

МОВОЗНАВСТВО

УДК 811.161.2'1'324'38 (038)

КІЛЬКІСНЕ ЗІСТАВЛЕННЯ ТЕКСТІВ (НА МАТЕРІАЛІ РЕДАКЦІЙ 1884 ТА 1907 РОКІВ ПОВІСТІ ІВАНА ФРАНКА «ВОА CONSTRICTOR»)

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*

Уперше у франкознавстві підготовано кількісне зіставлення текстів повісті І. Франка «Воа constrictor» редакцій 1884 та 1907 років. Обчислено співвідношення основних квантитативних параметрів обох творів (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), і відповідно простежено, як змінилися статистичні показники ідіостилу І. Франка від 1884 до 1907 року. Виявлено, що лексика редакцій відрізняється більше, ніж на половину, тобто письменник не лише розширив твір, а й суттєво переробив його, зокрема збільшивши його діалогічність.

Ключові слова: редакція твору, корпус текстів, маркування, частотний словник, кількісні характеристики тексту, діалогічність тексту, частини мови.

Квантитативне порівняння текстів у світовій філології застосовують віддавна. За його допомогою виявляють спільну та відмінну лексику [21; 32; 40], «відстань між текстами» [1; 21], ключові слова [39; 41; 42] тощо. 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, Ф. Достоєвського, французьких символістів, польських романтиків, натуралістів та ін. Французький дослідник П. Гуірауд порівняв лексику символістів П. Корнеля, Ж. Расіна, Ш. Бодлера, С. Малларме, Л. Валері й Г. Аполлінера зі загальним частотним словником французької мови [35, s. 78, 83, 100, 102, 104]. Е. Стахурський порівняв лексику поеми А. Міцкевича «Пан Тадеуш» із іншими романтичними поемами [41], Т. Скубаланка зробила подібне дослідження зі Ю. Словацьким на тлі традиції [39], Ф. Чермак порівняв частотний словник К. Чапека із корпусом сучасних чеських текстів [40], а колектив під керівництвом А. Шайкевича зіставив частотність лексики Ф. Достоєвського із корпусами російської мови [32].

Як зазначив польський мовознавець Е. Стахурський, серед складнощів застосування цього методу «лексичного просіювання» (*metoda przesiewania leksykalnego*) є великі затрати часу та праці, до того ж цей метод «є дуже нудний, і, можливо, тому дослідники застосовують його неохоче» [41, с. 18]. У пізнішій своїй праці він підкреслив його необхідність: «Якщо ми хочемо точніше визначити рівень лексичної подібності творів, мусимо виразити його кількісно» [42, с. 19].

Мета нашого дослідження – виявити, наскільки редакції «*Voа constrictor*» відрізняються між собою у кількісному аспекті, як співвідносяться статистичні характеристики обох текстів, а також простежити, як змінилися квантитативні показники ідіостилю І. Франка від 1884 до 1907 року.

Повість «*Voа constrictor*» І. Франка належить до творів бориславського циклу, який отримав цікавий опис дослідників здебільшого з погляду екзистенційної проблематики та психологічного опису героїв [14], естетичної концепції дійсності [25], жанрової специфіки [20; 26], реконструкції мовлення соціальних груп [31] і соціального статусу особи [15], парадигматичних лексичних відношень [19] тощо. Предметом зіставлення двох редакцій повісті були текстуально-стилістичні відмінності [17, с. 453–469], характеристика бориславського й дрогибицького локального колориту [16]. Кількісну інтерпретацію зіставлення текстів двох редакцій повісті у статті зроблено вперше.

Історія написання повісті Івана Франка «*Voа constrictor*». Повість «*Voа constrictor*» опубліковано вперше 1878 року в журналі «Громадський друг» (№ 2), який видавали М. Павлик та І. Франко (і тому він був конфіскований цензурою), а також у збірниках «Дзвін» та «Молот», що були продовженням цього журналу. Пізніше, 1884 року, твір вийшов окремим виданням [27] накладом редакції журналу «Зоря», з яким тоді співпрацював автор. Детальне текстологічне порівняння видань 1878 та 1884 років, виконано у коментарях 50-томного зібрання творів, дало підстави стверджувати, що «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» [17, с. 453]. Проте, незважаючи на усі відмінності, ці два тексти вважаються різними редакціями-видозмінами того самого твору, який розміщено у 14 томі [28].

Протягом 1905–1907 років письменник докорінно переробив повість, яка вийшла окремою книгою [30]. Цей текст умовно називається «другою редакцією» твору. Редакційна колегія 50-томника зазначає, що «І. Франко значно розширив твір. Докладніше змальовано дитинство Германа, введено образ Вольфа, з яким Герман зайнявся гендлярством, зовсім інакше подано образ Рифки та її сім'ї, одруження Германа; їх син тут зовється не Готліб, а Дувідко, відсутня сцена, де Готліб душить батька. Знято також розповідь про Івана Півторака та його сім'ю. Інакше й закінчення повісті – Герман гине від випадкового вибуху у штольні» [18, с. 507]. Тому цю другу редакцію передруковано як самостійний твір у 22 томі за вказаним прижиттєвим виданням [29].

