

РОМАН ІВАНА ФРАНКА «ДЛЯ ДОМАШНЬОГО ОГНИЩА» КРИЗЬ ПРИЗМУ ЧАСТОТНОГО СЛОВНИКА

У статті описано методику та принципи укладання частотного словника роману І. Франка «Для домашнього огнища». Отримано статистичні параметри лексики твору, зокрема індекси різноманітності (багатство словника), винятковості, концентрації, співвідношення між рангом слова та величиною покриття тексту тощо. На основі частотних словників романів І. Франка «Перехресні стежки» (2007) та «Для домашнього огнища» (2010) здійснено порівняння основних кількісних характеристик вказаних текстів.

Ключові слова: частотність слова, частотний словник, ранг, *harax legomena*, індекси різноманітності, винятковості, концентрації, покриття тексту.

Частотність лексичної одиниці — важлива характеристика слова, оскільки вона свідчить про активність його функціонування в тексті, вагу в статистичній структурі тексту тощо. На цьому наголошують як українські, так і зарубіжні мовознавці¹. З'явилась і нова галузь лінгвістичної науки — статистична лексикографія, що займається, зокрема, теорією та практикою укладання частотних словників. Такий інтерес до статистичного обстеження великих сукупностей текстів зумовлений, з одного боку, необхідністю глибше пізнати закономірності функціонування мовних одиниць у тексті. З другого боку, виникнення статистичної лексикографії зумовлене суспільними потребами, практичним застосу-

¹ Дарчук Н. П. Статистичні характеристики лексики як відображення структури тексту // Мовознавчі студії.— К., 1976.— С. 97–102; Дарчук Н. П., Грязнухіна Т. О. Частотний словник сучасної української публіцистики // Мовознавство.— 1996.— № 4–5.— С. 15–19; Муравицька М. П. Статистичні лінгвістичні дослідження та їх розвиток в українському мовознавстві // Мовознавство.— 1967.— № 5.— С. 47–58; Wimmer W., Altmann G. Review Article : On vocabulary richness // Journ. of Quantitative Linguistics.— 1999.—Vol. 6. N 1.— P. 1–9; Wimmer G. at al. Úvod do analýzy textov.— Bratislava, 2003.— 344 s.; Popesku I.-I., Altmann G. Some aspects of word frequencies // Glottometrics.— 13.— 2006.— P. 23–46; Köhler R., Altmann G. Aims and Methods of Quantitative Linguistics // Problems of Quantitative Linguistics.— Chernivci, 2005.— P. 12–42; Тулдава Ю. П. Проблемы и методы квантитативно-системного исследования лексики.— Таллин, 1987.— С. 62–64; Alekseev P. M. Frequency dictionaries // Quantitative linguistics : an international handbook / Ed. by R. Köhler, G. Altmann, R. G. Piotrowski.— Berlin ; New York, 2005.— P. 312–324.; Алексеев П. М. Частотные словари : Учеб. пособие.— СПб., 2001.— 156 с.; Шайкевич А. Я. О статистическом словаре языка Достоевского // Русский язык в научном освещении.— М., 2001.— № 2.— С. 122–149; Шайкевич А. Я. Количественные методы в языкознании // Лингвистический энциклопедический словарь / Ред. Н. Л. Ярцева.— М., 1990.— С. 231–232; Pawłowski A. Metody kwantytatywne w sekwencyjnej analizie tekstu.— Warszawa, 2001.— 168 s.; Ruszkowski M. Statystyka w badaniach stylistyczno-składniowych.— Kielce, 2004.— 144 s.

ванням надбань мовознавства для удосконалення системи стенографії і відбору лексичного мінімуму для словників різного типу, раціонального використання пам'яті комп'ютера для автоматичного опрацювання текстової інформації, теорії та практики перекладознавства².

Загальне застосування і мета частотного словника (ЧС) полягають в опрацюванні ймовірно-статистичної моделі системного лінгвістичного об'єкта, а саме мови (підмови, стилю, ідіолекту)³. Проте основне завдання ЧС — це насамперед теоретичні дослідження мови, на базі якої укладено частотний словник. Серед тенденцій та перспектив розвитку статистичної лексикографії П. Алексеев вказує на залучення до дослідження великої кількості мов та підмов, перехід від загального до часткового. Кількісно-статистичний опис ідіолекту письменника природно входить у цей контекст.

Таким чином, на стику авторської та статистичної лексикографії виник частотний словник мови письменника, і це закономірно, оскільки для опису ідіолекту письменника необхідні не лише якісні, а й кількісні характеристики лексики його творів. Саме тому для найвидатніших майстрів пера різних часів і народів укладали частотні словники, де наводили кількість уживань слова в обстежених текстах. Так, існують ЧС Арістотеля, В. Шекспіра, В. Гюго, Дж. Джойса, К. Чапека, Ф. Достоєвського, Л. Толстого, А. Чехова, М. Лермонтова, М. Горького⁴, болгарського поета Н. Вапцарова⁵, польського письменника К. Бачинського⁶, сербських прозаїків М. Павича, Д. Албахарія, М. Данойлича, С. Велмар-Янкович⁷ тощо.

