

УДК 811.161.2'1'324'38(038) І. Франко

ПРЯМА Й АВТОРСЬКА МОВА ВЕЛИКОЇ ПРОЗИ ІВАНА ФРАНКА: ЛІНГВОСТАТИСТИЧНЕ ДОСЛІДЖЕННЯ У КОНТЕКСТІ КОРПУСНОЇ ЛІНГВІСТИКИ

Соломія Бук

*Львівський національний університет імені Івана Франка,
кафедра загального мовознавства,
вул. Університетська, 1, ауд. 343, 79000 Львів, Україна,
тел.: +380 (32) 2394756, e-mail: solomija@gmail.com*

Презентовано кількісне зіставлення співвідношень прямої та авторської мови у великій прозі І. Франка (що може слугувати статистичним параметром ідіостилу), зокрема його порівняно з українською художньою прозою ХХ ст. Праця містить короткий опис досвіду маркування цих типів тексту в лексикографії та корпусній лінгвістиці, а також обчислення багатства словника у цих пластах лексики художньої спадщини письменника. Дослідження виконано на матеріалі електронного корпусу текстів І. Франка.

Ключові слова: пряма мова, авторська мова, обсяг тексту, слововживання, корпус текстів, маркування корпусу, лексичне багатство словника.

Мова як абстрактна семіотична система та її конкретна реалізація у мовленні як усному, так і письмовому (тексті) – унікальний об'єкт міждисциплінарних досліджень ХХІ ст. Застосування до її вивчення математичного інструментарію з комп'ютерними технологіями сприяє переосмисленню наукового статусу лінгвістичної дисципліни.

Пряма мова у творі, тобто мовлення персонажів, викликала і викликає науковий інтерес серед лінгвістів. Дихотомія прямої та авторської мови є важливим компонентом дослідження лінгво-стильових особливостей як цілих функціональних стилів [1, 14, 20], так й ідіолекту окремого письменника [13, 15, 22]. Через пряму мову автор подає мовну характеристику персонажів [18, 24]. Різномасштабному дослідженню цього пласту лексики присвячено багато праць, зокрема, наскільки істотно можуть різнитися частотні характеристики слів у цих двох площинах мови художнього твору довів «Частотний словник сучасної української художньої прози» за редакцією В. С. Перебийніс [25]. Кількісне співвідношення прямої й авторської мови у творах І. Франка об'єктом наукових студій стало вперше. Дослідження виконано в межах комплексної квантитативної параметризації творів письменника на матеріалі електронного корпусу текстів його великої прози, окремими результатами якого уже стали частотні словники романів «Борислав сміється» [8], «Захар Беркут» [6], «Для домашнього огнища» [7], «Основи суспільності» [9], «Перехресні стежки» [12] (а також онлайн конкорданс [10, 11]), Великий шум [3].

У статті порівняно пропорції прямої та авторської мови у великій прозі І. Франка, уточнено дефініцію терміносполук *пряма й авторська мова*, здійснено короткий опис досвіду маркування цієї антиномії в лексикографії та корпусній лінгвістиці, викладено концептуальні засади анування прямої мови в корпусі текстів письменника, подано кількісне співвідношення наративної та діалогічної частини тексту в його романах, окреслено багатство словника у прямій та авторській мові великої прози І. Франка.

Дефініція прямої мови. Пряму мову у статті трактуємо як дослівно відтворене мовлення самого автора або іншої особи, різновид чужої мови, переданої від тієї особи, якій вона належить [2, с. 538]. Вона зберігає індивідуальні та стилістичні особливості мовлення того, чиє висловлення відтворено: діалектні риси, повтори, паузи тощо. Чужу мову протиставлено авторській, тобто тій, яка належить оповідачеві, мовцю. Напівпряма мова та невласне-пряма мова належить до мови автора, в яку вони вплітаються, виконуючи, як і власне пряма мова, розмаїті стилістичні функції: комунікативно-естетичну, створення характерів, соціальної характеристики героїв тощо. Вважаємо, що основу художньої прози творить авторський наратив, а мовлення персонажів входить у неї як один зі суттєвих компонентів.

Розмежовуючи текст твору на дві частини – пряму та авторську мову, – листи, записки, телеграми і т. п. відносимо до першого типу текстів, хоча вони, безперечно, створюють окремі мовленнєві жанри, які у подальшому варто аналізувати окремо.

