

6. Гоголь Н.В. Вий. Миргород (Собрание сочинений: в 6-ти т.) / Николай Васильевич Гоголь. – М.: ГиХЛ, 1959. – . – Т.2. – 351с.
7. Голобуцький В. О. Запорізька Січ в останні часи свого існування (1734-1775pp.) // В. О. Голобуцький. – Дніпропетровськ: Січ, 2004. – 421, [323-329] с.
8. Данилевский Г. П. Предисловие к 6-му изданию (от автора) (Сочинения: в 9-ти т.) / Григорий Петрович Данилевский. – С.-Петербург: Издание А.Ф.Маркса, 1890. – Т.1. – 1890.- 208с.
9. Данилевский Г. П. Примечания к роману «Мирович» (Сочинения: в 24-х т.) / Григорий Петрович Данилевский. – С.-Петербург: Издание А.Ф.Маркса, 1901. – Т.11. – 1901.- 137, [129] с.
10. Данилевский Г. П. Уманская резня. (Последние запорожцы). (1768–1775 г.) Историческая повесть (Сочинения: в 24-х т.) / Григорий Петрович Данилевский. – С.-Петербург: Издание А.Ф.Маркса, 1901. – Т.12. – 1901.- 203, [112-201] с.
11. Зеленый Н. В. Город Барвенково. История// Николай Зеленый. Режим доступа: <http://barvincom.com.ua/history2.htm>
12. И.И. Иванова. Национальная ментальность и национальная философия. Кантовские чтения в КРСУ (22 апреля 2004 г.); Общечеловеческое и национальное в философии: II международная научно-практическая конференция КРСУ (27-28 мая 2004 г.). Материалы выступлений / Под общ.ред. И.И. Ивановой. — Бишкек, 2004. — С.285-298. Режим доступа: http://anthropology.ru/ru/texts/ivanova_ii/commnat02_26.html
13. Кейда Ф.Ф. Художня інтерпретація гайдамачини в українській літературі XIX - XX століть: автореф. дис. д-ра філол. наук: 10.01.01«Українська література» / Кейда Ф.Ф. - Київ. нац. ун-т ім. Т.Шевченка, 2001. — 36 с.
14. Костомаров Н. Материалы для истории Коливищины или резни 1768 г. // Николай Костомаров// Киевская старина. – 1882. - № 3. Режим доступа: <http://www.vostlit.narod.ru/Texts/Dokumenty/Ukraine/XVIII/1760-1780/Koliivshina/Dok1/text.htm>
15. Малкина В.Я. Поэтика исторического романа. Проблема инварианта и типология жанра: на материале русской литературы XIX - начала XX века: автореф. дис. на соис. науч. степени канд. фил. наук. 10.01.01 «Русская литература» / Виктория Яковлевна Малкина. - М., 2001. - 19, [6] с.
16. Мірчук П. Коліївщина.Гайдамацьке повстання 1768 р.// Петро Мірчук. Режим доступа:<http://exlibris.uaweb.org/hajdamaky/index.html>
17. Мордовцев Д.Л. Гайдамачина. Исторична монографія / Данило Мордовцев. – Дніпропетровськ : «Січ», 2004. – 359 с.
18. Петрунина Н. Орест Сомов и его проза / Наталья Петрунина. – 1984. Режим доступа:http://az.lib.ru/s/somow_o_m/text_0220.shtml
19. Половцев А.А. Данилевский Г.П. Русский биографический словарь в 25 т. / под ред. А. А. Половцова. - СПб.: Изд. Русского исторического общества 1896—1918, Т.8. – 1898. - 870с.
20. Романова Т.В. Данилевский Г.П. Русские писатели 1800-1917. Биографический словарь / [гл. редактор П. А. Николаев]. - М.: "Просвещение", 1990. - . – (Серия биогр. словарей: Русские писатели. 11-20вв.). Т. 2, - 654с.
21. Рудич Ф.М. Політологія: Підручник / Фелікс Михайлович Рудич. – К.: Либідь. 2004. – 480с.
22. 100 найвідоміших образів української міфології / [Завадська В.,Музиченко Я., Таланчук О., Шапак О.]. наук. ред. Олена Таланчук. –К.: Орфей, 2002. – (100 найвідоміших). – 448, [] с.
23. Томпсон Єва М. Трубадури імперії. Російська література і колоніалізм/ Єва М. Томпсон [пер. з англ. М. Корчінської]. – К.: Вид-во Соломії Павличко «Основи»,2006. – 368, [71] с.
24. Трубачев С. Данилевский Г.П. Биографический очерк [Сочинения: в 24-х т.] / Григорий Петрович Данилевский. – С.-Петербург: Издание А.Ф.Маркса, 1901. – Т.1. 1901. - 132, [71] с.
25. Чужое: опыты преодоления. Очерки из истории культуры Средиземноморья. [Гл. ред. серии Ш.М.Шукуров]. – М.: Алетея, 1999. – 384с, [10-14] с.
26. Шерман О.М. Художня специфіка російського історичного роману другої половини XIX століття і проблеми розвитку жанру: дис. канд. філол. наук: 10.01.02 «Російська література»/ Шерман Олена Михайлівна. – Львів, 1999. – 194с.
27. Шевченко Т.Г. Гайдамаки//Тарас Григорович Шевченко. – К., 1967. – 171с.
28. Эткинд А. Русская литература, XIX век: Роман внутренней колонизации // Александр Эткинд «НЛО» 2003, №59. Режим доступа:<http://magazines.russ.ru/nlo/2003/59/etk.html>

Соломія БУК

ІНТЕРПРЕТАЦІЯ ЛЕКСИКИ ОСТАНЬОГО РОМАНУ ІВАНА ФРАНКА “ВЕЛИКИЙ ШУМ” У КІЛЬКІСНО-СТАТИСТИЧНОМУ АСПЕКТІ

На підставі частотного словника роману І.Франка “Великий шум”, укладеного за методикою комплексного квантитативного опису творів письменника, отримано важливі статистичні характеристики тексту, такі як його обсяг, кількість різних словоформ та слів, індекс різноманітності (багатство словника), середня повторюваність слова в тексті, кількість нарах legomena, індекс винятковості у тексті та у словнику, індекс концентрації, співвідношення між рангом слова та величиною покриття тексту. Текст проаналізовано за

частотними зонами, внаслідок чого з'ясовано, яка лексика має найбільш активне функціонування; виявлено значеннєво найважливіші для аналізованого тексту слова самостійних частин мови, тематично пов'язаних зі селом та його соціальними та моральними проблемами. Важливим результатом дослідження є те, що навіть серед низькочастотної лексики загальноживана має абсолютну кількісну перевагу над пасивною.