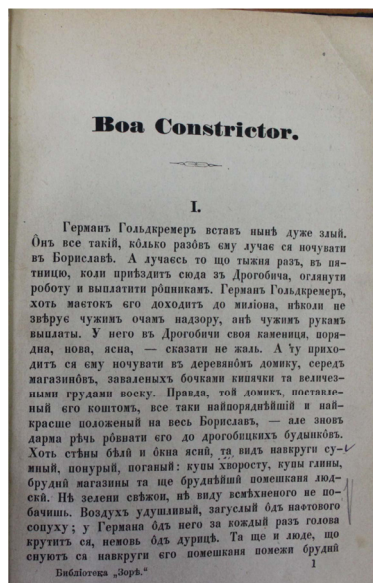


Рис. 1. Титул і перша сторінка повісті І. Франка «Boa constrictor» 1884 року [27]

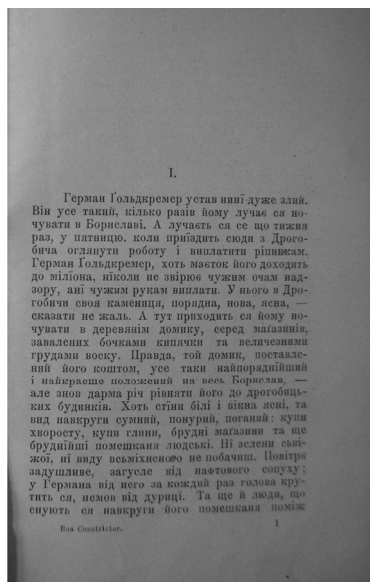


Рис. 2. Титул і перша сторінка повісті І. Франка «Boa constrictor» 1907 року [30]

Оскільки корпусний підхід до текстологічного аналізу стає все більш поширеним в останні десятиліття, зіставлення двох редакцій твору «Boa constrictor» виконано на основі електронного корпусу їх текстів (КТ). У термінології корпусної лінгвістики

найважливішими обов'язковими атрибутами КТ є машиночитаність та маркування (анотація), тобто присвоєння мовним одиницям тексту лінгвістичної інформації [36, s. 4]. Проект створення КТ І. Франка, який реалізується у Львівському університеті [5], розпочався із великої прози письменника, що на сьогодні завдяки морфологічному маркуванню уможливило лексичну та граматичну інвентаризацію творів, а також їх квантитативний опис [2; 3; 6–8; 10–12]. В основу КТ редакцій 1884 та 1907 років лягли видання твору відповідно 1884 року [27], передруковане у п'ятдесятитомному зібранні творів письменника [28] та видання 1907 року [30], передруковане у цьому ж зібранні [29].

За методикою та принципами, спільними для усіх творів І. Франка, на основі КТ [34] було укладено частотні словники (ЧС) обох редакцій повісті «*Boa constrictor*», які й стали матеріалом порівняння у статті. Лематизацію (зведення слововживань до початкової форми) виконано напівавтоматичним способом із використанням комп'ютерного лематизатора, що містить словник словоформ згаданих творів І. Франка (до яких укладено ЧС) із відповідними до них лемами.

Таблиця 1

Частотний словник повісті І. Франка « <i>Boa constrictor</i> » 1884 (зразок)		
Ранг	Слово (лема)	Частота
1	ВІН	836
2	І	820
3	НЕ	548
4	В	454
5	НА	422
6	ТОЙ	372
7	З	318
8	ГЕРМАН (ім'я)	308
9	ЩО (займ.)	268
10	БУТИ	263
11	ДО	262
12	А(спол.)	258
13	ЙОГО (присв. ч. р.)	245
14	ЩО (спол.)	194
15	ВЕСЬ	192

Таблиця 2

Частотний словник повісті І. Франка « <i>Boa constrictor</i> » 1907 (зразок)		
Ранг	Слово (лема)	Частота
1	І	1037
2	ВІН	868
3	НЕ	809
4	НА	613
5	В	493
6	ТОЙ	422
7	З	421
8	БУТИ	417
9	А (спол.)	400
10	ДО	376
11	ЩО (займ.)	369
12	ТА (спол.)	364
13	ГЕРМАН (ім'я)	319
14	ВОНИ	314
15	У	310

Окремим словом вважалася послідовність літер (апостроф і дефіс розглядають як літеру) між двома пробілами чи розділовими знаками, тому складні числівники розглянуто як різні слова. У тексті повісті методом контекстуального аналізу розмежовано лексичну та лексико-граматичну омонімію (*то* (спол., частка), *собі* (займ., частка), *так* (частка, присл.), *часом* (ім., присл.), *корч* (рослина, спазм), *власне* (частка, прикм.) тощо), лексеми із синонімічними морфемами подано у різних словникових статтях (*люди і люде*, *перстень і перстінь*). Зазвичай, у ЧС фонетичні варіанти зводять до одного реєстрового слова, проте для зіставлення ЧС їх краще мати окремо, оскільки

один із фонетичних варіантів слова потенційно може трапитися в одному тексті, а інший – в іншому.