Досвід маркування прямої мови в лексикографії та корпусній лінгвістиці. Робота з корпусами, тобто з великими масивами текстів в електронному вигляді, де марковано лінгвістичну інформацію [16, 17], «давно уже стала одним із основних, якщо не головним методом лінгвістичних досліджень, за допомогою якого можуть вирішуватися найрізноманітніші завдання» [21].

У сучасній лінгвістиці мало квантитативних досліджень прямої мови у художній прозі. Хоча відмінність між авторським наративом і мовленням персонажів суттєва й очевидна, але корпуси не завжди розділяють їх маркуванням, часто вважаючи художню прозу гомогенним жанром. Як зазначає К. Акселссон, пряму мову у корпусі часто складно знайти, до того ж немає інформації про кількість її лексем, тому складно виконати частотні обчислення [27, с. 191]. Так, авторитетний Британський національний корпус [28], як і багато інших, не пропонує цієї інформації.

З іншого боку, створено цілі корпуси текстів діалогічного мовлення, як, наприклад, корпус англійських діалогів 1560–1760 років. [29]. Міжнародний стандарт TEI [33], Браунівський корпус [31], Англійсько-норвезький паралельний корпус [32], Чеський народний корпус [30], Російський національний корпус [19], корпус творів М. Достоевського [26, с. 21] по-різному маркують ці види мовлення в художньому творі. Міжнародний стандарт кодування текстової інформації TEI (англійською Text Encoding Initiative – проект кодування текстів), який «репрезентує ті ознаки тексту, які потребують експліцитної ідентифікації з метою сприяти текстовому опрацюванню за допомогою комп'ютерних програм» [33], пряму мову трактує одним із елементів універсальної структури тексту і пропонує для її позначення декілька окремих маркерів. Наприклад, в Англійсько-норвезькому паралельному корпусі [32] таке маркування виконано частково автоматично, частково вручну. Російський національний корпус до прямої мови застосовує синтаксичне маркування [19].

В українській лексикографії Частотний словник сучасної української художньої прози за редакцією В. Перебийніс [25] також розрізняє ці два способи висловлювання в тексті і тому подає частоту кожного слова як окремо для прямої й авторської мови, так і сумарну (див. табл. 1.).

Таблиця 1

Зразок подання інформації про пряму та авторську мову
у Частотному словнику сучасної української художньої прози [25]

Слово, словоформа, граматичні характеристики		Пряма мова		Авторська мова		n	R	r		
		n	p, %	n	p, %					
*АБСОЛЮТНО	присл.	8	0,056	3	0,0008	11	6	11	0,022	0,0065
АВАНГАРД	ім.	2	0,0014	–	–	2	1	1	0,004	0,0039
авангард	наз.	1	0,0007	–	–	1	1	1	0,002	–
авангард	зн.	1	0,0007	–	–	1	1	1	0,002	–
АВАНТЮРА	ім.	3	0,0021	2	0,0005	5	2	3	0,01	0,0059
авантюра		2	0,0014	–	–	2	1	1	0,004	0,0039
авантюру		1	0,0007	–	–	1	1	1	0,002	–
авантюр		–	–	1	0,0003	1	1	1	0,002	–
авантюри	зн. мн.	–	–	1	0,0003	1	1	1	0,002	–
АВАРІЙНИЙ		–	–	2	0,0005	2	2	2	0,004	0,0027
аварійний	зн.	–	–	1	0,0003	1	1	1	0,002	–
аварійною		–	–	1	0,0003	1	1	1	0,002	–
АВАРІЯ	ім.	3	0,0021	6	0,0016	9	3	6	0,018	0,0076
аварія		2	0,0014	–	–	2	1	1	0,004	0,0039
аварії	род.	–	–	4	0,0011	4	2	3	0,008	0,0048
аварій		–	–	2	0,0005	2	1	1	0,004	0,0039

n – абсолютна частота, p – відносна частота, R – кількість джерел, – середня частота, r – кількість мінімальних вибірок, – міра коливання середньої частоти