Ключові слова: частотний словник, корпус тексту, статистичні характеристики тексту, частотні зони, багатство лексики.

S. N. Buk. Interpretation of the vocabulary of the last novel of Ivan Franko "The Great Noise" in the quantitative-statistical aspect

On the basis of the frequency dictionary of the novel of Ivan Franko "The Great Noise", compiled using the methodology of complex quantitative description of writer's texts, important statistical parameters of text are obtained. These are: text size, number of different wordforms and words, variety index (vocabulary richness), mean word frequency, number of hapax legomena, exclusivity index for the text and the dictionary, concentration index, and the relation between the word rank and text coverage. The text was analyzed with respect to frequency zones and it was clarified which vocabulary has the most active functioning. Within the words of independent parts of speech thematically related to countryside and its social and moral problems those being semantically most important for the analyzed text were identified. An important result of the study is that even in the low-frequency vocabulary the words of general use are absolutely predominant over the passive vocabulary.

Ключевые слова: частотный словарь, корпус текста, статистические характеристики текста, частотные зоны, богатство лексики.

S. N. Buk. Interpretation of the vocabulary of Velykyj Shum [The Great Noise], the last novel by Ivan Franko, in the quantitative-statistical aspect

On the basis of the frequency dictionary of Velykyj Shum [The Great Noise], a novel by Ivan Franko, which was compiled using the methodology of complex quantitative description of writer's texts, important statistical parameters of text are obtained. These are: text size, number of different wordforms and words, variety index (vocabulary richness), mean word frequency, number of hapax legomena, exclusivity index for the text and the dictionary, concentration index, and the relation between the word rank and text coverage. The text was analyzed with respect to frequency zones and it was clarified which vocabulary has the most active functioning. Within the words of independent parts of speech thematically related to countryside and its social and moral problems those being semantically most important for the analyzed text were identified. An important result of the study is that even in the low-frequency vocabulary the words of general use are absolutely predominant over the passive vocabulary.

Key words: frequency dictionary, text corpus, statistical parameters of text, frequency zones, vocabulary richness.

1. Вступ. У світовій лінгвістиці ХХ-ХХІ ст. поживався інтерес до математичної та статистичної складової в описі мови та мовлення. На думку польського лінгвіста А. Павловського, причиною популярності застосування кількісних методів у мовознавстві є “верифікаційність, формальний ригоризм, а також високий рівень узагальненості законів фізики і тверджень математики, що гарантують надійність висновків та остаточних підсумків” [43, 92].

У процесі вивчення мови письменника науковцям, як правило, складно точно вказати, у чому полягає багатство лексики певного автора, хоча інтуїтивно цю якість відчують [32, 78]. Розподіл слів у тексті має певну регулярність і творить його статистичну структуру. Унаочнює її частотний словник (далі — ЧС), тобто лексикографічна праця, що наводить частоту певної мовної одиниці в обстежених текстах. Розвиток комп'ютерної техніки пришвидшив укладання ЧС ідіотетичних мов, функціональних стилів, творчості письменників тощо.

“Можливість використання статистичних методів у стилістиці ґрунтується на тому, що ... текст є результатом добору певних одиниць із загальнонародної мови. ... дослідник стилю повинен простежити всі особливості авторського добору, рухаючись в протилежному — порівняно з автором — напрямку: від тексту — до правил його побудови, від потоку слів — до закономірностей їх добору” [24, 23-24]. Існує вже певна традиція використання кількісних методів до виявлення особливостей ідіолекту письменника [10; 11; 19; 33; 36] і, зокрема, на першому етапі з цією метою до його творів укладають ЧС [16; 26; 27; 37-42; 44; 45].

Деякі лінгвостатистичні дослідження виконано й на матеріалі творчості І. Франка, зокрема видання “Лексика поетичних творів І. Франка” [17] містить інформацію про кількість вживань кожної лексеми у поезії письменника, Л. Полюга на цій основі здійснив статистичний аналіз поетичної лексики цього автора [23], дав опис у цьому ключі поетичним антропонімам [22]. Укладено ЧС й до прозових творів І. Франка, зокрема до романів “Перехресні стежки” [9], “Для домашнього огнища” [6], “Основи суспільності” [7]. Останній твір І. Франка великого жанру — “Великий шум” — був об'єктом дослідження філологічних студій, наприклад [3; 12; 14] тощо, проте у кількісному ракурсі аналізується вперше.

Мета статті — дати кількісно-статистичну інтерпретацію лексики останнього роману Івана Франка “Великий шум” — зумовила постановку таких завдань: укласти ЧС твору, виявити статистичні

параметри тексту твору, зокрема, багатство словника, індекси винятковості, концентрації, співвідношення між рангом слова та величиною покриття тексту; проаналізувати лексику твору за частотними зонами. Актуальність теми та новизна роботи полягає насамперед у комплексному лінгвостатистичному описі роману. Постановка проблеми у вказаному ракурсі дозволяє виявити на новому матеріалі досі не досліджені особливості ідіолекту І. Франка.

Дослідження продовжує традицію української лінгвостатистики, репрезентовану, зокрема, іменами В. Перебийніс [20; 21; 35], Л. Алексієнко [1], Н. Дарчук [13; 34] тощо.

2. Електронний корпус тексту роману “Великий шум” для укладання ЧС твору. З метою укладання ЧС роману “Великий шум” попередньо було створено електронний корпус тексту (далі — КТ) цього твору. В основу лягло академічне 50-томне видання творів І. Франка 1979 р. [29], а також останнє прижиттєве видання роману 1907 [30]. У “Літературно-науковому вістнику”, звідки здійснено передрук київського видання 1907 р., цензура зробила купюру того місця, де описано подробиці інтимного життя старшої дочки пана Суботи, яку відновлено у зібранні творів 1979 р. за автографом.

У текстах є певна непослідовність написання імен та прізвиськ персонажів: *Олефір¹ Онефір, Халавко/Халавка, Степан Чапля / Стефан Чапля*. Можливо, редактори 50-томного видання не уніфікували цих розбіжностей через те, що вони незначні. Загалом, тексти видань 1907 та 1979 рр. ідентичні за винятком деяких правописних відмінностей. До речі, правопис роману “Великий шум” 1907 р. збігається із виданням роману “Перехресні стежки” 1900 р. [31].