В українській лексикографії інформацію про частоту вживання слова містять словники мови Т. Шевченка⁸ та Г. Квітки-Основ'яненка⁹, словопокажчики

² Rybicki J. Stylometria komputerowa w służbie tłumacza (na przykładzie własnych przekładów) // Rocznik przekładoznawczy : Studia nad teorią, praktyką i dydaktyką przekładu.— 2007–2008.— N 3/4.

³ Alekseev P. M. Op. cit.

⁴ Colin B., Rutten C. Aristotele, categoriae: index verborum, listes de fréquence.— Liège, 1993; Complete Shakespeare wordlist.— http://en.wiktionary.org/wiki/Wiktionary:Frequency_lists/Complete_Shakespeare_wordlist; Brune E. Le vocabulaire de Victor Hugo.— Genève, 1988; Hanley M. L. Word index to James Joyce's «Ulysses».— Madison, 1951; Ulysses by J. Joyce. A Ranked Concordance.— <http://www.doc.ic.ac.uk/~rac101/concord/texts/ulysses/>; Slovník Karla Čapka / Hl. ed. F. Čermák. Ustav Českého národního korpusu.— Praha, 2007.— 715 s.; Шайкевич А. Я., Андрущенко В. М., Ребецькая Н. А. Статистический словарь языка Достоевского.— М., 2003.— 880 с.; Частотный словарь романа Л. Н. Толстого «Война и мир» / Великодворская З. Н. и др. (ред.).— Тула, 1978; Гребенищikov А. О. Частотный словарь рассказов А. П. Чехова / Под ред. Г. Я. Мартыненко.— СПб., 1999; Частотный словарь языка М. Ю. Лермонтова // Лермонтовская энциклопедия.— М., 1981.— С. 717–774; Алексеев П. Частотный словарь автобиографической трилогии М. Горького.— СПб., 1996.— 205 с.

⁵ Честотен речник на Вапцаровата поезия / Крилова Г. В., Азарова А. А., Захаревич Е. А. и др.— Велико Търново, 1996.

⁶ Balowski M. Lista frekwencyjna poezji, prozy i dramatu Krzystofa Kamila Baczyckiego.— Prohovice, 1997.

⁷ Vasić S. Polazne osnove novije srpske proze. Knj. 2. Hazarski rečnik Milorada Pavića: frekvencijski rečnik.— Beograd, 1998; Vasić S. Polazne osnove novije srpske proze — frekvencijski rečnik romana Mamac Davida Albaharija.— Beograd, 2001; Vasić S. Polazne osnove novije srpske proze — frekvencijski rečnik romana Oslobođioci i izdajnici Milorada Danojlića.— Beograd, 2002; Vasić S. Polazne osnove novije srpske proze — frekvencijski rečnik romana Bezdno Svetlane Velmar-Janković.— Beograd, 2000.

⁸ Словник мови Шевченка : У 2 т. / Ред. В. С. Ващенко.— К., 1964.

⁹ Словник мови творів Г. Квітки-Основ'яненка : У 3 т. / Відп. ред. М. А. Жовтобрюх.— Х., 1978.

до творів І. Котляревського¹⁰, Т. Шевченка¹¹, Лесі Українки¹², Ю. Федьковича¹³, А. Тесленка¹⁴ та інших письменників.

Електронний частотний словник сучасної української поетичної мови (обсягом 300 тис. слововживань) доступний в Інтернеті¹⁵. Він охоплює твори таких українських митців ХХ ст., як Ю. Андрухович, М. Вінграновський, М. Воробйов, І. Драч, І. Жиленко, О. Забужко, С. Йовенко, Т. Коломієць, Л. Костенко, І. Малкович, Б. Олійник, Д. Павличко, В. Стус, Л. Скирда.

Перші засади словникового опису творів І. Франка заклав І. Ковалик¹⁶, зокрема вийшла друком «Лексика поетичних творів Івана Франка»¹⁷, що подає всі лексеми (35 000) поезії письменника за алфавітом із вказівкою на частотність кожної з них. На основі цієї праці виконано низку досліджень, наприклад статистичний аналіз лексики¹⁸, кількісний аналіз антропонімів¹⁹ тощо.

Нині реалізується проект комплексного квантитативного опису творів І. Франка²⁰, що розпочався з етапу кількісної параметризації великої прози письменника. Він передбачає поетапне укладання ЧС романів (повістей) автора (зокрема, вийшов друком ЧС роману «Перехресні стежки»²¹), з подальшим їх зіставленням і виявленням багатства Франкового мовлення, особливостей добору мовних засобів, підтвердженням чи спростуванням статистичних законів, що також вносить уточнення до ідіолекту письменника.