Засади маркування прямої мови у корпусі текстів Івана Франка. Проект створення корпусу текстів І. Франка, який реалізується у Львівському університеті [4], передбачає спеціальне маркування прямої мови, яке належить до структурного типу [5]. В електронному корпусі текстів письменника формальними ознаками прямої мови прийнято лапки (приклад 1), тире перед і після слів автора (приклад 2), а також тире на початку абзацу та кінець абзацу (приклад 3):

(1) *А коли скінчиться дивовижна повість, то малі й старі, зітхаючи, шепчуть: «Ах, яка ж то красна байка!»* («Захар Беркут»¹)

(2) *– Ти, Бенедю, – спитав якийсь замазаний глиною робітник, – а то на яку пам'ятку того щигля сюда вивісили?»* («Борислав сміється»)

(3) *– Я Ціцк Шуберт, знаєш? Моя бабка там жила з твоєю мамою, знаєш?»* («Воа constrictor»)

Маркування прямої мови виконано напівавтоматично, оскільки лапки також можуть слугувати оформленням цитат (приклад 4), метафоричного значення чи іронії (приклади 5–7), а також власних назв (приклади 8, 9).

(4) *«Blut ist ein ganz besondrer Saft», – брнів йому в голові тетівський вірш, оті іронічні та при тім глибоко символічні слова Мефістофелеві* («Для домашнього огнища»)

(5) *«Взагалі тут Герман перший раз пізнав, що то за народ тоті «гої», і його хлоп'ячий розум швидко покмітив, що в кожного з «жидівської віри», так сказати, двоє лиць ...»* («Воа constrictor»)

(6) *...він [Антін Субота] представляв із себе те, що називають «гарний тип польського шляхтича»* («Великий шум»)

(7) *Лице «сусіди» скривилося легким іронічним сміхом* («Петрії й Довбушуки»)

¹ Тут і далі уривки великої прози І. Франка взято з електронного корпусу текстів, тому у паспортизації не подаємо номерів сторінок.

(8) *І от коли по довгих літах, переглядаючи «Спархіяльні відомості», побачив, що опорожнена парафія Торки, де жила його давніша улюблена, в серці його не обізвася ніякий голос, а тільки в умі почали метатися цифри більших доходів («Основи суспільності»)*

(9) *Нараз на одній елегантній, блискучій прочитав назву: «Waldemar Baron von Reuchlingen» і затремтів так, немовби пожалив собі пальці кропивою («Для домашнього огнища»)*

Тире можуть виконувати функцію заміщення дієслівної частки «бути» і т. д.

До того ж трапляються випадки однієї прямої мови у межах іншої (приклад 10, 11), а також двох «вкладених» одна в одну прямих мови в іншій (приклад 12):

(10) – *А де ж би, дома. Чекали, чекали на паннуцію, а далі кажуть до мене: «Е, – кажуть, – Параско, мабуть наша паннуція десь там і заночує в гостині, а мені ту годі голодувати та на ню чекати. Давай їсти, що там маєш». Іно я наклала в тарілку пирогіє і взялась сметану збирати, аж і паннуція над'їхали («Не спитавши броду»).*

(11) – *Та нічого особливого! – відбрїхувався Гадина. – Те, що кажу, те й бачив. Застукає хтось до дверей, о. Нестор запитав: «Хто там?» Відповідає: «Я, Деменюк». О. Нестор вітворив. Увійшов Деменюк, оба почали розмовляти... («Основи суспільності»).*

(12) – *Так, так! Ха, ха, ха! Знаєте, ми з жінкою вчора реготалися так!.. Треба вам було чути, з яким обуренням оповідали нам ті люди... Як тільки вернули з міста, зараз до мене. «Просимо ласки панської, ми би хотіли спитатися, що то за адукат такий. Ми йому віддали свою справу за ту толоку, а він нам каже: «Ви дурні. – Так нам казав, їй-богу! – Ви дурні! Що вам, – каже, – правуватися з паном за якусь толоку? Що то та толока варта? Тьфу! Я вам не те скажу. Дайте мені тисячу ринських, а я вам виправую весь панський маєток, з будинками, фільварками і лісами». Ха, ха, ха! («Перехресні стежки»²)*

Слова автора можуть відокремлюватися в прямій мові й дужками (приклад 13):

(13) – *... Десять років минуло (вона при сих словах обернулася знов до мужа), – мої злі демони або покинули сю сукню, покинули мене, і в такім разі нинішній день буде новим зворотом у моім житті і варт того, щоб зустріти його празнично. ... («Перехресні стежки»)*

Прямую мовою ми вважали також відповідно оформлені думки (приклади 14, 15):

(14) *«Се ж комедія, не жадне дійсне сватання, – думалось йому. – Де ж таки се чувано й видано!» («Великий шум»)*

(15) *«Тепер прийдеться загибати!» – блиснуло в думці в Мирослави... («Захар Беркут»)*

Співвідношення прямої та авторської мови в романах Івана Франка.

