

## МОВОЗНАВСТВО

УДК 811.161.2'1'324'38 (038)

### ЛІНГВОСТАТИСТИЧНИЙ ОПИС “НЕ СПИТАВШИ БРОДУ” ІВАНА ФРАНКА

Соломія БУК

*Львівський національний університет імені Івана Франка,  
кафедра загального мовознавства,  
вул. Університетська, 1, к. 343, Львів, Україна, 79000  
тел.: (032) 239 47 56, e-mail: solomija@gmail.com*

Проаналізовано роман І. Франка “Не спитавши броду” з позицій лінгвостатистики, який зроблено в мовознавстві вперше. На матеріалі електронного корпусу тексту, що має зовнішнє та внутрішнє маркування, укладено частотний словник твору, з якого отримано такі його параметри, як обсяг тексту і словника, індекси різноманітності та концентрації, співвідношення прямої та авторської мови, співвідношення частин мови у словнику й тексті, індекси епітетизації, дієслівних означень, ступінь номінальності.

*Ключові слова:* лінгвостатистика, корпус текстів, маркування, статистичні параметри стилю, квантитативні характеристики тексту, слововживання, лема, багатство словника.

Лінгвостатистичні дослідження – новітній підхід до вивчення текстів, проте має свою традицію як в українській, так і в зарубіжній науці, і відлік може починати від античного світу. Суттєвий внесок у розвиток цієї галузі мовознавства зробили Габріель Альтман (Gabriel Altmann), Рейнгард Кьолер (Reinhard Köhler) (Німеччина), Петер Гжибек (Peter Grzybek), Еммеріх Келіх (Emmerich Kelih) (Австрія), Гейза Віммер (Geiza Wimmer) (Словаччина), Адам Павловскі (Adam Pawłowski), Ядвіга Самбор (Jadwiga Sambor) (Польща), Валентина Перебийніс, Наталія Дарчук, Віктор Левицький (Україна), Юхан Тулдава (Естонія), Раймунд Пиотровський, Анатолій Шайкевич (Росія) та багато інших.

Лінгвостатистично вже описано творчість таких письменників, як В. Шекспір [28], Дж. Джойс [31], К. Чапек [36], О. Бжезіна [32], Б. Грабал [29], К. Бачинський [27], М. Павич [38], М. Данойлич [37], Н. Вапцаров [16], Ф. Достоевський [26], А. Чехов [15], А. Вознесенський [1], сучасна українська поезія [24] та художня проза середини ХХ ст. [18; 25] та багатьох інших.

У Львівському національному університеті імені Івана Франка триває проект створення корпусу текстів І. Франка, на основі якого заплановано зробити комплексну

квантитативну параметризацію творів письменника [4; 5]. Під цим радикалом вже описано такі твори, як “Борислав сміється” [7], “Захар Беркут” [2], “Для домашнього огнища” [6], “Основи суспільності” [8], “Великий шум” [3], а також “Перехресні стежки” [10–12]. Статистичні параметри лексики роману “Не спитавши броду” у цій статті вперше стануть предметом окремого дослідження.

У статті подано історію реконструкції твору, коротко описано укладення електронного корпусу тексту і його маркування, зокрема прямої й авторської мови та морфологічних характеристик слів, викладено статистичні характеристики твору з увагою до кількісних реляцій між частинами мови (індекси епітетизації та дієслівних означень, ступінь номінальності). Зразок Частотного словника роману І. Франка “Не спитавши броду” запропоновано в Додатку.

**Історія реконструкції тексту “Не спитавши броду”.** “Не спитавши броду” належить до незакінчених творів і має цікаву історію створення та життя. Над твором І. Франко працював упродовж 1880-х років, планував надрукувати його спочатку в журналі “Поступ” (листи до М. Драгоманова від 31 жовтня, 20 листопада 1886 року), а потім, після поліційної конфіскації навіть проспекту журналу, – в альманасі, який також видати не вдалося. Тому роман “Не спитавши броду” залишився неопублікованим, тільки окремі сім його уривків були надруковані в різних виданнях й зажили самостійним життям як окремі оповідання: “На лоні природи”, “Тава і Вовкун”, “Борис Граб”, “Геній”, “Гершко Гольдмахер”, “Тава”, “Дріада”. Оповідання “Тава” І. Франко видав навіть окремою книжкою [19] та переклав польською мовою [30].

Уперше як цілий твір роман, який реконструював М. Возняк у супроводі його статті “Спроба відбудови невикінченої повісті” [14], вийшов у журналі “Червоний шлях” [23]. Окремою книгою твір виданий 1966 року у реконструкції Г. Вервеса [22], чому передувала стаття франкознавця “Незакінчена повість Івана Франка “Не спитавши броду” (До проблеми українсько-польських громадських зв’язків)” [13].