Передбачається, що ЧС І. Франка дасть кількісну характеристику одиницям мовлення письменника, а тим самим детальний та вичерпний перелік лексичних одиниць, що може стати базою для укладання словника його мови тлумачного типу.

Роман «Для домашнього огнища» займає важливе місце у творчості І. Франка. Це єдиний великий прозовий твір, написаний не від номера до номера, а відразу весь — від першого до останнього рядка²². І. Франко створив його спершу по-

¹⁰ Ващенко В. С., Медведєв Ф. П., Петрова П. О. Лексика «Енеїди» І. Котляревського : Показчик слововживання. — Х., 1955. — 207 с.; Бурячок А. А., Залишко А. Т. та ін. Лексика п'єс та од І. П. Котляревського / За ред. А. Бурячка. — К., 1974. — 54 с.

¹¹ Ващенко В. С., Петрова П. О. Шевченкова лексика : Словопоказчик до поезій Т. Г. Шевченка. — К., 1961. — 106 с.

¹² Бойко М. Ф. Словопоказчик драматичних творів Лесі Українки. — К., 1961. — 93 с.

¹³ Юрій Федькович : Словопоказчик мови творів письменника // Гол. ред. К. М. Лук'янюк. — Чернівці, 2004. — 188 с.

¹⁴ Сизько Т. Лексика мови Архипа Тесленка : Словопоказчик оповідань. — Д., 1970. — 101 с.

¹⁵ Частотний словник сучасної української поетичної мови (2003–2008). — www.mova.info/Page2.aspx?11=89.

¹⁶ Ковалик І. І. Принципи укладання Словника мови творів Івана Франка // Українське літературознавство. Іван Франко. Статті і матеріали. — Л., 1968. — Вип. 5. — С. 174–183; Ковалик І. І. Наукові філологічні основи укладання і побудови Словника мови художніх творів Івана Франка // Там же. — 1972. — Вип. 17. — С. 3–10.; Ковалик І. І. Словник мови художніх творів Івана Франка : Пробний зошит // Там же. — 1976. — Вип. 26. — С. 63–99.

¹⁷ Полюга Л. М. Лексика поетичних творів Івана Франка : Метод. вказівки з розвитку лексики // Ковалик І. І., Ощипко І. Й. — Л., 1990. — 264 с.

¹⁸ Полюга Л. Статистичний аналіз лексики поетичних творів І. Франка // Іван Франко і національне відродження. — Л., 1991. — С. 164–166.

¹⁹ Полюга Л. М. До питання про опрацювання української літературної ономастики // Ономастика і апелятиви : Зб. наук. пр. — Д., 1998. — С. 42–47.

²⁰ Бук С. Квантитативна параметризація текстів Івана Франка: спроба проекту // Іван Франко. Студії та матеріали : Зб. наук. статей. — Л., 2010 (у друці).

²¹ Бук С., Ровенчак А. Частотний словник повісті І. Франка «Перехресні стежки» // Стежками Франкового тексту (комунікативні, стилістичні та лексикографічні виміри роману «Перехресні стежки») / Наук. ред. Ф. С. Бацевич. — Л., 2007. — С. 138–369.

²² Мороз М. «Для домашнього огнища» (творча та видавнича історія повісті І. Франка) // Рад. літературознавство. — 1982. — № 4. — С. 30.

льською мовою, перебуваючи у Відні під час славістичних студій у В. Ягича і готуючись до захисту докторату. В. Щурат, який мешкав тоді разом з ним, згадує: «... Франко творив при мені повість “Для домашнього огнища” на основі голосного у Львові судового процесу. Міркуючи, що повість може бути для поляків цікавіша ніж для українців, написав її зразу по-польськи зимою 1892 та й передав її до Варшави через Зигмунда Василевського ...»²³.

У листі до Я. Карловича І. Франко пише: «Я, звичайно, як добрий батько, хотів би це моє дітище добре продати тим більше, що, залишивши роботу у “Kurjer’i Lw[owskim]” і готуючись до захисту докторату, потребую грошенят як для свого перебування у Відні, так і для утримання дружини і дітей у Львові»²⁴, далі просить подати твір до кількох варшавських редакцій, щоб отримати «дещо більший аванс (30–50 %)»²⁵. Існує думка, що саме гостра потреба в грошах змусила І. Франка написати цей твір по-польськи — зі сподіванням надрукувати його у варшавському виданні, де можна розраховувати на «дещо більший аванс»²⁶. Проте польською мовою твору тоді так надруковано й не було — не пройшов цензури.

У вересні 1893 р. І. Франко почав перекладати текст українською. Оскільки «зміна мови передбачає зміну читацької аудиторії, ... певну зміну усього читацького поля загалом»²⁷, то й перекладав роман автор не дослівно, а переорієнтовуючись на українського читача з його характерними смаками. Насамперед написав присвяту О. Мирному, створюючи алузії на його «Повію». Україномовний твір уперше повністю надруковано 1897 р. у серії «Літературно-наукова бібліотека» (кн. 11).