Нарація автора і мовлення персонажів у різних романах І. Франка мають різну кількісну пропорцію (див. табл. 2).

Таблиця 2

Співвідношення прямої та авторської мови в романах Івана Франка

романи	пряма мова		авторська мова		разом	
	слововживань	%	слововживань	%	слововживань	%
«Воа constrictor», 1878	3 143	12,4	22 284	87,6	25 427	100
«Борислав сміється»	33 330	43,0	44 125	57,0	77 455	100
«Захар Беркут»	17 229	34,3	32 994	65,7	50 223	100

² До речі, саме у формі прямої мови автор подає ключові слова для розуміння теми й назви цього твору: «Стрічаються перехресні стежки на широкім степу та й знов розбігаються! Таке буде й наше. Що вона мені тепер, і що я їй? Нічогісінько. Перелетні тіні, що мигнуть понад долиною і не лишать по собі нічого-нічогісінько...» («Перехресні стежки»).

«Не спитавши броду»	21 590	43,9	27 580	56,1	49 170	100
«Для домашнього огнища»	18 914	42,2	25 927	57,8	44 841	100
«Основи суспільності»	28 278	42,1	38 894	57,9	67 172	100
«Перехресні стежки»	45 513	48,5	48 377	51,5	93 890	100
«Великий шум»	18183	49,1	18822	50,9	37 005	100
«Петрії й Довбушки», 1909– 1912	16 219	30,7	36 532	69,3	52 751	100
РАЗОМ:	202 399	40,6	295 535	59,4	497 934	100

У результаті порівняння даних усіх дев'яти україномовних романів І. Франка виявлено, що найменшу кількість прямої мови має «Воа constrictor» (12,4 %), що зумовлено особливостями нарації твору. «Захар ...» і «Петрії ...» – історичні твори, які потребують більше авторського опису, відповідно діалогів у них менше: 34,3 % і 30,7 %. Інші Франкові твори великої форми присвячені опису сучасного письменникової життя («Борислав сміється», «Не спитавши броду», «Для домашнього огнища», «Основи суспільності», «Перехресні стежки», «Великий шум»), і кількість діалогів у них суттєво вища: 42,1–49,1 %. Наочно це зображено на рисунку 1.

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

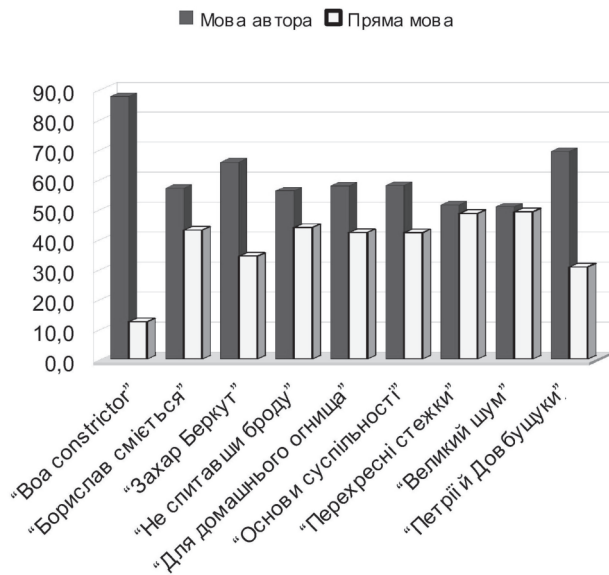


Рис. 1. Пропорція прямої та авторської мови в романах Івана Франка

Щоб виявити, наскільки ці параметри характеризують індивідуальний стиль письма І. Франка, важливо було би зіставити їх з даними інших письменників, зокрема сучасників, але, на жаль, аналогічних досліджень в україністиці виявити не вдалося. Проте Частотний словник сучасної української художньої прози подає таку інформацію. Співвідношення прямої та авторської мови в його джерельній базі становить 28,2 % до 71,8 % відповідно [25, т. 1, с. 15], тобто пряма мова займає трохи менше третини розглянутих текстів. Оскільки аналізований словник укладено на масиві

25 книжок 22 різних авторів, то цю пропорцію можна вважати середньостатистичним для української художньої прози середини ХХ ст. загалом.