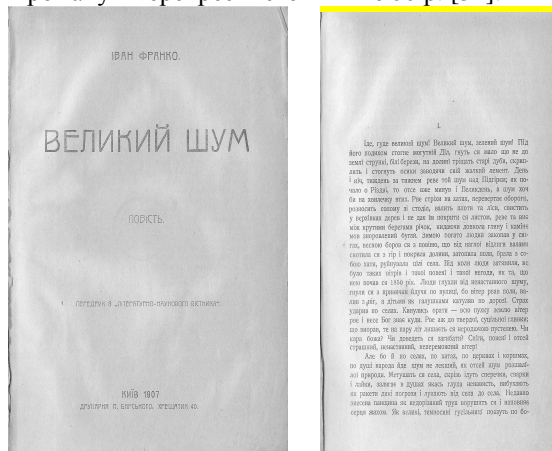


Рис. 1. Сторінки оригінального видання роману І. Франка “Великий шум”, 1907: титул і стор. 3.

КТ роману укладено відповідно до принципів, розроблених для повного КТ І. Франка [5], а отже, він отримав морфологічну, ономастичну, структурну анотації [8]. Таким чином, кожному слововживанню у тексті було присвоєно тег, що вказує на належність до певної частини мови та його початкову форму. Власні назви у творі додатково отримали відповідну позначку про тип оніма: прізвище, ім'я, країна, місто, село, вулиця тощо. Таке маркування тексту уможливило використовувати процесі роботи алгоритми і програми автоматичного та статистичного аналізу лексики. Подаємо зразок з морфологічно- та ономастично анотованого КТ роману Івана Франка “Великий шум”.

ВЕЛИКИЙ<J|ВЕЛИКИЙ> ШУМ<N|ШУМ>
ПОВІСТЬ<N|ПОВІСТЬ>

<c>I

Іде<V|ІТИ>, гуде<V|ГУДІТИ> великий<J|ВЕЛИКИЙ> шум<N|ШУМ>! Великий<J|ВЕЛИКИЙ> шум<N|ШУМ>, зелений<J|ЗЕЛЕНИЙ> шум<N|ШУМ>! Під<AR|ПІД> його<P|ВІН> подихом<N|ПОДИХ> стогне<V|СТОГНАТИ> могутній<J|МОГУТНІЙ> Діл<N|ДІЛ(місцевість)>, гнуться<V|ГНУТИСЯ> мало<D|МАЛО(присл.)> що<AP|ЩО(част.)> не<AR|НЕ> до<AR|ДО> землі<N|ЗЕМЛЯ> стрункі<J|СТРУНКІЙ> білі<J|БІЛИЙ> берези<N|БЕРЕЗА>, на<AR|НА> долині<N|ДОЛИНА> тріщать<V|ТРИЩАТИ> старі<J|СТАРИЙ> дуби<N|ДУБ>, скриплять<V|СКРИПТИ> і<AC|І> стогнуть<V|СТОГНАТИ> осики<N|ОСИКА>, завдячи<VD|ЗАВОДИТИ> свій<P|СВІЙ> жалкий<J|ЖАЛКИЙ> лемент<N|ЛЕМЕНТ>. ...

3. Засади укладання та структура ЧС роману Івана Франка “Великий шум”. Оскільки ЧС роману Івана Франка “Великий шум” — один з етапів реалізації проекту комплексного квантитативного опису творів І. Франка [4], то методика та принципи його укладання, а також схема лематизації (зведення словоформи до її початкової, словникової форми) є такими самими, як і в попередніх ЧС творів письменника.

¹ Походить від скороченої форми імені Елевферій/Елевтерій, від грецьк. ἑλεύθερος — вільний.

Так, у творі розрізнено омонімію, зведено до однієї лексеми фонетичні варіанти. Не зводились до однини лише множинні іменники (*Грушатичі* (село), *кайдани*, *зносини*, *манатки*, *хрестини*, *итани*, *теревені*), а також ті, що збірно позначають осіб жіночої та чоловічої статі (*грушатичани*, *горішняни*, *долішняни*, *двораки*, *мазури*, *молодята*).

Мовлення одного з головних персонажів твору — о. Квінтіліана — сповнене специфічним суржиком, стилізованим під церковно-слов'янську мову, що надає персонажеві особливого колориту. Ті слова з його лексику, що мають українські відповідники, лематизовано як українські, інші ж (*будет*, *вещах*, *вторженіє*, *гидравличний*, *гоненіє*, *како* тощо) супроводжуються у словнику відповідним коментарем *старослов*. Найпростіше було робити таке розрізнення тоді, коли у словах була неукраїнська графіка (*воистину*, *и* (у значенні спол. *і*), *изобразити*).

Макроструктура ЧС містить частотний список словоформ, частотний та алфавітно-частотний список слів, зведених до початкової форми.

Мікроструктура ЧС містить інформацію про ранг (порядковий номер слова у списку, розташованому за спадом частот), абсолютну (кількість вживання слова у тексті) та відносну (відношення абсолютної частоти слова до обсягу тексту) частоти слова, а також міру покриття тексту. Величину (міру) покриття тексту для певного рангу обчислюють як відношення суми абсолютних частот усіх слів з меншими рангами до загальної кількості слів у тексті. Обчислити, яку частку тексту (покриття) становлять слова з найбільшою частотою дає змогу розміщення слів у ЧС за спадом частот. У додатку подано першу сотню найчастотніших слів роману у вигляді ЧС.

4. Деякі результати статистичних підрахунків. Загальний обсяг тексту роману (N) — 37 005 слововживань (для порівняння обсяг “Перехресних стежок” — 93 888, “Основ суспільності” — 67 174, “Для домашнього вогнища” — 44 840), серед них 11 065 різних словоформ ($V_{\text{ф}}$, обсяг словника словоформ). Всього різних слів (лексем або лематизованих слів) у романі 6 471.