Роман піднімає важливу проблему суспільства²⁸, яка не втратила актуальності й сьогодні. Головна героїня Анеля Ангарович, опинившись із двома дітьми без засобів для існування, коли чоловік-військовий був на війні в Боснії, стає частиною «бізнесу», що займається вербуванням-продажем молодих дівчат у публічні будинки (1992 р. за сюжетом твору знято художній фільм з однойменною назвою режисером Б. Савченком).

ЧС роману укладено за виданням повного зібрання творів І. Франка в 50-ти т. (К., 1979) з урахуванням прижиттєвого видання 1897 р.²⁹ У цій статті квантитативний опис його лексики здійснено вперше.

Як відомо, у другій половині XIX ст. існували різні правописні системи, але жодна з них не була обов’язковою³⁰. Тому в процесі роботи над ЧС роману «Для домашнього огнища» виявлено основні відмінності між текстами вказаних видань. Так, відновлено написання літери *г* у словах *гудзик*, *гешефт*, *цуг*, *Бота мі!* (серб. *Боже мій*), *тетівський*, *легейда*. На відміну від прижиттєвого видання

²³ Щурат В. Франків спосіб творення // Спогади про Івана Франка / Упоряд., вступ. ст. і примітки М. Гнатюка.— Л., 1997.— С. 281.

²⁴ Франко І. До Яна Карловича // Зібрання творів : У 50 т.— К., 1979.— Т. 49.— С. 368 (Тут і далі переклад наш).

²⁵ Там же.

²⁶ Мороз М. Зазнач. праця.— С. 29–30; Тодчук Н. Роман І. Франка «Для домашнього огнища»: простір і час.— Л., 2002.— С. 14.

²⁷ Тодчук Н. Зазнач. праця.— С. 27.

²⁸ Вервес Г. Д. Проблематика польських повістей Івана Франка // Рад. літературознавство.— 1963.— № 3.— С. 64–82.

²⁹ Франко І. Для домашнього огнища.— Л., 1897.— 149 с.

³⁰ Терлак З. Мовно-правописне редагування творів Івана Франка і проблема збереження автентичності авторського тексту // Іван Франко: дух, наука, думка, воля : Матеріали Міжнар. наук. конгресу, присвяченого 150-річчю від дня народження Івана Франка (Львів, 27 вересня — 1 жовтня 2006 р.).— Л., 2010.— Т. 2.— С. 22.

«Перехресних стежок»³¹, через літеру *z*, а не *r* написано слова *грунт*, *гвалт*, *морз* (міра площі) та ін. Можна було очікувати фарингальне *r* у власних назвах *Рейхлінген*, *Големб'я*, проте в першодруці використано фрикативне *z*.

Також виявлено значну кількість інших правописних відмінностей, які, проте, не впливають на статистичну структуру тексту на лексичному рівні: використання на місці сучасної *i* після голосних літери *i* або буквосполучення *ii* (*єі*, *боїть ся*, *ійі*, *прийїде*, *вспокоїти(ся)*, *твоїм*, *другої*, *приїзду*, *Боснії*); написання *i* подекуди на місці сучасного *и* (*систематично*, *трївожились*, *трївоги*, *корідора*) та в род. в. однини іменників жінок р. (*пам'яті*, *осени*, *хвили*); використання суфікса та закінчення наз. в. іменників середнього роду *-н(є)*, *-т(є)*, *-с(є)* на місці сучасного *-н(я)*, *-т(я)*, *-с(я)* (*чутє*, *прочутє*, *бажанє*, *розчаруванє*, *привитанє*, *волосє*, але *зворушеня*) та ін.

Зафіксовано також відмінності, які впливають на кількісні характеристики лексем: написання зворотної дієслівної частки *-ся* окремо (а *-сь* — разом), слова *нів* та частки *-небудь* — окремо (*нів години*, *що небудь*), частки *ж* — разом (*робиж*), складених прислівників типу *раз-у-раз* — через дефіс; написання сучасних прислівників, що починаються колишніми прийменниками, окремо (*поперед*, *з близька*, *на ліво*, *що хвиля*, *з далека*, *до дому*) тощо. Цікаво, що частка *-ся* у творі трапляється 1285 разів, це друге (!) за частотою значення після *i / ii*.

Привертає увагу, що в оригінальному тексті польське прізвище *Szablińska* фонетично наближене до української вимови: *Шаблінська*, тоді як у виданні 1979 р. подано *Шаблінська*, а ім'я головної героїні *Aniela* в оригіналі наближене до польського звучання: *Анєля*, на протизвагу *Анєля* в 50-томнику. За винятком наведених різнописань тексти видань загалом ідентичні.

Роман «Для домашнього огнища» — другий за порядком твір письменника, лексика якого отримала комплексну кількісну характеристику у вигляді частотного словника. ЧС роману «Для домашнього огнища» укладено на підставі лінгвостатистичного аналізу тексту твору і являє собою впорядкований список слів, забезпечений даними про частоту їх уживаності в тексті. Окремим словом вважаємо послідовність літер (тут апостроф і дефіс розглядаються як літера) між двома пропусками чи розділовими знаками, тому складні числівники виступали як різні слова. ЧС твору подає інформацію про словникові одиниці (леми) і про словоформи: парадигматичні форми і фонетичні варіанти слів.