Однією з основних вимог для коректного порівняння статистичних характеристик частотних словників є та, що вони повинні бути укладеними на матеріалі однакового обсягу, адже у процесі порівняння кількісних показників різних за обсягом текстів було помічено, що збільшення обсягу лексем не прямо пропорційне до збільшення обсягу тексту: «кількість слів і словоформ не збільшується такими самими темпами, як довжина тексту, а дещо відстає від неї» [25, т. 1, с. 15]. Обсяг обстежених романів І. Франка та корпусу текстів для укладання частотного словника художньої прози майже ідентичні: 497 934 і 500 357 слововживань відповідно, тому їх показники можна зіставляти без додаткових шкалювань. Отже, обсяг прямої мови творів І. Франка порівняно з українською художньою прозою ХХ ст. суттєво вищий, тут це співвідношення в середньому становить 40,6 % до 59,4 %. Наочно це зображено на рисунку 2.

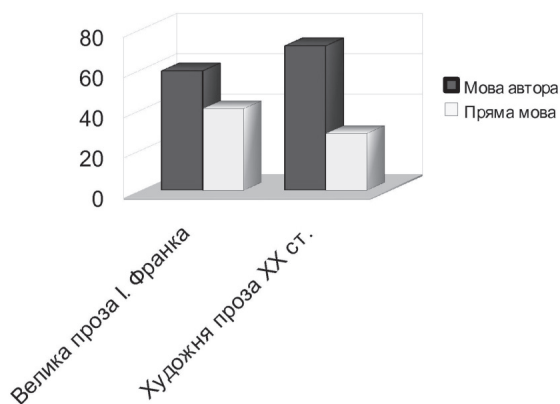


Рис. 2. Співвідношення прямої та авторської мови в романах Івана Франка та в українській художній прозі ХХ ст.

Вважається, що головна роль у будь-якому стилі належить авторському мовленню (яке творить основний корпус текстів і вирішує основні інформативні, комунікативні й естетичні завдання), а елементи чужого мовлення – це своєрідна інкрустація, що урізноманітнює авторське і надає йому різні стилістичні відтінки. У випадку великої прози І. Франка, де обсяг прямої мови перевищує 40% і в окремих випадках наближається до половини тексту (49 % у «Великому шумі»), це вже не можна назвати інкрустацією, це величезний рівноправний пласт лексики, без якого художній твір просто не може існувати. Пропорція наративної та діалогічної частини у творах різних письменників є різною, отже, може вважатися статистичним параметром індивідуального авторського стилю [23].

Багатство словника у прямій та авторській мові великої прози Івана Франка. Багатство словника (або індекс різноманітності) B – це одна з найважливіших статистичних параметрів тексту загалом та ідіостилу письменника зокрема. Його обчислюють як відношення обсягу словника лексем (тобто лематизованих, зведених до початкової форми слів) V до обсягу тексту N : $B = V/N$.

Для великої прози І. Франка індекс різноманітності подано в таблиці 3.

Укладачі частотного словника сучасної української художньої прози зазначають, що показник багатства лексики зменшується зі збільшенням обсягу тексту [25, т. 1, с. 15]. Наші підрахунки, представлені в таблицях 2 і 3, підтверджують такі висновки: «Перехресні стежки» за найбільшого обсягу прямої мови (45 513 слововживань) мають найнижчий індекс різноманітності лексики (0,13), а «Воа constrictor» за найменшого обсягу прямої мови (3 143 слововживань) має найвищий індекс різноманітності лексики (0,31).