**Корпус тексту “Не спитавши броду” та його маркування.** В основу електронного корпусу твору “Не спитавши броду” ліг варіант, який опублікував М. Возняк [23], передрукований у 50-томному виданні творів письменника [21] з урахуванням пропущених уривків із оповідання “На лоні природи” [20].

Зовнішнє маркування корпусу подає зовнішню стосовно нього інформацію: бібліографічний опис, освіта й походження автора, час і тривалість написання твору тощо. Структурна анотація інформує про частини тексту (I–IX), віршовані вставки, поклики (із додатковою інформацією, чи вони авторські, чи зроблені редакцією), пряму та авторську мову і т. д. [9]. У романі зафіксовано два уривки “руської народної пісні”, що починається словами “Де ж ти, милий, пробуваєш...”. Не зафіксовано жодного авторського пояснення чи поклику. Цікаво, що І. Франко не вважав за потрібне перекладати єврейські, німецькі чи польські діалоги своїх персонажів, що може слугувати припущенням, що ці мови були зрозумілими сучасним реципієнтам письменника. Усі поклики та пояснення у тексті є редакційними, наприклад: “Вечори під липою” (польс. – *Ред.*), і не входять в загальний обсяг тексту, оскільки їх мова не належить І. Франкові.

*Розмежування прямої та авторської мови у творі також становить окремий нау-*

ковий інтерес, оскільки художня проза – не гомогенний жанр, як часто трактують його дослідники, а інтегральне поєднання розмовно-побутового та наративного. Їх співвідношення для аналізованого твору подано в таблиці 1.

Таблиця 1

Співвідношення прямої та авторської мови  
в романі І. Франка “Не спитавши броду”

пряма мова		авторська мова		разом	
слововживань	%	слововживань	%	слововживань	%
21 590	43,9	27 580	56,1	49 170	100

Із виявлених даних видно, що пряма мова відіграє важливу роль, оскільки займає значну частину твору (майже 44 %). Ці показники цікаво зіставити з іншими творами великої форми І. Франка. Подібне співвідношення прямої та авторської мови мають й романи письменника про сучасне йому життя: “Борислав сміється”, “Для домашнього огнища”, “Основи суспільності”, “Перехресні стежки”, “Великий шум” (42,1–49,1 %). Найменшу кількість прямої мови має “*Voа constrictor*” (12,4 %), що зумовлено особливостями нарації твору. “Захар Беркут” і “Петрії й Довбуцуки” – історичні твори, які потребують більше авторського опису, відповідно діалогів у них менше: 34,3 % і 30,7 %. Наочно це зображено на рисунку 1.

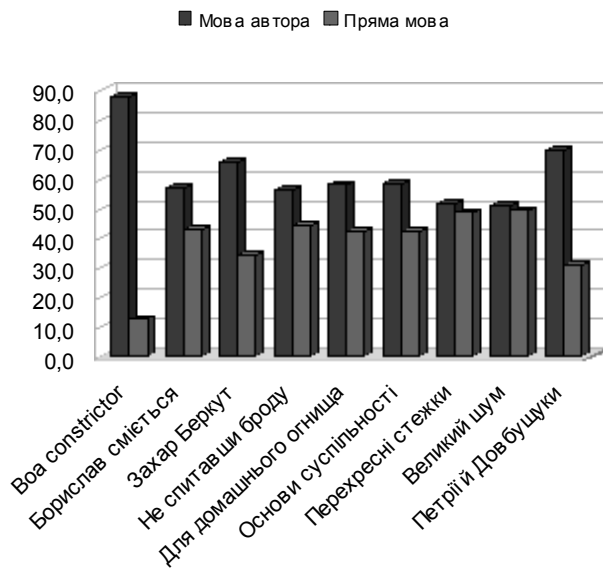


Рис. 1. Співвідношення прямої та авторської мови в романах І. Франка

Загалом, середня кількість прямої мови у всіх великих прозових творах І. Франка становить 40,6 %, а це свідчить про високий рівень їх діалогічності.

До внутрішньої анотації входить, зокрема, **морфологічна**, яка дає змогу автоматично отримати багату інформацію про словоформи, лексеми (леми), частини мови тощо.