У науковій літературі існують різні підходи до порівняння ЧС, наприклад, В. Перебийніс запропонувала поняття «нульовий стиль» [21], М. Арапов зі співавторами висунув теорію «кількісної відстані між словниками» [1]. Для зручності порівняння ЧС обох редакцій повісті І. Франка комп'ютерна програма звела слова обох текстів у єдиний список із трьома подальшими колонками: у першій зазначено частотність слова у редакції твору 1884 року, у другій – у редакції 1907 року, у третій – сумарна частотність (див. табл. 3).

Таблиця 3

Зведений частотний словник повісті І. Франка «*Voа constrictor*» редакцій 1884 та 1907 років

Ранг	Слово (лема)	Частота 1884	Частота 1907	Сумарна частотність
1	І	820	1037	1857
2	ВІН	836	868	1704
3	НЕ	548	809	1357
4	НА	422	613	1035
5	В	454	493	947
6	ТОЙ	372	422	794
7	З	318	421	739
8	БУТИ	263	417	680
9	А (спол.)	258	400	658
10	ДО	262	376	638

Порівняння кількісних характеристик тексту редакцій 1884 й 1907 років повісті Івана Франка «*Voа constrictor*». У науковій літературі основними кількісними характеристиками твору називають обсяг тексту і словника, індекс різноманітності (багатство словника), індекс винятковості тексту й словника, індекс концентрації тексту й словника і т. ін. Статистична структура тексту вважається важливою складовою опису ідіостилю письменника. За даними ЧС обох редакцій повісті обчислено їх квантитативні показники (див. табл. 4).

Обсяг тексту повісті І. Франка 1884 року – 25 427 слововживань, а обсяг тексту 1907 року – 34 215, відповідно, друга редакція на 8 788 слововживань довша за першу, а це більше, ніж на третину (на 34,6%). Кількість слів, спільних для обох текстів, – лише 3 252 (це становить 64,9% від першої редакції і 52,1% від другої редакції). Такі дані можна вважати свідченням того, що в основу тексту твору 1907 року І. Франко не просто взяв текст 1884 року, доповнивши його, але суттєво переробив, видозмінивши лексику фактично на половину (47,9% слів у ньому нових). Серед першої сотні спільних лексем найчастотнішими, оминаючи службову лексику та займенники (ці слова високочастотні незалежно від тексту), є очікувано ім'я головного героя *Герман* (627), місто, де розгортаються події, *Борислав* (79), а також іменники на позначення органів та частин тіла *око* (157), *рука* (130), *голова* (121), *лице* (84); дієслова на позначення мислення, чуття, змоги *знати* (221), *могти* (148), *чутти* (117), *мати* (116), *бачити* (106), *сказати* (76);

часові поняття *раз* ім. (213), *тепер* (133), *другий* (122), *час* (115), *день* (98), *життя* (90); лексеми, пов'язані з роботою та її оплатою *яма* (152), *гроші* (112), *стояти* (105), *тільки* (98), *хвиля* (95), *почати*, *швидко* (82).

Таблиця 4

Кількісні характеристики редакцій 1884 й 1907 років повісті І. Франка «*Воа constrictor*»

Кількісні характеристики тексту	1884	1907
Обсяг тексту, слововживань	25 427	34 215
Обсяг словника, лексем	5 007	6 239
Багатство словника ¹	0,197	0,182
Середня повторюваність слова у тексті ²	5,08	5,48
Нарах <i>legomena</i> ³	2 802	3 432
Індекс винятковості тексту ⁴	0,11	0,10
Індекс винятковості словника ⁵	0,56	0,55
Високочастотна лексика ($f \geq 10$) у тексті	16 178	22 850
Високочастотна лексика ($f \geq 10$) у словнику	376	492
Індекс концентрації тексту ⁶	0,636	0,668
Індекс концентрації словника ⁷	0,075	0,079

1 809 слів є оригінальними в редакції 1884 року, найбільше серед них вживаються *Матій* (46), *к* (16), *враження* (14), *огромний* (10), *кибель*, *нинішній*, *Півторак* (9). Лише у редакції 1907 року функціонує 2 987 слів, між якими власні назви *Вольф* (72), *Дувідко* (39), *Мендель* (29), *Лютувиська* (27), а також загальні назви *мовити* (58), *лише* (37), *бартишівник* (34), *хоч* (33), *бац* ім. (25), *аби* (24), *гарний*, *знайти* (22), *колиба*, *яр* (21), *потім*, *ярмарок*, *восковий* (20) – найчастотніші.

