: 0,067 respectively). Also, with the increase of the text size, its lexical richness decreases; for texts with the same volume the index of diversity is higher for the author's speech compared to the direct speech. Different versions of the same text have different vocabulary richness (Boa constrictor 1884 and 1905–1907).

Discussion: Big potential has comparison the vocabulary richness of FLPF with similar characteristics of his other works of different genres and with other authors.

Keywords: vocabulary richness (Type Token Ratio (TTR), index of diversity), frequency dictionary, linguistic statistical parameterization of text, vocabulary size, text size, text corpus, linguistic statistics, applied linguistics, Ukrainian language, Ivan Franko.

Vitae

Solomija Buk is Doctor of Sciences in Philology (Doctor Habilitated), Associate Professor (Docent), Associate Professor of Department for General Linguistics, Ivan Franko National University of Lviv. Research interests include corpus, statistical, computational linguistics, linguistic semantics and pragmatics, lexicography, philosophy of science, psychology of happiness.

Correspondence: solomija@gmail.com

Надійшла до редакції 07 квітня 2022 року
Рекомендована до друку 22 квітня 2022 року