Кожна словоформа у процесі лематизації отримала свою початкову форму, з чого в результаті комп'ютерна програма згенерувала два списки: частотний список словоформ та частотний список лем (слів у початковій формі). Зразок частотного словника твору подано в кінці статті. У процесі підготовчої роботи над текстом зафіксовано та виправлено друкарську неточність: слово *доклано* виправлено на *докладно*. Для словоформи *ригорозум* запропоновано лему “*ригорозум* або *ригороз*”, оскільки з контексту (“...то був найстарший син Трацьких, укінчений правник і трохи вже чи не доктор прав (мав ще одно **ригорозум** робити), Густав Трацький”) складно однозначно виявити форму однини. У корпусі розрізнено омонімію. Шляхом контекстуального аналізу виявлено такі омоніми: *а* (вигук, спол.), *або* (спол., частка), *батьків* (прикм., ім.), *біг* (ім., дієсл.), *братів* (прикм., ім.), *будуце* (ім., прикм.), *все* (присл., займ.), *гори* (ім., присл.), *граб/Граб* (загал. і власна назва), *де* (присл., частка), *діти* (ім., дієсл.), *доктор* (лікар і наук. ступінь), *долі* (ім., присл.), *дорога* (ім., прикм.), *жаль* (ім., служб.), *жидків* (ім., прикм.), *захід* (дія, сторона світу), *зимою* (ім., присл.), *її, його, їх* (особовий та присв. займ.), *касієрова* (ім., прикм.), *коло* (ім., прийм.), *корч* (рослина, спазм), *коса* (волосся, знаряддя), *крихітку* (ім., присл.), *круг* (ім., присл.), *лютий* (ім., прикм.), *мило* (ім., присл.), *минуше* (ім., прикм.), *ніж* (ім., частка), *о* (вигук, прийм.), *образ* (вигляд, почуття), *палати* (ім., дієсл.), *паничів* (ім., прикм.), *перед* (ім., прийм.), *передом* (ім., присл.), *поверх* (ім., присл.), *повій* (ім. (рослина), дієсл.), *поза* (ім., прийм.), *попасти* (куди і про худобу), *професорова* (ім., прикм.), *прошле* (ім., прикм.), *рано* (ім., присл.), *святий* (ім., прикм.), *свято* (ім., присл.), *слід* (ім., присл.), *собі* (займ., частка), *справа* (ім., присл.), *столова* (ім., прикм.), *та* (спол., частка), *так* (присл., частка), *тепло* (ім., присл.), *це* (частка, займ.), *чи* (частка, спол.), *чому* (присл., займ.), *шкода* (ім., служб.), *що* (займ., спол., частка), *як* (присл., спол., частка) та інші. У словнику для їх розрізнення подано вказівку на частиномовну належність або на їх значення.

Фонетичні варіанти слів зведено до основної форми: *б/би/би-м; в/у; бачитися/бачитись; вбити/убити; вважатись/уважатися; весь/весь/увесь; лякатися/лякатись* тощо.

#### Лінгвостатистична характеристика роману І. Франка “Не спитавши броду”.

У процесі квантитативної параметризації великої прози І. Франка [4] виробилася схема опису статистичних параметрів тексту, відповідно до якої подамо нижче характеристику твору. Обсяг тексту (N) – **49 170 слововживань, серед яких 27 580 слововживань належить авторській, а 21 590 – прямій мові.** За обсягом “Не спитавши броду” перебуває на шостому місці після романів “Перехресні стежки” (93 890), “Борислав сміється” (77 455), “Основи суспільності” (67 172), “Петрії й Довбушука” (52 751), “Захар Беркут” (50 223), перед творами “Для домашнього огнища” (44 841), “Великий шум” (37 005), “*Voа constrictor*” (25 427). **Як видно з наведених даних, за обсягом твір найбільше наближений до “Захара Беркута”, а оскільки статистичні характеристики коректно порівнювати лише з тими, які отримані на вибірці такого ж обсягу, то зіставлення зроблено саме з цим твором за результатами, опублікованими у [2].**

Загальна кількість лем у тексті “Не спитавши броду” (V) – 7 140. Завдяки обсягу словника лем обчислено індекс різноманітності (багатство словника,  $V/N$ ) – 0,15 (у “Захарі Беркуті” – 0,13) та середню повторюваність слова в тексті ( $N/V$ ) – 6,9, тобто

в середньому кожне слово в обстеженому тексті вживається майже 7 разів, що трохи менше, ніж у “Захарі Беркуті” (7,7 разів).

Кількість *hapax legomena* ( $V1$ ), тобто слів, що по одному разу трапились у творі, – 3 834. Завдяки цій величині обчислено показники варіативності лексики, тобто індекс винятковості для тексту ( $V1/N = 0,08$ ) і для словника ( $V1/V = 0,54$ ). Тобто *hapax legomena* становлять 7,8 % для тексту і 53,7 % для словника. Таке відсоткове співвідношення гапаксів, коли вони “покривають 40–60% будь-якого тексту”, А. Корнаї вважає загальною універсалиєю [34, р. 72].