Навіть серед найчастотніших слів уже помітно, як І. Франко ретельно працював над тим, щоби мова твору була сучасною: прийменник *к* з першої редакції послідовно виправив на *до* у другій редакції: «підійшли к пальмам» (1884) на «підійшли до пальм» (1907), «Ближче к передньому планові картини» (1884) на «Ближче до переднього плану малюнка» (1907); *огромний* на *великий* / *здоровенний*: «Яким же способом доробився Герман Гольдкремер такого *огромного* маєтку?» (1884) на «Яким же способом доробився

¹ Багатство словника обчислюють як відношення обсягу словника лексем до обсягу тексту.

² Середню повторюваність слова в тексті обчислюють як відношення обсягу тексту до обсягу словника лексем.

³ Нарах *legomena* – це кількість слів, що трапилися у тексті один раз.

⁴ Індекс винятковості тексту обчислюють як відношення кількості *hарах legomena* до обсягу тексту.

⁵ Індекс винятковості словника обчислюють як відношення кількості *hарах legomena* до обсягу словника лексем.

⁶ Індекс концентрації тексту обчислюють як відношення кількості високочастотних слів до обсягу тексту.

⁷ Індекс концентрації словника обчислюють як відношення кількості високочастотних слів до обсягу словника лексем.

Герман Гольдкремер такого *великого* маєтку?» (1907); «...з *огромною* булавою...» (1884) на «...з *здоровенною* булавою...» (1907). Також порівняймо: «Була се досить *хороша* і вірно *списана* картина тропічної, індійської околиці» (1884) на «Була се досить *гарна* і вірно *змальована* картина тропічної, індійської околиці» (1907). Отже, на позначення тих самих понять І. Франко використовував різні лексеми у виданнях 1884 і 1907 років, що свідчить про еволюцію його стилю: *к – до, враження – вражіння, огромный – великий / здоровенний, сьогодні – сьогодні, хороший – гарний, способности – здібності, найти – знайти, опісля – потім*, і т. д. відповідно.

Загалом, словник обох редакцій становить 8 048 різних слів, а це більше, ніж у цілому словнику мови Т. Шевченка (6 116 слів [13])!

Обсяг словника лексем зростає неспівмірно зі зростанням довжини тексту: чим довший текст, тим менше у ньому з'являється нових слів [22, с. 143]. Це підтверджується отриманими показниками для текстів «*Voа constrictor*» різних розмірів. Як видно з таблиці 3, у довшому тексті багатство словника менше – 0,182 (проти 0,197 в коротшому), відповідно більша середня повторюваність слів у тексті – 5,48 (проти 5,08 в коротшому), нижчий індекс винятковості тексту (0,10 проти 0,11) й словника (0,55 проти 0,56), тому вищий індекс концентрації тексту (0,668 проти 0,636) й словника (0,079 проти 0,075). Детальний аналіз порівняння параметрів, залежних від довжини тексту (індекси різноманітності, винятковості, концентрації) буде предметом окремого лінгвостатистичного дослідження.

Проте величина середньої повторюваності дуже узагальнена, адже кількість слів, вжитих у творі 1884 року один раз (*hаrах legomenа, V₁*) – 2 802, а 1907 року – 3 432 тобто в обох випадках вони займають більше, ніж половину словника (55,96% та 55,00% відповідно). Саме в них криється основне багатство лексики письменника.

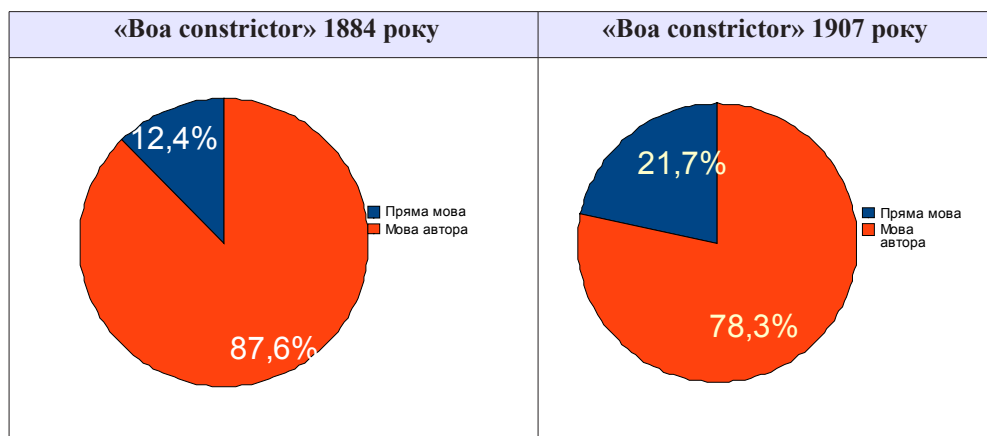
Рівень діалогічності та описовості у повісті. Текст художнього твору становлять здебільшого два головні різновиди мовлення: мова автора і пряма мова. В основі прямої мови лежить інтенція письменника передати слова й думки персонажа від його імені, з цією метою він часто стилізував пряму мову, використовуючи у ній діалектизми, професіоналізми, тобто намагався індивідуалізувати її. В основі ж мови автора лежить опис, бо письменник від свого імені (чи від імені оповідача) описав дію, обставини, думки та вчинки персонажа. Отже, розмежування цих двох вимірів твору дуже важливе для вивчення авторського ідіолекту, тому в КТ пряма та авторська мова мають відповідне маркування, і частотність уживань лексичних одиниць підраховано окремо для авторської та прямої мови. Під рівнем діалогічності розуміємо кількісне співвідношення обсягу прямої мови (монологів, діалогів, полілогів) до загального обсягу твору. Під рівнем описовості маємо на увазі кількісну реляцію обсягу авторської мови до загального обсягу твору. Графічно співвідношення авторської та прямої мови зображено у таблиці 5 та на рисунку 3.