Слід зазначити, що 50 найчастотніших слів твору становлять 40 % усіх слововживань¹, тобто більше третини усього тексту. 100 найчастотніших слів покривають майже половину (48,8 %) тексту. 1000 найчастотніших слів покривають 77,7 % тексту. Це означає, що для особи, що не володіє українською мовою, достатньо вивчити цих 1000 слів, щоби розуміти із кожних 1 000 слів 777 без звернення до словника, тобто розуміти цей текст І. Франка більше, ніж на три чверті. Якщо продовжити цей розгляд, то виявиться, що 1 700 найчастотніших слів відповідають 84,2% тексту роману, а на решту 4 760 слів лексики Франкового твору припадає тільки 15,8% тексту. Це слова, що рідко вживаються, серед них 1092 із частотністю 2 та 3368 слів із частотністю 1 (так звані *hарах legomena*, від *д.-грецьк.* ἅπαξ λεγόμενον “сказане один раз”). Кількість *hарах legomena* складає 56,7% (більше за половину!) словника і всього близько 10 % тексту.

Однією з найбільш важливих характеристик тексту, яку дає змогу визначити ЧС, є **багатство словника** (індекс різноманітності). Його обчислюють як відношення обсягу словника лексем до обсягу тексту (V/N), і для “Великого шуму” цей показник становить 0,175. Багатство лексики роману Франка здебільшого міститься саме в цих малочастотних словах: саме за рахунок них у тексті високий **індекс винятковості**. Для тексту його обчислюють як відношення кількості *hарах legomena* (V_1) до обсягу тексту (V_1/N) — 0,099. Для словника індекс винятковості обчислюють як відношення кількості *hарах legomena* (V_1) до обсягу словника (V_1/V) — 0,57. Два останні числа — показники варіативності лексики.

Протилежним до індексу винятковості є **індекс концентрації** — відношення кількості слів з високою частотою (10 і вище, їх у тексті $N_{10} = 25\ 456$, у словнику — $V_{10} = 479$) до обсягу тексту ($N_{10}/N = 0,688$) або словника ($V_{10}/V = 0,074$).

Із відношення обсягу тексту до обсягу словника лексем (N/V) отримуємо **середню повторюваність слова в тексті** — 5,7. Іншими словами, кожне слово в досліджуваному творі в середньому вжито приблизно 6 разів.

Відносно невелика кількість високочастотної лексики у словнику лем (і, відповідно, низький індекс концентрації) та порівняно велика кількість слів із частотою 1 (і, відповідно, високий індекс винятковості) свідчать про неабияке різноманіття лексики роману.

5. Одним із важливих завдань створення ЧС є виявлення **частотних зон словника**. Як зазначає Ю. Тулдава [25, 65], стратифікація лексики за частотою вживання має велике значення для лінгвістики, педагогіки та психології (визначення складності тексту, укладення словника-мінімуму, проведення психологічних експериментів), автоматичного опрацювання тексту. Кількісні характеристики слів, що стосуються певної частотної зони, багато в чому корелюють з якісними властивостями цих слів, наприклад з такими властивостями, як нейтральність (загальноновживаність), тематичність (ключові слова), інформативність тощо. Існують різні способи визначення меж цих частотних зон. Для простоти умовно поділимо лексику роману на чотири зони: **найчастотніша лексика** (перших 104 слова,

¹ Опис здійснено за методикою, запропонованою у [32].

частотність 1-50), **лексика із верхньою середньою частотністю** (ранг 105-479, частота 49-10), **лексика із нижньою середньою частотністю** (ранг 480-2802, частота 9-2) **низькочастотна лексика** (ранг 480-2802, частота 9-2), **парах legomena** (частота 1).

Таблиця 1.

Частотні зони лексики роману І.Франка “Великий шум”				
Зона частот	слів у тексті		слів у словнику	
Частота ≥ 50	18256	49,33%	104	1,61%
Частота 10-49	7200	19,46%	375	5,80%
Частота 2-9	7881	21,30%	2324	35,91%
Частота 1	3668	9,91%	3668	56,68%
	37005	100,00%	6471	100,00%

5.1. Найчастотніша лексика роману — це 104 слова (ранг 1-104), що мають найінтенсивніше функціонування в тексті (частотність 1 585-50). Як було сказано вище, займаючи всього 1,61 % словника, вони покривають майже половину тексту.

Першими за частотою вживання у романі очікувано виявились **службові частини мови**: сполучники *і/й* (1585), *а* (531), *та* (311), *що* (271) тощо; частка *не* (942); прийменники *в/у* (729), *на* (689), *з/із/зі/зо* (594), *до* (453) тощо. Це універсальна характеристика будь-якого тексту [13; 34; 35].

Серед **самостійних частин мови** лідирують **займенники** *він* (749), *я* (603), *той* (449), *що* (370), *вони* (333); *ви* (329), *свій* (323), *весь/увесь* (272), *сей* (269); **дієслова із найзагальнішим значенням** *бути* (410); *мати* (139); *могти* (90); *дати* (66); **дієслова мовлення** *сказати* (107), *говорити* (86), *мовити* (65); **мислення** *знати* (129), *хотіти* (93); **перцепції** *бачити* (64), *чути* (64); **пересування** *йти/ити* (69), *піти* (55).

В **іменниках** першість мають ввічливе звертання у Галичині того часу до світських осіб *пан* (322) і до духовних осіб *о. (отець)* (81); назви персонажів *Дум'як* (132), *Квінтіліан* (89), *війт* (60), *Субота* (66), *Кость* (50). Серед частин тіла в І. Франка тут, як і в інших творах, найбільш уживаними є *рука* (88), *око* (68), *голова* (50). Окремого розгляду у порівнянні з іншими творами в майбутньому потребують номінативи *слово* (76), *Бог* (56), *стіл* (54), що також належать до цієї зони роману.

Слід зауважити, що серед найчастотнішої лексики твору є лише декілька **прислівників** *так* (95), *як* (77), *коли* (71), *де* (50) і немає **прикметників**. Показово, що перший прикметник — *панський* — має ранг 106 із частотою 48.

На позначення слів/понять, особливо важливих для творчості конкретних письменників, дослідники вживають різні терміни, наприклад: ключові слова [28], ідіоглоси [15], текстами [2], ключові (базові) концепти [18] та ін. Проте науковці сходяться на думці, що вони мають високу частотність у тексті і несуть особливу концентрацію думки. Як правило, у письменницьких словниках їм справедливо надають особливу увагу, зокрема подають окремим списком. Для роману І. Франка “Великий шум”, дія якого відбувається у **селі**, саме воно становить собою важливий тематичний центр, що відобразилося в лексиці, з ним пов'язані: *село* (90, у порівнянні *місто* (20)), *люди* (89), *паничина* (88), *селяни* (88), *громада* (52), *коршма* (51). Ці значеннево найважливіші слова твору було виявлено завдяки частотному списку.