Кількість слів, які у тексті мають частотність вищу за 10 ( $V10T$ ), – 35 938 (73,09 %). Кількість слів, які у словнику мають частотність вищу за 10 ( $V10$ ), – 625 (8,75 %). Ці характеристики беруть участь у визначенні індексів концентрації у тексті ( $V10T/N = 0,73$ ) і словнику ( $V10/V = 0,09$ ).

*Співвідношення частин мови* у тексті може слугувати статистичним параметром індивідуального авторського стилю, а також особливістю конкретного твору [17, с. 152]. Оскільки корпус текстів роману має морфологічне маркування (застосовано класичний поділ слів на частини мови, до службових зараховано прийменник, сполучник, частку та, умовно, вигук), для кожної частини мови було автоматично отримано її частотність у словнику (реєстрі) і тексті роману (див. табл. 2).

Таблиця 2

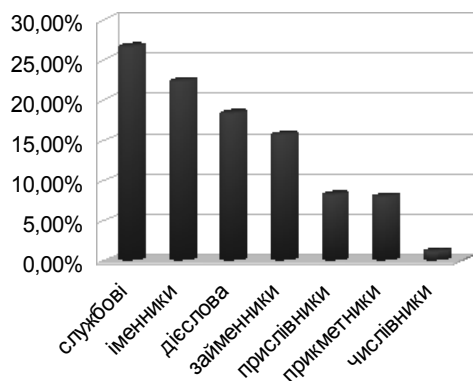
Відносна частотність частин мови у словнику і тексті роману І. Франка “Не спитавши броду”

частини мови	слів у тексті		слів у словнику	
службові	13092	26,63%	211	2,96%
іменники	10961	22,29%	2383	33,38%
дієслова	9007	18,32%	2447	34,27%
прикметники	3890	7,91%	1362	19,08%
займенники	7680	15,62%	64	0,90%
прислівники	4008	8,15%	614	8,60%
числівники	530	1,08%	57	0,80%
фрагменти	2	0,00%	2	0,03%
	<b>49170</b>	<b>100,00%</b>	<b>7140</b>	<b>100,00%</b>

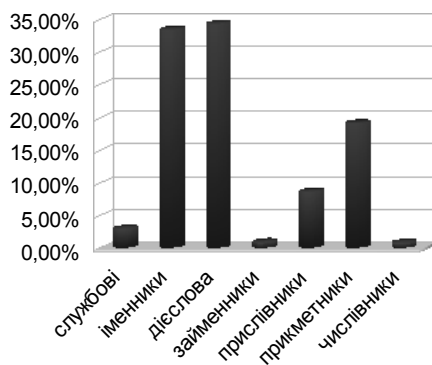
Як видно із таблиці 2, найчастотнішими словами у тексті очікувано є службові частини мови: займаючи у словнику менше 3 %, у тексті вони функціонують найактивніше і охоплюють більше, ніж його чверть (26,63 %). Подібну високу активність у тексті має ще один замкнений клас слів – займенники: маючи менше відсотка у словнику, вони займають майже 16 % тексту. Приблизно однакову частку у словнику (відповідно 7,1–8,8% і 0,6–0,9%) й тексті (відповідно 7,0–9,1% і 0,8–1,3%) покривають прислівники та числівники: 8,15/8,60 та 1,08/0,80 відповідно.

Найрізноманітнішими виявились іменники, дієслова та прикметники: їх відносна кількість у словнику навпаки перевищує відносну кількість у тексті. Іменники становлять фактично 23 % тексту і 33 % словника, відносна кількість дієслів у тексті (34 %) вдвічі перевищує їх відносну кількість у словнику (18 %), у прикметників перевищує

майже втричі: 19 % / 8 %. Саме від цих морфологічних класів залежить багатство лексики твору. Гістограми 1 і 2 наочно показують цю ситуацію:



Гістограма 1: Відсотковий розподіл частин мови у *тексті* роману “Не спитавши броду”



Гістограма 2: Відсотковий розподіл частин мови у *реєстрі словника* роману “Не спитавши броду”

У лінгвостатистиці також обчислюють **кількісні реляції між частинами мови**, вважаючи їх одним із елементів статистичної характеристики тексту. До них зараховують індекс іменних означень, тобто індекс епітетизації (відношення суми вживань іменників до суми вживань прикметників), індекс дієслівних означень (відношення суми вживань прислівників до суми вживань дієслів) [33, s. 128], ступінь номінальності (відношення суми вживань іменників до суми вживань дієслів) [35, s. 50]. Вказані величини для роману І. Франка “Не спитавши броду” подано у порівнянні з повістю “Захар Беркут”, та середнім показником для прози І. Франка у табл. 3.