У першій редакції твору відношення прямої до авторської мови становить 12,4% : 87,6%, а в другій – 21,7% : 87,3% (див. табл. 4), тобто рівень діалогічності у тексті 1907 року зріс майже вдвічі. Отже, обсяг другої редакції повісті розширився значною мірою завдяки збільшенню кількості діалогів.

Таблиця 5

**Кількісне співвідношення авторської та прямої мови
у повісті І. Франка «Воа constrictor» редакцій 1884 та 1907 років**

	«Воа constrictor», 1884			«Воа constrictor», 1907		
	<i>пряма</i>	<i>авторська</i>	<i>разом</i>	<i>пряма</i>	<i>авторська</i>	<i>разом</i>
Обсяг тексту,	3 142	22 285	25 427	7 435	26 780	34 215
слововживань	(12,4%)	(87,6%)	(100%)	(21,7%)	(78,3%)	(100%)



*Рис. 3. Співвідношення авторської і прямої мови
у повісті І. Франка «Воа constrictor» редакцій 1884 і 1907 років*

Частин мови та їх співвідношення у повісті Івана Франка «Воа constrictor» редакцій 1884 та 1907 року. Квантитативні показники частин мови у тексті дають багату інформацію про індивідуальний стиль письменника, адже якщо лексику автор підбирає свідомо, то граматичні категорії, в яких він уживає слова, належать до підсвідомої площини. Окрім того, співвідношення частин мови та граматичних форм у тексті становлять його статистичну структуру на морфологічному рівні. На цьому наголошують В. Пербийніс [23], В. Тищенко [24], М. Рушковський [37], Т. Скубалянка [38]. В аналізі великої прози І. Франка ми дотримувалися класичного підходу до виділення частин мови (див. табл. 6).

Частини мови у текстах обох редакцій функціонують подібно, і тому мають близькі кількісні показники. Незначно відрізняються лише службові частини мови (23,30% у 1884 році і 25,51% у 1907 році). Натомість у редакції 1907 року приблизно на один відсоток менше займенників (13,92%), прикметників (8,25%), прислівників (8,76), ніж у першій редакції (14,54%, 9,35% і 9,19% відповідно).

Таблиця 6

**Розподіл лексем за частинами мови
повісті І. Франка «Воа constrictor» у редакціях 1884 і 1907 років**

частини мови	слів у тексті		слів у словнику	
	Редакція 1884	Редакція 1907	Редакція 1884	Редакція 1907
службові ¹	23,30%	25,51%	3,75%	3,43%
іменники	24,64%	24,38%	31,52%	34,43%
дієслова	17,61%	17,82%	34,45%	33,58%
займенники	14,54%	13,92%	1,44%	1,01%
прикметники	9,35%	8,25%	18,95%	17,94%
прислівники	9,19%	8,76%	8,79%	8,53%
числівники	1,34%	1,35%	0,96%	1,04%
фрагменти	0,03%	0,01%	0,14%	0,05%
	100%	100%	100%	100%

У словнику другої редакції є на три відсотки більше іменників (34,43%), ніж у словнику першої (31,52%), але на один відсоток менше дієслів (34,45%) та прикметників (18,95), ніж у першому тексті (33,58% і 17,94% відповідно).

За поданими вище даними можна обчислити кількісні реляції між частинами мови, що також є об'єктивною характеристикою тексту, зокрема індекси епітетизації (відношення кількості іменників до кількості прикметників) [33; 37], дієслівних означень (відношення кількості прислівників до кількості дієслів) та ступінь номінальності (відношення кількості іменників до кількості дієслів) [37, с. 50]. За аналогічною методикою проаналізовано співвідношення цих основних морфологічних класів шести творів великої прози І. Франка («Борислав сміється», «Захар Беркут», «Для домашнього огнища», «Основи суспільності», «Перехресні стежки», «Великий шум») [9], тому цікаво отримані результати порівняти із досліджуваною повістю (див. табл. 7).