Водночас укладачі припускають, що «така різниця у збільшенні кількості слів і словоформ просто є свідченням того, що пряма мова менш різноманітна лексично, ніж авторська» [25, т. 1, с. 15]. Дані з таблиці 3 засвідчують таку саму тенденцію. Це можна зауважити як на творах з майже однаковим співвідношенням прямої й авторської мови («Перехресні стежки», «Великий шум»), так і на різних творах. Зокрема, за приблизно однакового обсягу прямої мови в «Бориславі сміється» (33 330) та авторської в «Захарі Беркуті» (32 994), індекс різноманітності вищий в авторській: 0,12 та 0,17 відповідно. Ще більше відрізняється багатство словника на користь наративного тексту у «Великому шумі» (0,25) у порівнянні з прямою мовою «Для домашнього» (0,17) у той час, як їх обсяг майже однаковий: 18 822 та 18 914 відповідно.

Таблиця 3

Багатство словника у прямій та авторській мові великої прози Івана Франка

романи	пряма мова		авторська мова		разом	
	багатство	слововж.	багатство	слововж.	багатство	слововж.
«Воа constrictor», 1878	0,31	3 143	0,21	22 284	0,20	25 427
«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
«Великий шум»	0,18	18183	0,25	18822	0,18	37 005
«Петрії й Довбушки»	0,18	16 219	0,17	36 532	0,14	52 751

Отже, корпус текстів Івана Франка, що маркує пряму та авторську мову, дає змогу досліджувати їх у кількісних та якісних вимірах. Електронне автоматичне опрацювання значно полегшило їх наукові дослідження. Воно показало, що художня проза – не гомогенний стиль, а інтегральне поєднання двох субжанрів: розповідного та діалогічного. Оскільки кількісне співвідношення цих пластів лексики буде відрізнятися у кожного конкретного письменника, то воно може вважатися статистичним параметром ідіостилу.

Запропоноване у статті зіставлення обсягу прямої та авторської мови за конкретними творами апостеріорно засвідчує, що воно залежить від тематики твору: історичний роман потребує ширшого авторського наративу, а твір про сучасність має більше діалогічної інтеракції персонажів. Щоб повніше та адекватніше відобразити ці закономірності, отримані результати треба зіставити з аналогічними дослідженнями

інших авторів. У великій прозі І. Франка виявлено суттєво більшу частку прямої мови (40,6 : 59,4 %) порівняно з художньою прозою ХХ ст. (28,2 : 71,8 %). Цей факт набуває особливого значення з огляду на комплексну квантитативну параметризацію прози І. Франка. 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.

У перспективі буде цікаво поділити пряму мову на діалогічну, монологічну і внутрішній діалог. Є припущення, що в межах цих текстових піджанрів лексика буде відрізнятися. На відміну від авторської мови, орієнтованої, за винятком деяких оповідних жанрів, на писемну форму, мова персонажів відображає особливості усної розмовної мови, тобто сукупність прямої мови усіх персонажів у текстах Франкової спадщини в далекосяжній перспективі дасть змогу реконструювати розмовний стиль західного варіанту української літературної мови зламу ХІХ–ХХ ст.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. *Арцишевська А. Л.* Лінгвопрагматична типологія репрезентуючого компонента в системі форм репродукції чужої мови (на матеріалі англломовних оповідань ХХ століття): автореф. дис. на здобуття наук. ступеня канд. філол. наук : 10.02.04 / А. Л. Арцишевська; Львів. нац. ун-т ім. І. Франка. – Львів, 2001.
2. *Багмут А. Й.* Пряма мова / А. Й. Багмут // Українська мова : енциклопедія / редкол. : В. М. Русанівський та ін. – К., 2000. – С. 538.
3. *Бук С.* Повість Івана Франка «Захар Беркут» у світлі статистичної лінгвістики / С. Бук // Науковий часопис Національного педагогічного університету імені М. П. Драгоманова : збірник наукових праць. – К. : НПУ імені М. П. Драгоманова, 2010. – Вип. 6. – С. 172–176. – Серія 10. Проблеми граматики і лексикології української мови.
4. *Бук С.* Інтерпретація лексики роману Івана Франка «Великий шум» у кількісно-статистичному аспекті / С. Бук // *Studia Methodologica*. – Тернопіль : ТНПУ ім. В. Гнатюка, 2010. – Вип. 30. – С. 145–153.
5. *Бук С.* Корпус текстів Івана Франка : спроба визначення основних параметрів / С. Бук // *Прикладна лінгвістика та лінгвістичні технології : MegaLing-2006* : збірник наукових праць. – К., 2007. – С. 72–82.
6. *Бук С. Н.* Структурне анування у корпусі текстів (на прикладі прози Івана Франка) / С. Н. Бук // *Українська мова*. – 2009. – № 3. – С. 59–71.
7. *Бук С.* Роман Івана Франка «Для домашнього огнища» крізь призму частотного словника / С. Бук // *Мовознавство*. – 2011. – № 4. – С. 56–66. Препринт arXiv:1006.0153v1 [cs.CL]. – Режим доступу : <<http://arxiv.org/abs/1006.0153>>.
8. *Бук С.* Статистична структура роману І. Франка «Борислав сміється» / С. Бук // *Ученые записки Таврического национального университета им. В. И. Вернадского*. – 2010. – С. 114–118. – Серія «Филологія».
9. *Бук С.* Статистичні характеристики роману Івана Франка «Основи суспільності» (на основі частотного словника твору) / С. Бук // *Вісник : Проблеми української термінології*. – Львів : Національний університет «Львівська політехніка». – 2010. – № 676. – С. 90–93.