Таблиця 3

Кількісні реляції між частинами мови у творах І. Франка “Не спитавши броду”, “Захар Беркут” та в українській художній прозі

	<i>Не спитавши броду</i>	<i>Захар Беркут</i>	Проза І. Франка	Українська проза
індекс епітетизації	2,82	2,69	3,13	3,00
індекс дієслівних означень	0,44	0,46	0,46	0,46
ступінь номінальності	1,22	1,56	1,30	1,41

У “Не спитавши броду” на один іменник припадає дещо більше прикметників/епітетів (2,8), ніж в “Захарі Беркуті” (2,7), проте менше, ніж у великій прозі письмен-

ника (3,1) й українській прозі загалом (3). Індекс дієслівних означень показує, скільки прислівників припадає на одне дієслово: у досліджуваному тексті цей показник трохи менший (4,4 прислівників на 10 дієслів), ніж у “Захарі Беркуті” і великій прозі письменника загалом (4,6 прислівників на 10 дієслів). Індекс номінальності у “Не спитавши броду” – 1,22; це означає, що на одне дієслово у тексті є 1,22 іменника (в “Захарі Беркуті” – 1,56, у великій прозі І. Франка загалом – 1,3, у прозі ХХ ст. – 1,4). Хоча індекси епітетизації, номіналізації та дієслівних означень є лише одним із багатьох інструментів, використовуваних для стилістичного аналізу тексту, їх можна вважати потужним інструментом для доповнення якісного аналізу тексту.

З погляду статистичної лінгвістики роман І. Франка “Не спитавши броду” отримав кількісну параметризацію, що є, поряд із домінантним у сучасному мовознавстві якісним аналізом, невід’ємною частиною комплексного дослідження твору. Проведена на матеріалі електронного маркованого корпусу текстів, вона охоплює такі характеристики, як обсяг тексту, кількість різних слів, індекс різноманітності (багатство словника), середня повторюваність слова в тексті, кількість *hapax legomena*, індекс винятковості у тексті та у словнику, індекс концентрації, індекси епітетизації, номіналізації та дієслівних означень. Завдяки такому опису визначено місце роману серед інших творів великої форми І. Франка. У перспективі заплановано також окремо дослідити покриття тексту високо- та низькочастотною лексикою, співвідношення між рангом слова та величиною покриття тексту, семантичні особливості лексики в різних частотних зонах словника, кількісні особливості функціонування власних назв та інших груп слів у контексті великої прози І. Франка загалом.

Послідовно змалювавши лінгвостатистичні портрети усіх творів І. Франка (застосовуючи однакові принципи та методи дослідження), ми отримаємо комплексне уявлення про статистичну структуру тексту стилю письменника, що було би неможливим без урахування роману “Не спитавши броду”.

#### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. *Баевский В. С.* Частотная структура лексики А. Вознесенского (“Антимиры”, лирика) / В. С. Баевский // *Русская авторская лексикография XIX–XX веков.* – М., 2003. – С. 111–112.
2. *Бук С.* “Захар Беркут” Івана Франка у світлі статистичної лінгвістики / С. Бук // *Науковий часопис національного педагогічного університету імені М. П. Драгоманова.* – 2010. – Вип. 6. – С. 172–176. – (Серія 10. Проблеми граматики і лексикології української мови).
3. *Бук С.* Інтерпретація лексики роману Івана Франка “Великий шум” у кількісно-статистичному аспекті / С. Бук // *Studia Methodologica.* – 2010. – Вип. 30. – С. 145–153.
4. *Бук С.* Квантитативна параметризація текстів Івана Франка: спроба проекту / С. Бук // *Іван Франко : студії та матеріали.* – Львів, 2011. (у друці).
5. *Бук С.* Корпус текстів Івана Франка : спроба визначення основних параметрів / С. Бук // *Прикладна лінгвістика та лінгвістичні технології : MegaLing-2006.* – К., 2007. – С. 72–82.
6. *Бук С.* Роман Івана Франка “Для домашнього огнища” крізь призму частотного слов-