Таблиця 7

**Кількісні реляції частин мови повісті І. Франка «Воа constrictor»
у редакціях 1884 і 1907 років порівняно з великою прозою І. Франка²**

	Редакція 1884	Редакція 1907	Проза І. Франка
Індекс епітетизації	2,64	2,95	3,13
Індекс дієслівних означень	0,52	0,49	0,46
Ступінь номінальності	1,39	1,37	1,30

Індекс епітетизації, як впливає з його дефініції, вказує, скільки іменників у тексті припадає на один прикметник, тобто, чим показник більший, тим менше прикметників

¹ Сполучники, прийменники, частки, вигуки.

² Дані подано із [9].

припадає на один іменник. У першій редакції «*Voа constrictor*» цей індекс найменший (2,64), а це означає, що у ньому використано найбільше епітетів (1 епітет на 2,64 іменника). Загалом, у відсотковому відношенні повість має суттєво більше епітетів-означень на іменник, ніж у середньому інші твори І. Франка великої прози (1 епітет на 3,13 іменника). **Індекс дієслівних означень** прямо пропорційно виявляє, скільки прислівників у тексті випадає на одне дієслово. Знову редакція 1884 року має найбагатше відсоткове відношення: 0,52 прислівника на 1 дієслово, тоді як у редакції 1907 року та в середньому у прозі І. Франка ці показники відповідно 0,49 та 0,46. Аналогічно **ступінь номінальності** означає кількість іменників на одне дієслово. У першому тексті твору є 1,39 іменника на одне дієслово, у другому – 1,37, загалом у прозі – 1,3.

Отже, кількісні реляції частин мови повісті І. Франка «*Voа constrictor*» свідчать, що редакція повісті 1884 року має відсотково найбільше епітетів-означень, дієслівних означень, а також найвищий ступінь номінальності як порівняно з редакцією твору 1907 року, так і порівняно зі середніми показниками великої прози І. Франка.

Оскільки перша редакція повісті «*Voа constrictor*», опублікована 1884, належить до перших творів великої форми І. Франка, а друга, опублікована 1907 року, – до останніх, то їх зіставлення виявило еволюцію творчої манери письма автора. Для аналізованих текстів властиві різні кількісні характеристики. Ці відмінності зумовлені не лише різним обсягом текстів (від обсягу, наприклад, залежать багатство словника, індекс винятковості словника й тексту, які вищі у тексті 1884 року, оскільки його обсяг менший), а й різним стилем написання (співвідношення частин мови, реляції між ними). Перша редакція повісті має відсотково найбільше епітетів-означень, дієслівних означень, а також найвищий ступінь номінальності як у порівнянні з редакцією твору 1907 року, так і порівняно зі середніми показниками великої прози І. Франка. Як свідчать реляції між частинами мови у текстах 1884 та 1907 років, старший І. Франко надавав перевагу службовим частинам мови більше (25,51%), ніж молодий (23,30%), натомість молодий використовував дещо більше прикметників (9,35%), ніж старший (8,25%); відсоткове ж співвідношення усіх інших частин мови майже однакове.

Лексика порівнюваних текстів відрізняється більше, ніж на половину. Зіставлення ЧС обох редакцій виявило, що письменник не лише розширив твір, а й суттєво переробив його, оскільки друга редакція має тільки 52 % лексики спільної із першою редакцією, 48 % у ній становлять нові слова, яких не було у першій. Цікаво зауважити, що на позначення тих самих понять Франко використовував різні лексеми у виданнях 1884 і 1907 років, що свідчить про еволюцію його стилю: *к – до, враження – вражіння, огромний – великий / здоровенний, сьогодні – сьогодні, хороший – гарний, способності – здібності, найти – знайти, опісля – потім*, і т. д. відповідно.

У другій редакції повісті «*Voа constrictor*» рівень діалогічності зріс майже удвічі, що виражено у збільшенні обсягу прямої мови від 3 142 до 7 435 слововживань, а це становить відповідно 12,4% і 21,7% тексту.