Привертає увагу, що перших 100 слів і зараз належать до загальноживаної нейтральної лексики (окрім, хіба що, займенника *сей*).

5.2. Лексика із верхньою середньою частотністю (ранг 105-479, частота 49-10) займає 5,8 % словника та 19,5 % тексту і за складом суттєво не відрізняється від **лексики із нижньою середньою частотністю** (ранг 480- 2802, частота 9-2).

Сюди потрапили **власні назви головних персонажів**: *Галя* (37), *Гордієра* (34), *Яць* (34), *Коваль* (прізвище, 30), *Чапля* (прізвище, 26), *Халавка* (13), *Женя* (10) та головний топонім *Грушатичі* (село, 19). Досить вагому групу становлять **загальні назви людей, їх соціального стану, терміни спорідненості**: *староста* (44), *чоловік* (38), *батько* (37), *мужик* (37), *жінка*, *комісар* (36), *хлоп* (32), *дитина*, *дочка*, *панна* (25), *народ* (24), *діти* (23), *дідич* (22), *цісар* (22), *піп* (21), *капітан* (20), *егомоть* (18), *сваха*, *таточко* (18), *пані*, *панотець* (17), *граф*, *ландсдрагони* (16), *дівчина*, *юрба*, *анархіст* (13), *пан-дідич* (12), *мати*, *матір*, *родина*, *сват*, *сендзя* (11), *отець* (священик), *парубок* (10).

Логічно було передбачити тут також **слова пов'язані із життям та побутом людей, зокрема селян**: *хата* (45), *двір* (45), *церква* (36), *дорога* (35), *земля* (28), *двері* (26), *горівка* (25), *ліс* (24), *робота* (23), *поле* (22), *трунт* (21), *місто* (20), *подвір'я* (19), *дім* (18), *кінь* (18), *худоба* (15), *сіни* (13), *пісня*, *спів*, *співати* (12), *гість*, *господар*, *праця*, *хліб* (11), *вілла* (10).

Універсальними для творів І. Франка є слова на позначення **часових проміжків**: *день* (34), *рік* (32), *літо* (25), *неділя* (15), *сьогодні* (15), *вечір* (12), *торік* (11), *ніч* (10) та **частин тіла**: *серце* (31), *лице* (26), *нога* (22), *вухо/ухо*, *долоня* (14), *зуб*, *плече* (12), *тіло* (11).

Слова з **абстрактним значенням** пов'язані з морально-етичними та соціальними проблемами, які

піднімає роман: *душа* (47), *життя* (36), *дух* (32), *серце* (32), *світ* (29), *воля* (26), *думка* (24), *правда* (23), *Господь* (15), *надія* (14), *кривда* (13), *біда*, *тверезість* (12), *власть* (10), *новина*, *повага*, *сумління*, *тон* (10).

У зоні середніх частот, порівняно з іншими, більше **прикметників**: *панський* (48), *старий* (41), *великий* (35), *новий* (34), *хлопський*, *царський* (21), *божий*, *довгий*, *молодий* (19), *білий*, *високий*, *власний*, *польський* (18), *церковний* (17), *простий* (16), *громадський*, *грушатицький* (15), *руський*, *святий* (15), *широкий* (15), *тяжкий*, *чорний* (14), *вельможний*, *людський*, *попівський*, *страшний*, *урядовий* (13), *весільний*, *близький*, *важкий* (12), *живий*, *здоровий*, *шлюбний* (11), *гарний*, *двірський*, *селянський*, *чистий* (10),

Серед **дієслів** кількісно найбільшою є група **мовленевої діяльності** *казати*, *слухати*, *просити* (32), *запитати*, *скрикнути* (27), *відповісти* (25), *крикнути* (23), *кричати* (21), *промовити* (16), *написати* (12), *веліти* (17), *бажати* (14), *благословити* (10). Також умовно виділяються групи дієслів **пересування та зміни положення** *вийти* (30), *сидіти* (26), *прийти* (25), *їхати* (21), *приходити* (12), *ввійти*, *сідати*, *сісти* (10); **мислення** *думати* (25), *вважати/уважати* (14), *розуміти* (11), *подумати* (12); **зорового сприйняття** *побачити* (26), *дивитися* (25); **фізичної дії** *давати* (25), *зробити* (24), *взяти/узяти* (23), *подати* (13), *держати* (12) та ін.

Цікаво, що у цій частотній зоні фактично немає діалектної лексики і є тільки декілька слів, що для сучасної мови архаїчні або стилістично забарвлені: *против* (присл., 28), *дідич* (22), *власть* (10), *цар* (22), *ландсдратони* (16), *канчук* (13), *рад* (10). Слова не містять застарілих словотвірних елементів.

5.3. До низькочастотної лексики (ранг 480-2802, частота 9-2) належить досить багато **топонімів**: *Чехія* (3), *Болеслав*, *Грушівка* (річка), *Дрогобич*, *Перемішель*, *Підгір'я*, *Поділля*, *Польща*, *Прага*, *Самбір* (2); **власних назв епізодичних персонажів**: *Гаврило*, *Гадина*, *Голуховський*, *Зарічний*, *Климентова*, *Хома* (2). Відносно велику групу лексики із частотою 2 становлять польські слова (пан Субота писав по-польськи, оскільки по-руськи не вмів): *Anton*, *Halina*, *Swoboda* (ще один цікавий випадок неуніфікованості: прізвище “Субота” польською послідовно написано “Swoboda”, а не “Sobota”), *błogosławieństwo*, *imię*, *mój*, *podpisywać*, *pop*, *pretensja*, *spadkobierca*, *warunek*, *wesele*, *wola*, *zgoda* та ін.