- ника / С. Бук. – 2010. Препринт arXiv:1006.0153v1 [cs.CL]. – Режим доступу: <http://arxiv.org/abs/1006.0153>.
7. Бук С. Статистична структура роману І. Франка “Борислав сміється” / С. Бук // Ученые записки Таврического национального университета им. В. И. Вернадского. – 2010. – Т. 23(62). – № 3. – С. 114–118. – (Серия “Филология”).
  8. Бук С. Статистичні характеристики роману Івана Франка “Основи суспільності” (на основі частотного словника твору) / С. Бук // Вісник : Проблеми української термінології. – 2010. – № 676. – С. 90–93.
  9. Бук С. Структурне анотування у корпусі текстів (на прикладі прози Івана Франка) / С. Бук // Українська мова. – 2009. – № 3. – С. 59–71.
  10. Бук С. Он-лайн конкорданс роману Івана Франка “Перехресні стежки” / С. Бук, А. Ровенчак // Іван Франко : дух, наука, думка, воля : матеріали Міжнародного наукового конгресу, присвяченого 150-річчю від дня народження Івана Франка (Львів, 27 вересня – 1 жовтня 2006 р.). – Львів, 2010. – Т. 2. – С. 203–211.
  11. Бук С. Он-лайн конкорданс роману Івана Франка “Перехресні стежки” / С. Бук, А. Ровенчак. – Режим доступу: <http://www.ktf.franko.lviv.ua/~andrij/science/Franko/concordance.html>. (2006–2011)
  12. Бук С. Частотний словник повісті І. Франка “Перехресні стежки” / С. Бук, А. Ровенчак // Стежками Франкового тексту (комунікативні, стилістичні та лексичні виміри роману “Перехресні стежки”). – Львів, 2007. – С. 138–369.
  12. Вервес Г. Незакінчена повість Івана Франка “Не спитавши броду” (До проблеми українсько-польських громадських зв’язків) / Г. Вервес // Славістичний збірник. – К., 1963. – С. 214–283.
  14. Возняк М. Спроба відбудови невикінченої повісті / М. Возняк // Червоний шлях. – 1929. – № 1. – С. 5–9.
  15. Гребенников А. О. Частотный словарь рассказов А. П. Чехова / А. О. Гребенников ; под ред. Г. Я. Мартыненко. – СПб., 1999.
  16. Честотен речник на Вапцаровата поезія / Г. В. Крилова, А. А. Азарова, Е. А. Захаревич, и др. – Велико Търново, 1996.
  17. Перебийніс В. С. Частотні словники та їх використання / В. С. Перебийніс, М. П. Муравицька, Н. П. Дарчук. – К., 1985.
  18. Статистичні параметри стилів / за ред. В. С. Перебийніс. – К., 1967.
  19. Франко І. Гава. Образок з життя підкарпатського народу / І. Франко. – Львів : Просвіта, 1888. – № 110. кн. 11.
  20. Франко І. На лоні природи / І. Франко // Зібр. творів : у 50-ти т. / І. Франко. – К., 1979. – Т. 18. – С. 154–163.
  21. Франко І. Не спитавши броду / І. Франко // Зібр. творів : у 50-ти т. / І. Франко. – К., 1979. – Т. 18. – С. 325–463.
  22. Франко І. Не спитавши броду. Повість (Реконструкція повісті і післямова Г. Д. Вервеса) / І. Франко. – К., 1966.
  23. Франко І. Не спитавши броду / І. Франко // Червоний шлях. – 1927. – № 1. – С. 5–54; 1929. – № 1. – С. 5–33; № 2. – С. 5–22.
  24. Частотний словник сучасної української поетичної мови. – Режим доступу: <http://www.mova.info/Page2.aspx?l1=89>. (2003–2011).
  25. Частотний словник сучасної української художньої прози у 2 т. / за ред. В. С. Перебийніс. – К., 1981.
  26. Шайкевич А. Я. Статистический словарь языка Достоевского / А. Я. Шайкевич, В. М. Андрющенко, Н. А. Ребецкая. – М., 2003.



27. *Balowski M.* Lista frekwencyjna poezji, prozy i dramatu Krzystofa Kamila Baczynskiego / M. Balowski. – Prohowice, 1997.
28. Complete Shakespeare wordlist. – Available from: <[http://en.wiktionary.org/wiki/Wiktionary:Frequency\\_lists/Complete\\_Shakespeare\\_wordlist](http://en.wiktionary.org/wiki/Wiktionary:Frequency_lists/Complete_Shakespeare_wordlist)>.
29. *Cvrček V.* Slovník Bohumila Hrabala / V. Cvrček, F. Čermák. – Praha, 2009.
30. *Franko I.* Hawa. Obrazek z natury / I. Franko // Kurjer Warszawski. – 1888. – № 31–33 (od 31 stycznia–2 lutego).
31. *Hanley M. L.* Word index to James Joyce's "Ulysses" / M. L. Hanley. – Madison: University of Wisconsin Press, 1951 ; Ulysses by J. Joyce: A Ranked Concordance <<http://www.doc.ic.ac.uk/~rac101/concord/texts/ulysses/>>.
32. *Holman P.* Frekvenční slovník básnického díla Otokara Březiny (něm. s tit. **Frequenzwörterbuch zum lyrischen Werk von Otokar Březina**), 2 sv. / P. Holman. – Köln am Rhein – Weimar – Wien, 1993.
33. *Kamińska-Szmaj I.* Części mowy w słowniku i tekście pięciu stylów funkcjonalnych polszczyzny pisanej (na materiale słownika frekwencyjnego) / I. Kamińska-Szmaj // Biuletyn Polskiego Towarzystwa Językoznawczego. – 1988. XLI. – S. 127–136.
34. *Kornai A.* Mathematical Linguistics / A. Kornai. – Springer, 2008.
35. *Ruszkowski M.* Wskaźnik epitetyzacji w badaniach stylistycznych / M. Ruszkowski // *Respectus Philologicus*. – 2004. – № 5(10). – S. 48–53.
36. Slovník Karla Čapka / Čermák F. (Hl. ed.). – Praha, 2007.
37. *Vasić S.* Polazne osnove novije srpske proze frekvencijski rečnik romana Oslobođici i izdajnici Milorada Đanojlića / S. Vasić. – Beograd, 2002.
38. *Vasić S.* Polazne osnove novije srpske proze. Knjiga II. Hazarski rečnik Milorada Pavića : frekvencijski rečnik / S. Vasić. – Beograd, 1998.