У ширшій перспективі заплановано провести комплексне статистичне зіставлення текстів великої прози І. Франка [4], а саме з'ясувати лексику, спільну для усіх творів, скільки кожен з текстів має однакових лексем один з одним, у яких зіставлюваних

парах текстів є найбільше спільних лексем, а значить: які твори найбільш подібні з лексичного погляду. У такому дослідженні виявлені квантитативні особливості обох редакцій повісті – перший крок до масштабніших висновків.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. *Арапов М. В.* Сравнение частотных словарей / М. В. Арапов, Л. И. Тер-Гаспрян, М. М. Херц // Научно-техническая информация. – 1978. – №4. – С. 20–29. – (Серия 2. Информационные процессы и системы).
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 словника / С. Бук // Мовознавство. – 2011. – № 4. – С. 56–66.
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) // Вісник : Проблеми української термінології. – 2010. – № 676. – С. 90–93.
9. *Бук С.* Частини мови у словнику і тексті Івана Франка (на матеріалі великої прози) / С. Бук // Лінгвістичні студії. – 2011. – Вип. 22. – С. 62–66.
10. *Бук С.* Частотний словник роману Івана Франка «Воа constrictor» / С. Бук // Стежками Франкового тексту (комунікативні, лінгвосеміотичні, когнітивні та лінгвостатистичні виміри прози) / Ф. С. Бацевич (наук. ред.), С. Н. Бук, Л. М. Процак, Л. Ю. Сваричевська, О. В. Ясіновська. – Львів : Видавничий центр ЛНУ імені Івана Франка, 2012. (у друці).
11. *Бук С.* Лінгвостатистичний опис «Не спитавши броду» Івана Франка / С. Бук // Вісник Львівського університету. – Львів : Львівський національний університет імені Івана Франка, 2011. – Вип. 55. – С. 230–241. – (Серія філологічна : Франкознавство).
12. *Бук С.* Частотний словник повісті І. Франка «Перехресні стежки» / С. Бук // Стежками Франкового тексту (комунікативні, стилістичні та лексичні виміри роману «Перехресні стежки»). – Львів, 2007. – С. 138–369.
13. *Ващенко В. С.* Словник мови Шевченка / В. С. Ващенко. – К. : Наукова думка, 1964. – Т. 1. – 484 с., Т. 2. – 566 с.
14. *Водяна Л. В.* Екзистенційна проблематика бориславського циклу творів І. Франка : автореф. дис... канд. філол. наук : 10.01.01 / Л. В. Водяна; Кіровоград. держ. пед. ун-т ім. В. Винниченка. – Кіровоград, 2008. – 20 с.
15. *Гавриш М. М.* Комунікативні засоби вираження соціального статусу особи в художній

- прозі Івана Франка : автореф. дис... канд. філол. наук : 10.02.01 / Гавриш М. М. ; Львівський національний ун-т ім. Івана Франка. – Львів, 2009. – 18 с.
16. *Гузар З.* До питання про локальний колорит у прозі І.Франка (Борислав і Дрогобич у двох редакціях повісті «*Voа constrictor*») / З. Гузар // Українське літературознавство : міжвузівський республіканський збірник. – Львів : В-во Львів. університету, 1972. – Вип. 17. – С. 106–111.
 17. Коментарі // Франко І. Зібр. творів : у 50 т. Т. 14 / І.Франко. – К. : Наукова думка, 1979. – С. 453–470.
 18. Коментарі // Франко І. Зібр. творів : у 50 т. Т. 22 / І.Франко. – К. : Наукова думка, 1979. – С. 506–507.
 19. *Ощипко И. И.* Лексическая синонимика в бориславских повестях и рассказах Ивана Франко / И. И. Ощипко : автореф. дисс. ... канд. филол. наук. – Львов, 1954. – 18 с.
 20. *Пастух Т.* Роман у системі прозових жанрів Івана Франка / Т. Пастух // Українське літературознавство : міжвузівський республіканський збірник. – Львів : В-во Львів. університету, 1996. – Вип. 62. – С. 100–108.
 21. *Перебийнос В. С.* Методы и уровни моделирования нулевого стиля / В. С. Перебийнос // Вопросы статистической стилистики. – К. : Наука, 1974. – С. 16–35.
 22. *Перебийніс В. С.* Частотні словники та їх використання / В. С. Перебийніс, М. П. Муравицька, Н. П. Дарчук. – К. : Наукова думка, 1985. – С. 143.
 23. Статистичні параметри стилів / за ред. В. С. Перебийніс. – К. : Наукова думка, 1967. – 260 с.
 24. *Тищенко В.* Частота частин мови в різних функціональних стилях сучасної української мови / В. Тищенко // Питання структурної лексикології – К. : Наукова думка, 1970. – С. 215–224.
 25. *Ткачук М. П.* Естетична концепція дійсності в Бориславському циклі І. Я. Франка : монографічне дослідження / М. П. Ткачук. – Тернопіль : Тернопільський педагогічний ін-т, 1992. – 123 с.
 26. *Ткачук М. П.* Жанрова структура прози Івана Франка (Бориславський цикл та романи з життя інтелігенції) / М. П. Ткачук : дис... д-ра філол. наук : 10.01.01 / Тернопільський держ. педагогічний ун-т. – Тернопіль, 1997. – 394 с.
 27. *Франко И.* *Voа constrictor* : Повѣсть Ивана Франка / И. Франко. – Львѡвъ : Накладомъ редакціи «Зорѣ» ; зъ печатнѣ Ставропігійского Института, 1884. – 114 с.
 28. *Франко І.* *Voа constrictor* // **Франко І. Зібр. творів : у 50 т. Т. 14 / І. Франко.** – К. : Наукова думка, 1979. – С. 370–441.
 29. *Франко І.* *Voа constrictor* // **Франко І. Зібр. творів : у 50 т. Т. 22 / І. Франко.** – К. : Наукова думка, 1979. – С. 109–207.
 30. *Франко І.* *Voа constrictor* : Нове, перероблене виданє / Іван Франко. – Львів : Накладом Українсько Руської видавничої спілки, 1907. – 150 с.
 31. *Ціхоцький І.* Мовні профілі «Бориславських студій» (Виробнича лексика у франковому ідіолекті) / І. Ціхоцький // Українське літературознавство. – 2006. – Вип. 68. – С. 237–246.
 32. *Шайкевич А. Я.* Статистический словарь языка Достоевского / А. Я. Шайкевич, В. М. Андрущенко, Н. А. Ребецкая. – М. : Языки славянской культуры 2003. – 880 с.
 33. *Buk S.* The Epithetization Index in a Work of Fiction (on the basis of the Text Corpus of Ivan Franko's Long Prose Fiction) / S. Buk // Practical Applications of Linguistic Research 2012. – Access mode: <http://palr.vipserv.org/conference/articles/sbuk.pdf> >
 34. *Čermák F.* New generation corpus-based frequency dictionaries : The case of Czech // International Journal of Corpus Linguistics. – 10. – 2005. – P. 453–467.
 35. *Guiraud P.* Les caractères statistiques du vocabulaire. Essai de methodologie / P. Guiraud. – Paris : Presses universitaires de France, 1954. – 116 p.