Формування частотного словника здійснено за графічним збігом лем, і кожна частина мови мала свою схему об'єднання словоформ під лемою: аналогічну, як і в частотних словниках художньої прози³², публіцистики³³, розмовно-побутового³⁴, наукового³⁵ та офіційно-ділового³⁶ стилів.

Іменник — до називного відмінка однини зведено форми всіх відмінків однини та множини. Частоту множинних іменників зведено до форми називного відмінка множини (*причандали*, *условини*, *зостанки*). Оскільки такі форми імен-

³¹ Франко І. Перехресні стежки : Повесть.— Л., 1900.— 320 с. (на обкладинці вказано 1899 рік).

³² Частотний словник сучасної української художньої прози / За ред. В. С. Перебийніс.— К., 1981.— Т. 1.— 863 с.; Т. 2.— 855 с.

³³ Дарчук Н. П., Грязнухіна Т. О. Значч. праця.

³⁴ Бук С. 3000 найчастотніших слів розмовно-побутового стилю сучасної української мови / Наук. ред. Ф. С. Бацевич.— Л., 2006.— 180 с.

³⁵ Там же.— 192 с.

³⁶ Бук С. Частотний словник офіційно-ділового стилю: принципи укладання і частотні характеристики // Лінгвістичні студії.— 2006.— Вип. 14.— С. 184–188.

ників, як *домінікани*, *єзуїти*, *турки* і їм подібні позначають осіб і чоловічої, і жіночої статі, їх не було зведено до чоловічого роду однини.

Прикметник — до називного відмінка однини чоловічого роду зведено відмінкові форми всіх родів у однині та множині, вищій і найвищій ступені порівняння, за винятком суплетивних форм, які зведено окремо до називного відмінка однини чоловічого роду вищого ступеня, напр.: *більший*, *найбільший* зведено до *більший*. Така традиція усталилася в більшості словопоказчиків і частотних словників української мови.

Займенник та числівник — зведено відмінкові форми відповідно до типу відмінювання.

Дієслово — до інфінітива зведено синтетичні форми часу (теперішній, минулий і майбутній), форми наказового способу і дієприслівник, а також безособові форми на *-но*, *-то*. Аналітичні форми часу ми вважали синтаксичними утвореннями, кожний складник яких зареєстровано як окреме слово.

Дієприкметник — до називного відмінка однини чоловічого роду зведено відмінкові форми всіх родів у однині та множині, оскільки розглядаємо його як різновид віддієслівного прикметника із властивими йому основними категоріями (рід, число, відмінок) та типовою синтаксичною роллю означення³⁷.

Прислівник — зведено вищій і найвищій ступені порівняння, за винятком суплетивних форм.

Лематизацію слів з частками *-но*, *-таки*, *-то*, слововживань з якими в тексті роману налічується кілька десятків, реалізовано таким чином: змінні частини мови зведено до початкових форм (*розповідж-но* до *розповідсти*), у незмінних ці частки збережено (*тим-то*, *якось-то*, *якби-то*, *все-таки*, *нарешиті-таки*, *зовсім-таки*...).

До початкової форми зведено фонетичні варіанти слів, де чергування початкових або кінцевих літер пов'язане з милозвучністю мови, а саме: дієслова з постфіксами *-ся* / *-сь*; сполучники *щоб* / *щоби*, *і* / *й*; частки *ж* / *же*, *б* / *би*, *ще* / *іще*; прийменники *у* / *в*, *з* / *із* / *зі* / *зо* та деякі інші; слова з відповідними префіксами (*вложити* / *уложити*, *весь* / *увесь* / *ввесь*, *всякий* / *усякий*), а також *вулиця* / *улиця*, *вуха* / *уха*.

Не зводилися до єдиного слова такі: *влюблений* / *улюблений*, *вклад* / *уклад*, *впадати* / *упадати*, оскільки вони мають різні значення:

...мов який *влюблений* хлопчина ... = *закоханий*;
...не доторкнувся своєї *улюбленої* легоміни ... = *який подобається більше, ніж інші*;
Твоя пенсія наразі буде йти на *вклади* в господарство = *внески*;
...цілим вихованням, цілим життям, *укладом*, ... попсовано її етичні основи = *встановлений порядок, що склався у житті, побуті, родині*;
І чого ж тут гніватись і в пафос *впадати*? = *пафосувати* (дія за значенням іменника);
Впадала на різні здогади ... = *здогадувалася* (дія за значенням іменника);
Супроти рішучого зізнання сих дівчат усяке підозріння супроти вашої жінки *упадає* = *зникає*.