Сюди власне входять деякі **діалектизми**: *вислугуватися*, *вівкання*, *витрублювати*, *вояк*, *вояцький*, *вразіння*, *вроді*, *готов*, *губерніальний*, *гумно*, *дослугуватися*, *жених*, *зажадати*, *їй-богу*, *концепт* (у значенні *план*, *схема*), *красний* (у значенні *гарний*), *лотаиш* (квітка), *лотерія*, *міліон*, *мінута*, *молодеж*, *молодіж*, *моримуха* (*мухомор*), *мундур*, *недовір'я*, *одвічальність*, *одур*, *паня*, *підгнітити*, *против* (присл.), *пакати* (палити люльку), *пожиття*, *празник*, *фотель* (2); **застаріла лексика**: *возний* (судовий урядовець), *клебаня* (так звуть у селі попівську резиденцію — пояснення І. Франка), *ландсдратонський* (кіннотний), *преліпотент* (уповноважений у справах громадського майна), *сівня* (козуб або мішок із зерном, який вішають на плече, коли засівають поле ручним способом), *староство*, *фельдфебель*,

5.4. Нарах legomena — слова із частотою 1 — складають 3 368 одиниць і займають 9,9% тексту та 56,7% (більше за половину!) словника. Тут абсолютно переважає **стилістично нейтральна лексика** (*аргумент*, *брунька*, *вартувати*, *гірський*, *доволі*, *іронізувати*, *кишеня*, *людяний*, *мляво*, *надалі*, *овес*, *палкий*, *пізно*, *темінь*, *фізичний*, *хлипати*, *цупко*), хоча трапляються **діалектизми**: *абишит* (звільнення зі служби), *горі* (угорі, угору), *тавра* (лігво), *дихавичний* (хворий на дихавицю, тобто астму), *загарливо* (палко), *здалий* (здатний, придатний), *злямпартований* (зруйнований розпусним життям), *зманерований* (що має неприродну манірну поведінку), *куферок* (валізка), *лакомий* (пожадливий), *напруго* (наполегливо, рішуче), *нагніт* (гноблення), *обарінок* (бублик), *омаста* (жир, приправа), *пленіпотент* (уповноважений у майнових справах), *фасоля*; **застаріла лексика**: *будуще*, *виїмок*, *віно*, *достава* (постачання), *ідовитий*, *много*, *оградити*, *танечник*, *танцюра* (танцюрист); **просторіччя**: *видко*, *гвалт*, *одітий*, *писок*, *струмент*. Слова, які можна розцінювати як **словотвірні варіанти сучасних літературних**, можуть стати темою окремого цікавого дослідження: *втека* (втеча), *гарнізоновий* (гарнізонний), *геркулічний* (геркулесовий), *дитинячий*, *дітський* (і *дитячий*), *дідьчий* (дідьків), *енергічно* (енергійно), *жартовливий* (жартівливий), *жіноцький* (жіночий), *жмінька* (жменька), *заробок* (заробіток), *зліворуч* (ліворуч), *легкомисно* (легковажно), *отверезіння* (протверезіння), *охотник* (охочий), *партеровий* (партерний),

6. Висновки. ЧС роману І. Франка “Великий шум”, який укладено за методикою комплексного квантитативного опису творів письменника, дав змогу отримати важливі статистичні характеристики тексту, такі як його обсяг, кількість різних словоформ та слів, індекс різноманітності (багатство словника), середня повторюваність слова в тексті, кількість нарах legomena, індекс винятковості у тексті та у словнику, індекс концентрації, співвідношення між рангом слова та величиною покриття тексту. Ці характеристики у перспективі цікаво буде зіставити з аналогічними квантитативними даними усіх інших романів І. Франка. Зведений ЧС усіх творів письменника покаже багатство його ідіолекту і уможливить порівняти його з багатством лексики інших письменників (як українських, так і зарубіжних), до творчості яких уже є ці дані (Т. Шевченко, Г. Квітка-Основ'яненко, Ф. Достоєвський, В. Гюго тощо). Аналіз тексту за частотними зонами з'ясував, яка лексика має найбільш активне функціонування, також

було виявлено значеннєво найважливіші для аналізованого тексту слова самостійних частин мови, тематично пов'язаних зі селом та його соціальними та моральними проблемами. Важливим результатом дослідження є те, що навіть серед низькочастотної лексики загальноживана має абсолютну кількісну перевагу над пасивною.

Таким чином, ЧС роману “Великий шум”, що самостійно окремо дає цікаві дані для інтерпретації, набуває особливого значення у світлі загального проекту як фрагмент комплексного аналізу Франкової творчості у статистичному та якісному аспектах.