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» [Електронний ресурс] / С. Бук, А. Ровенчак. – [Цит. 10 березня 2011]. – Режим доступу : <<http://www.ktf.franko.lviv.ua/~andrij/science/Franko/concordance.html>>.
12. Бук С. Частотний словник повісті І. Франка «Перехресні стежки» / С. Бук, А. Ровенчак // Стежками Франкового тексту (комунікативні, стилістичні та лексичні виміру роману «Перехресні стежки»). – Львів, 2007. – С. 138–369.
13. Волошук В. І. Лінгвостильові особливості ідіолекту З. Ленца в малих епічних жанрах. автореф. дис. на здобуття наук. ступеня канд. філол. наук : 10.02.04 / В. І. Волошук; Львів. нац. ун-т ім. І. Франка. – Львів, 2004.
14. Гріпцов В. І. Структурно-семантичні та прагматичні особливості речень з прямою мовою (на матеріалі сучасної іспаномовної художньої прози) : автореф. дис. на здобуття наук. ступеня канд. філол. наук : 10.02.05 / В. І. Гріпцов; Київ. нац. ун-т ім. Т. Шевченка. – К., 2008.
15. Гут Н. В. Слова автора при прямій мові в системі художнього тексту (на матеріалі творчості Григорія та Григора Тютюнників) : автореф. дис. на здобуття наук. ступеня канд. філол. наук : 10.02.01 / Н. В. Гут; Київ. нац. ун-т ім. Т. Шевченка, Ін-т філол. – К., 2009.
16. Демська-Кульчицька О. Основи національного корпусу української мови / О. Демська-Кульчицька. – К., 2005.
17. Корпусна лінгвістика / В. А. Широков, О. В. Бугаков, Т. О. Грязнухіна та ін. – К., 2005.
18. Масенко Л. Т. Мовна характеристика (персонажів) / Л. Т. Масенко // Українська мова : Енциклопедія / редкол. : В. М. Русанівський та ін. – К., 2000. – С. 361–362.
19. Национальный корпус русского языка. – [Цит. 10 марта 2011]. Режим доступа : <<http://www.ruscorpora.ru>>.
20. Пілик В. В. Лексико-граматична структура турецького газетного тексту : автореф. дис. на здобуття наук. ступеня канд. філол. наук : 10.02.13 / В. В. Пілик; НАН України. Ін-т сходознавства ім. А. Кримського. – К., 2007.
21. Сичинава Д. В. К задаче создания корпусов русского языка. [Цит. 10 марта 2011]. Режим доступа : <<http://www.mccme.ru/ling/mitrius/article.html>>.
22. Скляренко О. Б. Авторське та персонажне мовлення як одна із структурно-композиційних домінант оповідань І. Бахман (на матеріалі збірки «Das dreibigste Jahr») / О. Б. Скляренко // Культура народів Причорномор'я. – 2007. – № 110. – Т. 2. – С. 176–178.
23. Статистичні параметри стилів / за ред. В. С. Перебийніс. – К., 1967.
24. Ціхоцький І. Л. Мовна характеристика персонажів у прозі Івана Франка : автореф. дис. на здобуття наук. ступеня канд. філол. наук / І. Л. Ціхоцький; Львів : Вид. центр ЛНУ ім. Івана Франка, 2004. – 20 с.
25. Частотний словник сучасної української художньої прози : у 2 т. / за ред. Перебийніс В. С. – К., 1981.