Додаток

Частотний словник роману І. Франка "Не спитавши броду" (фрагмент\*)

Ранг	Лема	Абс. част.	Відн. част., %	Покриття тексту, %
1	І – 1601; Й – 331	1 932	3,929	3,93
2	НЕ	1 273	2,589	6,52
3	В – 775; У – 205	980	1,993	8,51
4	ВН	944	1,920	10,43
5	Я	811	1,649	12,08
6	НА	755	1,535	13,62
7	З – 703; ЗІ – 6; ЗО – 31; ІЗ – 4	744	1,513	15,13
8	А (спол.)	714	1,452	16,58
9	ТОЙ	709	1,442	18,02
10	БУТИ	633	1,287	19,31
11	ЩО (займ.)	532	1,082	20,39
12	ДО	509	1,035	21,43
13	ЩО (спол.)	456	0,927	22,36
14	ВОНА	453	0,921	23,28
15	ВИ	442	0,899	24,18
16	БОРИС (ім'я)	408	0,830	25,01

Ранг	Лема	Абс. част.	Відн. част., %	Покриття тексту, %
17	ТО (част.)	382	0,777	25,78
18	СКАЗАТИ	344	0,700	26,48
19	АЛЕ	337	0,685	27,17
20	СЕЙ	335	0,681	27,85
21	ТА (спол.)	329	0,669	28,52
22	Ж (част.) – 234; ЖЕ – 58	292	0,594	29,11
23	ВЕСЬ – 216; ВВЕСЬ – 1; У~ – 67	284	0,578	29,69
24	ТИ	272	0,553	30,24
25	ТАК (присл.)	264	0,537	30,78
26	СВІЙ	256	0,521	31,30
27	ТАКИЙ	255	0,519	31,82
28	ПАНІ	251	0,510	32,33
29	ЗА	247	0,502	32,83
30	ЙОГО (присв. займ. ч. р.)	247	0,502	33,33
31	МОГТИ	246	0,500	33,83
32	ЗНАТИ	231	0,470	34,30
33	ПО	214	0,435	34,74
34	ЯК (спол.)	211	0,429	35,17
35	КОЛИ	205	0,417	35,58
36	Б – 128; БИ – 72; БИ-М – 1	201	0,409	35,99
37	ВОНИ	198	0,403	36,40
38	ЩЕ – 180; ЩЕ – 2	182	0,370	36,77
39	КОТРИЙ	181	0,368	37,13
40	МИ	181	0,368	37,50
41	О (виг.)	172	0,350	37,85
42	ТІЛЬКИ (~О – 68)	172	0,350	38,20
43	ПАН	170	0,346	38,55
44	МІЙ	167	0,340	38,89
45	РУКА	166	0,338	39,23
46	НУ	161	0,327	39,55
47	ЧИ (спол.)	153	0,311	39,86
48	ЩОБ – 149; ЩОБИ – 3	152	0,309	40,17
49	ЯКИЙ	150	0,305	40,48
50	ОКО	148	0,301	40,78
51	ВЖЕ – 115; У~ – 27	142	0,289	41,07
52	ВІД	141	0,287	41,35
53	ОДИН	138	0,281	41,64
54	НІ	128	0,260	41,90