В одній словниковій статті подано слова *ледве* / *ледво*, *трохи* / *троха*, *стільки* / *стілько* і под., оскільки в 50-томному виданні творів І. Франка авторську форму (*троха*, *стілько* ...) залишено в мовленні персонажів, а сучасну літературну форму (*трохи*, *стільки* ...) подано в інших випадках³⁸.

³⁷ Вихованець І. Р. Частиномовний статус дієприкметників // Українська мова.— 2003.— № 2.— С. 54–60.

³⁸ Від редакційної колегії // Франко І. Зібрання творів : У 50 т.— К., 1979.— Т. 1.— С. 14–15.

У словнику розмежовано лексичну та лексико-граматичну омонімію методом контекстуального аналізу: *a* (сполучник / частка / вигук), *раз* (прислівник / іменник / сполучник), *та* (займенник / сполучник / частка), *так* (частка / прислівник / сполучник), *то* (сполучник / частка / займенник), *волів* (іменник / дієслово), *дні* (наз. в. мн. від *день* і род. в. одн. від *дно*) тощо. В омографах вказано наголос: *мук* і *м*а, *в*кликати і *виклик*ти тощо.

У ЧС для розрізнення омонімів у дужках подано або вказівку на значення — *ніч*до (присудк. сл.), або на частиномовну належність — *мати* (і.м.) і *мати* (дієсл.).

Додаток у кінці статті містить першу сотню найчастотніших лем (слів, зведених до словникової форми). Очікувано високий ранг мають, крім службових частин мови, назви персонажів *капітан*, *Анеля*, *Редліх*, *Юлія*; дієслова *бути*, *могти*, *мати*, *знати*; іменники *пан*, *пані*, *рука*, *діти*.

Наведемо деякі результати статистичних підрахунків:

1. Обсяг роману (N) — 44 840 слововживань.
2. Кількість різних словоформ (V_{Φ}) — 11 505.
3. Кількість різних слів (V) — 6486.
4. Багатство словника (індекс різноманітності), тобто відношення обсягу словника лексем до обсягу тексту (V/N), становить 0,145.
5. Середня повторюваність слова в тексті, тобто відношення обсягу тексту до обсягу словника лексем (N/V), — 6,9. Інакше кажучи, кожне слово в середньому вжито в досліджуваному тексті приблизно 7 разів.
6. Кількість *h*арак *l*еgоmеnа — слів із частотою 1 — ($V_1 = 3363$) становить 7,5% тексту і 51,85% (понад половину!) словника.
7. Індекс винятковості в тексті, тобто відношення V_1 до обсягу тексту (V_1/N), — 0,075.
8. Індекс винятковості у словнику, тобто відношення V_1 до обсягу словника (V_1/V), становить 0,52. Числа в пунктах 7, 8 — показники варіативності лексики.
9. Протилежним до індексу винятковості є індекс концентрації — відношення кількості слів з високою частотою (10 і вище, їх у романі — 598) до обсягу тексту ($V_{10}/N = 0,752$) або словника ($V_{10}/V = 0,092$).

Відносно невелика кількість високочастотної лексики (і, відповідно, низький індекс концентрації) та порівняно велика кількість слів із частотою 1 (і, відповідно, високий індекс винятковості) свідчать про неабияку різноманітність лексики роману.

Цікаво порівняти кількісні дані двох романів І. Франка — «Для домашнього огнища» та «Перехресні стежки» на основі їхніх ЧС. Зіставлення характеристик різних текстів має сенс лише у випадку приблизно однакової кількості слововживань в аналізованих матеріалах, оскільки збільшення обсягу словника відбувається не пропорційно до збільшення обсягу тексту, а дещо повільніше³⁹. Оскільки обсяг роману «Для домашнього огнища» більш як удвічі менший за обсяг «Перехресних стежок», то, наприклад, вищий індекс різноманітності його лексики чи більша частка *h*арак *l*еgоmеnа у словнику є очікуваними явищами. Ця проблема потребує окремого детальнішого дослідження. У табл. 1 наведено зіставлення деяких параметрів двох творів. Для коректного порівняння подано результати розрахунків на матеріалі перших 44 840 слововживань роману «Перехресні стежки», що відповідає обсягові роману «Для домашнього огнища».

³⁹ Частотний словник сучасної української художньої прози.— К., 1981.— Т. 1.— С. 14; Тулдава Ю. П. Зазнач. праця.— С. 62–64.