26. Шайкевич А. Я. Статистический словарь языка Достоевского / Шайкевич А. Я., Андриющенко В. М., Ребецкая Н. А. – М., 2003.
27. Axelsson K. Research on fiction dialogue : Problems and possible solutions // Andreas H. Jucker, Daniel Schreier, Marianne Hundt Corpora : Pragmatics and discourse : papers from the 29th International Conference on English Language Research on Computerized Corpora (ICAME 29), Ascona, Switzerland, 14–18 May 2008. Rodopi, 2009. – P. 189–201.
28. British National Corpus. – Cited 10 March 2011. – Available from : <<http://www.nat-corp.ox.ac.uk>>.
29. A Corpus of English Dialogues 1560–1760 – Uppsala universitet. – Cited 10 March 2011. – Available from : <www.engelska.uu.se/Forskning/sprakvetenskap/A_Corpus_of_English_Dialogues>
30. Český národní korpus. – Cited 10 March 2011. – Available from : <<http://ucnk.ff.cuni.cz>>.
31. Francis W. N. Brown Corpus Manual (1979) / W. N. Francis, H. Kucera. – Cited 10 March 2011. – Available from : <http://dilbilim.info/yukseklisans/Corpus%20Based/Brown_Corpus_Manual.pdf>.
32. Johansson S. English-Norwegian Parallel Corpus : Manual / S. Johansson, J. Ebeling, S. Oksefjell – Cited 10 March 2011. – Available from : <http://www.hf.uio.no/ilos/tjenester/omc/enpc/ENPCmanual.html#_Toc445194182>.
33. TEI : Text Encoding Initiative. P6 : Elements Available in All TEI Documents. – 2004. – Cited 10 March 2011. – Available from : <<http://www.tei-c.org/release/doc/tei-p4-doc/html/CO.html#COHQQ>>

Стаття надійшла до редколегії 28.01.2011

Прийнята до друку 9.02.2011

**DIRECT AND AUTHOR'S SPEECH
OF THE LONG PROSE FICTION BY IVAN FARNKO:
LINGVOSTATISTICAL STUDY IN THE CONTEXT OF CORPUS LINGUISTICS**

Solomiya BUK

*Ivan Franko National University of Lviv,
Department for General Linguistics,
1, Universytetska St., r. 343, 79000 Lviv, Ukraine,
tel.: +380 (32) 2394756, e-mail: solomija@gmail.com*

The paper presents the quantitative comparison of the direct and author's speech in the long prose fiction by Ivan Franko (which can serve as a statistical parameter of a writer individual style), in particular, the comparison with the Ukrainian prose of the 20th century is made. The work includes a brief description of the experience of marking these types of text in lexicography and corpus linguistics, as well as the calculation of richness in these lexical layers of writer's artistic heritage. The study is based on the electronic text corpus of Ivan Franko.

Key words: direct speech, author's speech, text size, tokens, text corpus, mark-up, lexical richness of vocabulary.

**ПРЯМАЯ И КОСВЕННАЯ РЕЧЬ БОЛЬШОЙ ПРОЗЫ ИВАНА ФРАНКО:
ЛИНГВОСТАТИСТИЧЕСКОЕ ИССЛЕДОВАНИЕ
В КОНТЕКСТЕ КОРПУСНОЙ ЛИНГВИСТИКИ**

Соломия БУК

*Львовский национальный университет имени Ивана Франко,
кафедра общего языкознания,
ул. Университетская, 1, ауд. 343, 79000 Львов, Украина,
тел.: +380 (32) 2394756, e-mail: solomija@gmail.com*

В статье представлено количественное сопоставление соотношений прямой и авторской речи в большой прозе И. Франко (что может служить статистическим параметром идиостиля), в частности, приведено сравнение с украинской художественной прозой XX века. Статья содержит краткое описание опыта маркировки этих типов текста в лексикографии и корпусной лингвистике, а также вычисления богатства словаря в этих пластах лексики художественного наследия писателя. Исследование выполнено на материале электронного корпуса текстов И. Франко.

Ключевые слова: прямая речь, косвенная (авторская) речь, объём текста, словоупотребление, корпус текстов, разметка корпуса, лексическое богатство словаря.