36. *McEnery T.* Corpus-Based Language Studies. An advanced resource book / T. McEnery, R. Xiao, Y. Tono. – London; New York : Routledge, 2006. – 408 p.
37. *Ruszkowski M.* Wskaźnik epitetyzacji w badaniach stylistycznych / M. Ruszkowski // *Respectus Philologicus*. – 2004. – № 5(10). – S. 48–53.
38. *Skubalanka T.* Założenia gramatyki stylistycznej / T. Skubalanka // *Типы описов грамматических языка*. – Wrocław, 1986. – S. 141–155.
39. *Skubalanka T.* Słownictwo poezji miłosnej Juliusza Słowackiego na tle tradycji / T. Skubalanka. – Toruń, 1966.
40. *Slovník Karla Čapka* / Hl. ed. František Čermák. – Praha : Nakladatelství Lidové noviny ; Ústav Českého národního korpusu, 2007. – 715 s.
41. *Stachurski E.* Słownictwo «Pana Tadeusza» Adama Mickiewicza na tle tekstów innych poematów romantycznych / E. Stachurski. – Kraków : Wydawnictwo Naukowe Akademii Pedagogicznej, 2005. – 262 s.
42. *Stachurski E.* Wyrazy-klucze Miskiewiczowskiej eposy / E. Stachurski // *Język polski LXXIX*. – 1999. – S. 15–29.

Стаття надійшла до редакції 05.04.2012

Прийнята до друку 20.04.2012

QUANTITATIVE COMPARISON OF TEXTS (ON THE MATERIAL OF THE 1884 AND 1907 EDITIONS OF THE NOVEL «BOA CONSTRICTOR» BY IVAN FRANKO)

Solomiya BUK

*The Ivan Franko National University of L'viv,
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 quantitative text comparison of the novel «Boa constrictor» by Ivan Franko from 1884 and 1907. Such an analysis is made for the first time. The following parameters were obtained for both editions: 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. These characteristics show the changes of I. Franko style from 1884 to 1907. The vocabulary of texts differ more than by a half, it means the writer not only expanded the text but also significantly reworked it, in particular increased the level of its dialogs.

Key words: edition of text, text corpus, corpus mark-up, frequency dictionary, quantitative characteristics of text, level of its dialogs, parts of speech.

**КОЛИЧЕСТВЕННОЕ СОПОСТАВЛЕНИЕ ТЕКСТОВ
(НА МАТЕРИАЛЕ РЕДАКЦИЙ 1884 И 1907 ГОДОВ
ПОВЕСТИ ИВАНА ФРАНКО «VOA CONSTRICTOR»)****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*

Впервые в франковедении сделано количественное сравнение повести Ивана Франко «*Voa constrictor*» редакций 1884 и 1907 годов. Вычислено соотношение главных количественных параметров обоих текстов (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), а тем самым прослежено изменение статистических показателей идиостиля И. Франко с 1884 по 1907. Выявлено, что лексика редакций отличается более, чем на половину, то есть писатель не только расширил произведение, но и существенно переработал его, в частности увеличил его диалогичность.

Ключевые слова: редакция произведения, корпус текстов, разметка, частотный словарь, количественные характеристики текста, диалогичность текста, части речи.