Таблиця 1

Зіставлення кількісних характеристик романів І. Франка «Для домашнього огнища» та «Перехресні стежки»			
Кількісні характеристики	Для домашнього огнища	Перехресні стежки (перші 44 840 слововживань)	Перехресні стежки (весь текст)
Обсяг роману (N)	44840	44840	93888
Кількість різних словоформ (V_{Φ})	11505	11506	19448
Кількість різних слів (V)	6486	6561	9978
Багатство словника (індекс різноманітності)	0,145	0,146	0,106
Середня повторюваність слова в тексті (N/V)	6,9	6,8	9,41
Кількість нарах legomena (V_1)	3363	3586	4907
Індекс винятковості у тексті (V_1/N)	0,075	0,080	0,0523
Індекс винятковості в словнику (V_1/V)	0,519	0,547	0,492
Індекс концентрації тексту (V_{10}/N)	0,752	0,731	0,795
Індекс концентрації словника (V_{10}/V)	0,092	0,087	0,113

Як видно з таблиці, для текстів І. Франка однакового розміру основні квантитативні показники фактично збігаються, а саме: кількість різних словоформ та слів, багатство словника, середня повторюваність слова в тексті тощо. Дещо відрізняються лише кількість нарах legomena (і через це — індекс винятковості в тексті та словнику), а також індекс концентрації тексту і словника.

Визначальною ознакою ЧС є те, що слова в ньому розміщують у порядку спадання частоти, тобто слово з найбільшою частотою має ранг 1, наступне — 2 і т. д. Таке подання інформації дає змогу обчислити, яку частку тексту (покриття) становлять слова з найбільшою частотою. Величину покриття тексту для певного рангу обчислюють як відношення суми абсолютних частот усіх слів з меншими рангами до загальної кількості слів у тексті. Співвідношення між рангом слова та покриттям тексту подано у табл. 2 і на рис. 1.

Таблиця 2

Співвідношення між рангом слова та величиною покриття тексту					
Ранг	Покриття	Ранг	Покриття	Ранг	Покриття
1	3,33	200	58,83	1500	83,99
5	11,78	300	63,87	2000	87,39
10	18,88	400	67,45	3000	91,96
25	31,13	500	70,25	4000	94,46
50	40,19	600	72,56	5000	96,69
75	45,65	750	75,39	6000	98,92
100	49,59	1000	79,01	6484	100,0

З табл. 2 видно, що перші за частотністю 25 слів покривають 31% тексту, перші 100 слів — 50%, 1000 слів — 79% тексту.

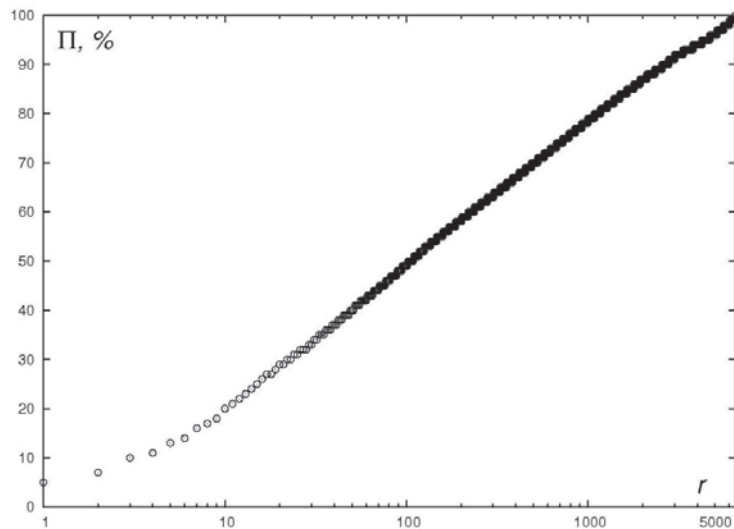


Рис. 1. Залежність величини покриття тексту (П) від рангу лема (r).

ЧС роману І. Франка «Для домашнього огнища», за класифікацією частотних словників, наведеною В. Перебийніс⁴⁰, є 1) за одиницями підрахунку — частотним словником слів та словоформ; 2) за обсягом вибірки — невеликим; 3) за характером вибірки — частотним словником конкретного твору; 4) за обсягом — повний; 5) за характером подання матеріалу — алфавітно-частотним і частотно-алфавітним; 6) за статистичними характеристиками одиниць — частотним словником абсолютної та відносної частоти без статистичних її оцінок.

Отже, ЧС роману дав змогу подати важливі статистичні характеристики тексту, такі як обсяг роману, кількість різних словоформ та слів, індекс різноманітності (багатство словника), середня повторюваність слова в тексті, кількість нарах *legomena*, індекс винятковості в тексті та у словнику, індекс концентрації, співвідношення між рангом слова та величиною покриття тексту. Крізь призму ЧС лексичний склад роману отримує нову інтерпретацію.

Таблиця 3

Перша сотня найчастотніших лем роману І. Франка «Для домашнього огнища»				
Ранг	Лема	Абс. част.	Відн. част., %	Покр., %
1	і—1393; и—101	1494	3,33	3,33
	-ся	1285	2,87	
2	він	1020	2,27	8,47
3	не	1000	2,23	10,70
4	в—728; у—241	969	2,16	12,86
5	я	800	1,78	14,65
6	з—620; із—54; зі—33; зо—13	720	1,61	16,25
7	на	703	1,57	17,82
8	той	589	1,31	19,13
9	сей	588	1,31	20,45
10	бути	585	1,30	21,75
11	що (займ.)	577	1,29	23,04
12	до	543	1,21	24,25
13	вона	529	1,18	15,43

⁴⁰ Перебийніс В. С. Частотний словник // Українська мова : Енциклопедія / Редкол.: В. М. Русанівський, О. О. Тараненко (співголови) та ін.— К., 2000.— С. 790.