ЛІТЕРАТУРА

1. *Алексієнко Л. А., Дарчук Н. П.* Принципи створення параметризованої бази даних за поетичними текстами Лесі Українки // *Леся Українка і сучасність (До 130-річчя від дня народження Лесі Українки): Зб. наук. пр.*— Луцьк: Волин. обл. друк., 2004.— С. 344-352.
2. *Бабайцева В. В., Инфантова Г. Г., Чесноков П. В.* [Рецензия на] Словарь языка Михаила Шолохова. М., 2005 // *Филологические науки.*— 2005.— № 5.— 108–113.
3. *Боднар В. Т.* Авторська рецептивна стратегія та система персонажів повісті Івана Франка “Великий шум” // *Вісник Житомирського державного університету імені Івана Франка.* № 26.— С. 76-78.
4. *Бук С.* Квантитативна параметризація текстів Івана Франка: спроба проекту // *Іван Франко: Студії та матеріали: Збірник наукових статей.* — Львів : ЛНУ імені Івана Франка, 2010 (у друці), див. препринт arXiv:1005.5466v1 [cs.CL].— Електронний ресурс <<http://arxiv.org/abs/1005.5466>>.
5. *Бук С.* Корпус текстів Івана Франка: спроба визначення основних параметрів // *Прикладна лінгвістика та лінгвістичні технології: MegaLing-2006:* Зб. наук. пр. — К.: Довіра, 2007.— С. 72–82.
6. *Бук С.* Роман Івана Франка “Для домашнього вогнища” крізь призму частотного словника / С. Н. Бук // *Препринт arXiv:1006.0153v1 [cs.CL].*— Електронний ресурс <<http://arxiv.org/abs/1006.0153>>.
7. *Бук С.* Статистичні характеристики роману Івана Франка “Основи суспільності” (на основі частотного словника твору) // *Вісник: Проблеми української термінології.*— Львів: Національний університет “Львівська політехніка”.— 2010. (у друці)
8. *Бук С.* Структурне анотування у корпусі текстів (на прикладі прози Івана Франка) // *Українська мова.*— 2009.— № 3.— С. 59-71.
9. *Бук С., Ровенчак А.* Частотний словник роману Івана Франка “Перехресні стежки” // *Стежками Франкового тексту (комунікативні, стилістичні та лексичні виміри роману “Перехресні стежки”).* — Львів: Видавничий центр ЛНУ імені Івана Франка, 2007. — С. 138-369.
10. *Вокальчук Г. М., Мамчич Т. І.* Частотний аналіз авторських лексичних новотворів у поетичному словнику “Молодої музи” // *Актуальні проблеми сучасної філології. Мовознавчі студії: зб. наук. праць Рівненськ. держ. гуманітарного ін-ту.*— Вип 14.— Рівне, 2006.— С. 44-47.
11. *Гаспаров Б. М., Гаспарова Э. М., Минц З. Г.* Статистический подход к исследованию плана содержания художественного текста // *Учен. зап. Тарт. ун-та.*— 1971.— Вып. 284: Тр. по знаковым системам. 5.— С. 288-309.
12. *Гуняк М. Ф.* Творчість Івана Франка початку ХХ ст. (особливості етико-антропологічного та історіософського синтезу): Автореф. дис... канд. філол. наук: 10.01.01 / Львівський держ. ун-т ім. Івана Франка. — Львів, 1999. — 20с.
13. *Дарчук Н. П.* Статистичні характеристики лексики як відображення структури тексту // *Мовознавчі студії.*— К.: Наукова думка, 1976.— С. 97–102.
14. *Дуркалевіч В. В.* Синтетизм творчого мислення Івана Франка (проза початку ХХ ст.): автореф. дис... канд. філол. наук: 10.01.06; Тернопільський національний педагогічний ун-т ім. Володимира Гнатюка. — Тернопіль, 2007. — 20 с.
15. *Караулов Ю.Н.* Понятие идиоглоссы и словарь языка Достоевского // *Слово Достоевского.* 2000: Сб. статей.— М.: Азбуковник, 2001.— С. 424-444.
16. *Карпова О. М.* Частотный словарь писателя как инструмент стилистического анализа художественного текста // *Анализ стилей художественной и научной литературы.*— Л., 1990.— С. 11-18.
17. *Ковалик І. І., Оцишко І. Й., Полюга Л. М. (уклад.)* Лексика поетичних творів Івана Франка. Методичні вказівки з розвитку лексики.— Львів: ЛНУ, 1990.— 264 с.
18. *Монахова Т. В.* Мова Валерія Шевчука: ключові концепти, корпус, тезаурус: Автореф. дис. ... канд. філол. наук.— К.: КНУ, 2008.— 18 с.
19. *Перебийніс В. С.* Статистична стилістика // *Українська мова: Енциклопедія / редкол.: В. М. Русанівський та інші.*— 2-ге вид., випр. і доп.— К.: В-во Українська енциклопедія ім. М. П. Бажана, 2004.— С. 644–645.
20. *Перебийніс В. С.* Статистичні методи для лінгвістів: Навчальний посібник.— Вінниця: “Нова книга”, 2002.— 168 с.
21. *Перебийніс В. С., Муравицька М. П., Дарчук Н. П.* Частотні словники та їх використання.— К.: Наукова думка, 1985.— 204 с.
22. *Полюга Л.* Антропоніми в лексичній системі поетичних творів І. Франка // *Іван Франко — письменник, мислитель, громадянин. Матеріали міжнародної наукової конференції (Львів 25-27 вересня 1996) /* Ред. М. Ільницький, Б. Якимович.— Львів: Світ, 1998.— С. 657–661.
23. *Полюга Л.* Статистичний аналіз лексики поетичних творів І. Франка // *Іван Франко і національне відродження.*— Львів: ЛДУ ім. І. Франка; Ін-т франкознавства, 1991.— С. 164–166.
24. *Статистичні параметри стилів /* за ред. В. С. Перебийніс.— К.: Наук. думка, 1967.— 260 с.
25. *Тулдава Ю. А.* Проблемы и методы системно-квантитативных исследований лексики / Отв. ред.

- Р. Г. Пиотровский.— Таллин: Валгус, 1987.— 207 с.
26. *Фоменко И. В.* Частотный словарь и интерпретация художественного текста // Лингвистика и поэтика: преодоление границ. Памяти Наталии Алексеевны Кожевниковой : сб. науч. тр. — М.: Тверь, 2008. — С. 50–64.
 27. *Фоменко И. В.* Частотный словарь как основа интерпретации романа Л. Добычина «Город Энн» // Добычинский сборник-3. Даугавпилс, 2001.— С. 98–107.
 28. *Форманова С. В.* Ключові слова у мовній картині світу Михайла Коцюбинського. Автореф. дис... канд. філол. наук: 10.02.01; НАН України. Ін-т мовознав. ім. О.О.Потебні. — К., 1999. — 17 с.
 29. *Франко І.* Великий шум // Зібрання творів у 50-ти томах.— Т. 22: Повісті та оповідання.— К.: Наукова думка, 1979.— С. 208–317.
 30. *Франко І.* Великий шум: Повість.— Київ: Друкарня П. Барського. (Передрук з Літературно-наукового вістника), 1907.— 116 с.
 31. *Франко І.* Перехресні стежки: Повість.— Львів: Видане ред. “Літературно-наукового вістника”, 1900.— 320 с. (на обкладинці вказано 1899 рік).
 32. *Фрумкина Р. М.* Статистическая структура лексики Пушкина // Вопросы языкознания.— 1960.— № 3.— С. 78–81.
 33. *Фрумкина Р. М.* Статистические методы изучения лексики.— М.: Наука, 1964.— 115 с.
 34. Частотний словник сучасної української поетичної мови (2003-2010)— Електронний ресурс <www.mova.info/Page2.aspx?I1=89>.
 35. Частотний словник сучасної української художньої прози / За ред. *Перебийніс В. С.*— К.: Наук. думка, 1981.— Т. 1.— 863 с.; Т. 2.— 855 с.
 36. *Шайкевич А. Я.* Персонажи романа “Идиот” в лингвостатистическом освещении // Достоевский и современность. Материалы IX Международных Старорусских чтений 1994.— Новгород, 1995.— С. 214-216.
 37. *Шайкевич А. Я., Андрищенко В. М., Ребецкая Н. А.* Статистический словарь языка Достоевского.— 2003.— 880 с.
 38. *Balowski M.* Lista frekwencyjna poezji, prozy i dramatu Krzystofa Kamila Baczyńskiego.— Prohovice: Pro, 1997.
 39. *Brunet E.* Le vocabulaire de Victor Hugo. Genève: Slatkine, 1988.
 40. *Colin B., Rutten C.* Aristote, categoriae: index verborum, listes de fréquence. Liège: Université de Liège, 1993.
 41. Complete Shakespeare wordlist. Available from <http://en.wiktionary.org/wiki/Wiktionary:Frequency_lists/Complete_Shakespeare_wordlist>
 42. *Hanley M. L.* Word index to James Joyce’s “Ulysses”. Madison: University of Wisconsin Press, 1951.
 43. *Pawłowski A.* O stosowalności metod ścisłych w badaniach języka // Efekt motyla. Humanisci wobec teorii chaosu / *Dorota Heck, Kordian Bakula* (red.). Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.— 2006.— S. 92-99.
 44. Slovník Karla Čapka / Hl. ed. František Čermák. — Praha : Nakladatelství Lidové noviny; Ústav Českého národního korpusu, 2007. — 715 s.
 45. *Vasić S.* Polazne osnove novije srpske proze. Knjiga II. Hazarski rečnik Milorada Pavića: frekvencijski rečnik.— Beograd: Kultura, 1998.