Ранг	Лема	Абс. част.	Відн. част., %	Покриття тексту, %
55	ДЛЯ	127	0,258	42,15
56	МОВ	120	0,244	42,40
57	СА́МИЙ	120	0,244	42,64
58	МАТИ (дісл.)	117	0,238	42,88
59	ГУСТЯ (ім'я)	115	0,234	43,11
60	Її (присв. займ.)	111	0,226	43,34
61	ЯКИЙСЬ	111	0,226	43,57
62	РАЗ (ім.)	106	0,216	43,78
63	СОБИ (част.)	106	0,216	44,00
64	ТОНЬО (ім'я)	103	0,209	44,21
65	ДУМАТИ	97	0,197	44,40
66	ГОВОРИТИ	96	0,195	44,60
67	ІТИ – 48; ЙТИ – 47	95	0,193	44,79
68	СЕБЕ	95	0,193	44,98
69	НЩО	92	0,187	45,17
70	ХОЧ	91	0,185	45,36
71	ХВИЛЯ	90	0,183	45,54
72	ДРУГИЙ	86	0,175	45,71
73	ХОПТИ	86	0,175	45,89
74	БО	84	0,171	46,06
75	СЛОВО	82	0,167	46,23
76	ТУ (ТУТ)	82	0,167	46,39
77	МІХОНСЬКА (прізви.)	81	0,165	46,56
78	ВАШ	80	0,163	46,72
79	ТЕПЕР	80	0,163	46,88
80	ХТО	80	0,163	47,05
81	ЩО (част.)	79	0,161	47,21
82	НЕМОВ	78	0,159	47,37
83	СКРИКНУТИ	78	0,159	47,52
84	КАЗАТИ	77	0,157	47,68
85	СЕРЦЕ	77	0,157	47,84
86	ЯК (присл.)	77	0,157	47,99
87	ЛИЦЕ	76	0,155	48,15
88	ЛЮБИТИ	76	0,155	48,30
89	РОБИТИ	75	0,153	48,46
90	ЧАС	75	0,153	48,61
91	ГОЛОВА	74	0,150	48,76
92	ТРЕБА	73	0,148	48,91
93	СТАРИЙ	72	0,146	49,05

Ранг	Лема	Абс. част.	Відн. част., %	Покриття тексту, %
94	ОГ	71	0,144	49,20
95	ПЕРЕД (прийм.)	71	0,144	49,34
96	БОГ	70	0,142	49,49
97	СТОЯТИ	70	0,142	49,63
98	ПТИ	69	0,140	49,77
99	АЖ	68	0,138	49,91
100	БРАТ	68	0,138	50,04
101	ДАТИ	68	0,138	50,18

\* Знак ~ замінює повторювану частину слова: стаття “ТІЛЬКИ (~О – 68) 172” означає, що форма ТІЛЬКО у тексті трапилася 68 разів, форма ТІЛЬКИ 172 – 68 = 104 рази; стаття “ВЖЕ – 115; У~ – 27” означає, що форма ВЖЕ трапилася в тексті 115 разів, а форма УЖЕ – 27 разів.

*Стаття надійшла до редколегії 10.09.2011  
Прийнята до друку 26.09.2011*

## LINGUOSTATISTIC DESCRIPTION OF “NE SPYTAVSHY BRODU” (“WITHOUT ASKING A WADE”) BY IVAN FRANKO

**Solomiya Buk**

*The Ivan Franko National University of Lviv,  
Department for General Linguistics,  
1, Universytetska Str., room 343, Lviv, Ukraine, UA-79000  
tel.: (032) 239 47 56, e-mail: solomija@gmail.com*

The article presents the analysis of the novel “Ne spytavshy brodu” (Without Asking a Wade) by Ivan Franko from the position of linguostatistics. Such an analysis is made for the first time. The frequency dictionary was compiled on the material of an electronic text corpus with an external and internal markup. The following parameters were obtained: text and vocabulary sizes, variety, exclusiveness, concentration indexes, the direct and author’s speech correlation, the parts of speech relations in the text and vocabulary, the indexes of epithetization, nominalization, and verbal definitions.

*Key words:* linguostatistics, text corpus, statistical parameters, quantitative characteristics of text, frequency dictionary, lemma, vocabulary richness.

## ЛИНГВОСТАТИСТИЧНИЙ ОПИСАНИЕ “НЕ СПРОСИВ БРОДЫ” ИВАНА ФРАНКО

Соломія Бук

*Львовский национальный университет имени Ивана Франко,  
кафедра общего языкознания,  
ул. Университетская, 1, к. 343, Львов, Украина, 79000  
тел.: (032) 239 47 56, e-mail: solomija@gmail.com*

Статья представляет анализ романа И. Франко “Не спитавши броду” (“Не спросив брода”) с позиций лингвостатистики, который осуществлен в языкознании впервые. На материале электронного корпуса текста, имеющего внешнюю и внутреннюю разметку, составлен частотный словарь произведения, с помощью которого получены такие его параметры, как объем текста и словаря, индексы разнообразия и концентрации, соотношение прямой и косвенной речи, соотношение частей речи в словаре и тексте, индексы эпитетизации, глагольных определений, степень номинальности.

*Ключевые слова:* лингвостатистика, корпус текстов, разметка, статистические параметры стиля, количественные характеристики текста, словоупотребление, лемма, богатство словаря.