14	що (спол.)	529	1,18	26,61
15	капітан	489	1,09	27,70
16	а (спол.)	478	1,07	28,76
17	ти	361	0,81	29,57
18	свій	276	0,62	30,19
19	ж—245; же—25	270	0,60	30,79
20	вони	255	0,57	31,36
21	могти	246	0,55	31,90
22	весь—154; увесь—87; ввесь—1	242	0,54	32,44
23	Анеля (ім'я)	236	0,53	32,97
24	такий	233	0,52	33,49
25	та (спол.)	226	0,50	33,99
26	знати	206	0,46	34,45
27	за	202	0,45	34,90
28	коли	202	0,45	35,35
29	би—162; б—39	201	0,45	35,80
30	про	198	0,44	36,24
31	мати (дієсл.)	188	0,42	36,66
32	якийсь	186	0,41	37,08
33	себе	183	0,41	37,49
34	ми	171	0,38	37,87
35	вже—120; уже—49	169	0,38	38,24
36	але	168	0,37	38,62
37	так (присл.)	164	0,37	38,99
38	мовити	161	0,36	39,34
39	ще—143; іще—18	161	0,36	39,70
40	який	154	0,34	40,05
41	ми	148	0,33	40,38
42	по	144	0,32	40,70
43	сам	137	0,31	41,00
44	як (спол.)	136	0,30	41,31
45	пан	135	0,30	41,61
46	від	134	0,30	41,91
47	щоби—110; щоб—24	134	0,30	42,21
48	тут	130	0,29	42,50
49	один	126	0,28	42,78
50	рука	124	0,28	43,05
51	тепер	124	0,28	43,33
52	там	122	0,27	43,60
53	хвиля	121	0,27	43,87
54	око	118	0,26	44,13
55	раз (ім.)	112	0,25	44,38
56	ні	111	0,25	44,63
57	діти (ім.)	103	0,23	44,86
58	Редліх (прізви.)	103	0,23	45,09
59	іти—42; йти—59	101	0,23	45,32
60	ви	100	0,22	45,54
61	перед (прийм.)	100	0,22	45,76
62	хотіти	100	0,22	45,99
63	жінка	99	0,22	46,21

64	для	97	0,22	46,72
65	слово	94	0,21	46,63
66	адже (част.)	93	0,21	46,84
67	котрий	91	0,20	47,04
68	то (спол.)	88	0,20	47,24
69	лице	86	0,19	47,43
70	сказати	86	0,19	47,62
71	ну	84	0,19	47,81
72	душа	81	0,18	47,99
73	голова	79	0,18	48,17
74	бачити	78	0,17	48,34
75	мусити	78	0,17	48,51
76	ніщо	78	0,17	48,69
77	тільки	77	0,17	48,86
78	щось (займ.)	77	0,17	49,03
79	то (част.)	76	0,17	49,20
80	чи (част.)	75	0,17	49,37
81	навіть	74	0,17	49,53
82	пані	74	0,17	49,70
83	зовсім	73	0,16	49,86
84	при	73	0,16	50,02
85	ані	72	0,16	50,19
86	без	71	0,16	50,34
87	дуже	71	0,16	50,50
88	Юлія (ім'я)	71	0,16	50,66
89	кілька	70	0,16	50,82
90	говорити	69	0,15	50,97
91	може	69	0,15	51,12
92	скрикнути	69	0,15	51,28
93	або (спол.)	68	0,15	51,43
94	мов	68	0,15	51,58
95	думка	67	0,15	51,73
96	тільки	66	0,15	51,88
97	голос	65	0,14	52,02
98	так (част.)	65	0,14	52,17
99	хто	65	0,14	52,31
100	час	65	0,14	52,46
101	знов	64	0,14	52,60
102	твій	64	0,14	52,74
103	чи (спол.)	64	0,14	52,89

(Львів)

S. N. BUK

IVAN FRANKO'S NOVEL «DLJA DOMASHNJOHO OHNYSHCHA» («FOR THE HEARTH») FROM THE FREQUENCY DICTIONARY PERSPECTIVE

In the article, the methodology and the principles of the compilation of the Frequency dictionary for Ivan Franko's novel «Dlja domashnjoho ohnyshcha» («For the Hearth») are described. The following statistical parameters of the novel vocabulary are obtained: variety, exclusiveness, concentration indexes, correlation between word rank and text coverage, etc. The main quantitative characteristics of Franko's novels «Perekhresni stezhky» («The Cross-Paths») and «Dlja domashnjoho ohnyshcha» are compared on the basis of their frequency dictionaries.

Key words: word frequency, frequency dictionary, rank, hapax legomena, variety, exclusiveness, concentration indexes, text coverage.