Додаток

**Перша сотня найчастотніших лем роману І. Франка
“Великий шум”**

Ранг	Лема	Абс. час т.	Відн. част., %	Покр., %
1	І—1265;Й—320	158		
	-СЯ	102	2,78	4.28
		9		
2	НЕ	942	2.55	6.83
3	ВІН	749	2.02	8.85
4	В—494;У—235	729	1.97	10.82
5	НА	689	1.86	12.68
6	Я	603	1.63	14.31
7	З—476;ІЗ— 62;ЗІ—41;ЗО—15	594	1.61	15.92
8	А(спол.)	531	1.43	17.35
9	ДО	453	1.22	18.58
10	ТОЙ	449	1.21	19.79
11	БУТИ	410	1.11	20.90
12	ЦЮ(займ.)	370	1.00	21.90
13	ВОНИ	333	0.90	22.80
14	ВИ	329	0.89	23.69
15	СВІЙ	323	0.87	24.56

Ранг	Лема	Абс. час т.	Відн. част., %	Покр., %
16	ПАН	322	0.87	25.43
17	ТА(спол.)	311	0.84	26.27
18	ВЕСЬ— 187;УВЕСЬ—85	272	0.74	27.01
19	ЩО(спол.)	271	0.73	27.74
20	СЕЙ	269	0.73	28.47
21	МИ	251	0.68	29.14
22	ЗА	246	0.66	29.81
23	ЯК(спол.)	228	0.62	30.43
24	ВОНА	226	0.61	31.04
25	ТАКИЙ	197	0.53	31.57
26	ПО	181	0.49	32.06
27	ТИ	169	0.46	32.51
28	ЩЕ—151;ЩЕ— 17	168	0.45	32.97
29	Ж(част.)—130;ЖЕ— 30	160	0.43	33.40
30	ЯКИЙ	160	0.43	33.83
31	АЛЕ	158	0.43	34.26

Ра нг	Лема	Абс · час т.	Відн. част., %	Покр., %
32	МІЙ	150	0.41	34.67
33	МАТИ(дієсл.)	139	0.38	35.04
34	СЕБЕ	135	0.36	35.41
35	ВІД	133	0.36	35.77
36	ТО(спол.)	133	0.36	36.12
37	ДУМ'ЯК(прізви.)	132	0.36	36.48
38	ЗНАТИ	129	0.35	36.83
39	ВЖЕ—95;УЖЕ— 30	125	0.34	37.17
40	САМ	125	0.34	37.51
41	Б—34;БИ—78	112	0.30	37.81
42	СКАЗАТИ	107	0.29	38.10
43	ОДИН	95	0.26	38.35
44	ТАК(присл.)	95	0.26	38.61
45	ХОПІТИ	93	0.25	38.86
46	ТО(част.)	92	0.25	39.11
47	ПРО	91	0.25	39.36
48	МОГТИ	90	0.24	39.60
49	СЕЛО	90	0.24	39.84
50	РУКА	88	0.24	40.08
51	ГОВОРИТИ	86	0.23	40.31
52	ЩОБ	84	0.23	40.54
53	НАШ	82	0.22	40.76
54	РАЗ(ім.)	82	0.22	40.98
55	ЧИ(спол.)	82	0.22	41.21
56	О.(ОТЕЦЬ)	81	0.22	41.42
57	ХТО	81	0.22	41.64
58	КВІНТІЛІАН(ім'я)	79	0.21	41.86
59	ЛЮДИ	79	0.21	42.07
60	ЯК(присл.)	77	0.21	42.28
61	СЛОВО	76	0.21	42.48
62	ТУТ	76	0.21	42.69
63	КОЛИ	71	0.19	42.88
64	ТЕПЕР	71	0.19	43.07
65	ІТИ—23;ЙТИ—46	69	0.19	43.26
66	ОКО	68	0.18	43.44
67	НУ	67	0.18	43.62
68	ЩО(част.)	67	0.18	43.80

Ра нг	Лема	Абс · час т.	Відн. част., %	Покр., %
69	ДАТИ	66	0.18	43.98
70	СУБОТА(прізви.)	66	0.18	44.16
71	МОВИТИ	65	0.18	44.34
72	БАЧИТИ	64	0.17	44.51
73	ВАШ	64	0.17	44.68
74	ПІД	64	0.17	44.86
75	ЧУТИ	64	0.17	45.03
76	АНІ	63	0.17	45.20
77	НІ	62	0.17	45.37
78	НІЩО	62	0.17	45.53
79	ЯКИЙСЬ	62	0.17	45.70
80	ВІЙТ	60	0.16	45.86
81	ЗНОВ	59	0.16	46.02
82	ПАНЩИНА	58	0.16	46.18
83	СЕЛЯНИ	58	0.16	46.34
84	БО	57	0.15	46.49
85	БОГ	56	0.15	46.64
86	БЕЗ	55	0.15	46.79
87	ПШТИ	55	0.15	46.94
88	ТА(част.)	55	0.15	47.09
89	АБИ	54	0.15	47.23
90	АЖ	54	0.15	47.38
91	ДЛЯ	54	0.15	47.53
92	СТІЛ	54	0.15	47.67
93	ХОЧ	54	0.15	47.82
94	МОВ	53	0.14	47.96
95	ПРИ	53	0.14	48.10
96	ГРОМАДА	52	0.14	48.24
97	ДРУГИЙ	51	0.14	48.38
98	КОРШИМА	51	0.14	48.52
99	ЧАС	51	0.14	48.66
100	ГОЛОВА	50	0.14	48.79
101	ДЕ(присл.)	50	0.14	48.93
102	ІНШИЙ	50	0.14	49.06
103	КОСТЬ(ім'я)	50	0.14	49.20
104	РОБИТИ	50	0.14	49.33