

**Валентина Ісидорівна (Сидорівна) Перебийніс  
(13.X.1923–21.XII.2020)**



**УДК 81'33=161.2(075)**

*Євгенія Карпіловська*

**Зачинателька української статистичної лексикографії:**

До 100-ліття від дня народження  
доктора філологічних наук, професора  
Валентини Ісидорівни (Сидорівни) Перебийніс  
(13.X.1923–21.XII.2020)

*У статті з нагоди 100-ліття від дня народження доктора філологічних наук, професора Валентини Ісидорівни (Сидорівни) Перебийніс проаналізовано укладений під її керівництвом новий для української лексикографії тип словника – «Частотний словник сучасної української художньої*

прози» в 2 т. (Київ, 1981) та осмислено роль Ювілярки в заснуванні і розвитку нової галузі українського словникарства – статистичної лексикографії. Продемонстровано принципи укладання частотного словника, підкреслено його значення для вивчення закономірностей і особливостей функціонування українського лексикону в текстах різних стилів.

**Ключові слова:** Валентина Ісидорівна (Сидорівна) Перебийніс, українська лексикографія, статистична лексикографія, частотний словник

13 жовтня 2023 року українці мовознавці відзначили 100-літній ювілей від дня народження доктора філологічних наук, професора, першої завідувачки відділу структурно-математичної лінгвістики Інституту мовознавства ім. О. О. Потебні АН УРСР (нині – НАН України), фундаторки разом із М. М. Пещак школи академічної структурної, прикладної та математичної лінгвістики Валентини Ісидорівни (Сидорівни) Перебийніс. Упродовж усього свого довгого й насиченого професійного життя Валентина Ісидорівна, германіст за базовою освітою і дипломом кандидата філологічних наук, дослідження англійської мови поєднувала з опрацюванням за допомогою структурних і математичних методів, комп'ютерних технологій фонетики, лексики і граматики української мови. Її докторська дисертація «Кількісні та якісні характеристики системи фонем сучасної української літературної мови» і численні праці з україністики є тому переконливим підтвердженням. Вона разом із колективом співробітників відділу структурно-математичної лінгвістики Інституту мовознавства ім. О. О. Потебні, який очолила в 1968 році, була зачинателькою нового напрямку в українській лексикографії – статистичної лексикографії, першою ластівкою якої став двотомний «Частотний словник сучасної української художньої прози» (далі – ЧС), що побачив світ у київському академічному видавництві «Наукова думка» 1981 року.

У тодішню ще докомп'ютерну еру цю працю уклали вручну. Огром обчисленого матеріалу вражає: понад 500 тисяч слів текстів були вручну обстежені й обраховані за допомогою спеціально створеної для цього нового типу словника системою статистичних методик. Обсяг текстів для підрахунків визначила реєстрова одиниця цього частотного словника української мови – слово. Фахівці з лінгвостатистики вже на той час встановили, що для одержання вірогідних частотних характеристик слова, тобто відомостей про його вживання в тексті та розподіл у ньому, потрібний масив слів, не менший за 300 тисяч одиниць. Обсяг у 500 тисяч слововживань Валентина Ісидорівна та її колеги прийняли за оптимальний. Прикметно, що як повноправні реєстрові одиниці до словника були введені дієприслівники, а також численні слова-юкстапозити на зразок *батько-герой* або *байдужо-спокійний*. Джерельну базу ЧС становили художні прозові твори 25 українських письменників, опубліковані в період між

1945–1970 рр. Дібрано було тексти різних жанрів, що представляли мовну практику різних регіонів України, зокрема твори Ірини Вільде, А. Головка, О. Гончара, О. Довженка, П. Загребельного, Івана Ле, Петра Панча, М. Стельмаха та ін. Було відібрано твори майстрів сучасної української прози, різноманітні за будовою та тематикою, письменників, мовну палітру котрих дослідники вважали показовою і зразковою для стану української літературної мови того періоду. Отже, ЧС є повним, середнім за обсягом реєстру частотним словником слів, для яких встановлені такі статистичні характеристики: абсолютна частота вживання в загальній вибірці, абсолютна й відносна частота вживання окремо в авторській та прямій мові (мові персонажів творів), середня частота вживання у вибірках у 1000 слів, показник міри коливання середньої частоти у вибірках з різних творів, показник кількості джерел, у яких зустрілася така реєстрова одиниця, та кількості мінімальних вибірок, у яких встановлювалася середня частота вживання слова. Для кожної реєстрової одиниці в ЧС указано її частиномовну належність, для багатозначних слів та омонімів – значення, в якому для нього властивий певний набір форм, що зафіксовані в обстежених текстах, а для змінюваних слів подані показники граматичних категорій у межах частин мови, незмінювані слова позначені символом \* (астериск, або зірочка).

Корпус ЧС складають 5 списків: 1) слова з абсолютною частотою вживання, яка дорівнює 2 і більше; 2) словоформи з абсолютною частотою вживання 1; 3) слова та словоформи, упорядковані за спадом частоти вживання (загальний список), 4) слова та словоформи, упорядковані за спадом частоти вживання в прямій та 5) в авторській мові. Перші два списки становлять так звані *алфавітно-частотні списки* слів, принципом побудови яких є розміщення одиниць в алфавітному порядку за їхнім початком. Три інших списки називають у статистичній лексикографії *ранговими*: зі спадом частоти слова в таких списках збільшується його ранг, або порядковий номер у такому списку. У першому списку для змінюваних слів встановлено вихідну форму, під якою об'єднано всі зафіксовані в текстах словоформи. Отже, ЧС дає його читачам уявлення не лише про ті слова, які вжито в опрацьованих джерелах, а й про обсяг їхніх уживаних форм. Наприклад, уживана в досліджених текстах словозмінна парадигма слова *українець* виглядає так: абсолютна частота його вживання дорівнює 18 (11 у прямій і 7 – в авторській мові); у текстах його представляють такі форми – *українець* (4 рази), *українца* (3), *українці* (називний відмінок множини, 5 разів), *українців* (родовий відмінок множини – 4 рази), *українців* (знахідний відмінок множини 1 раз), *українцями* (1 раз).

Загалом у ЧС подано статистичні характеристики близько 33 тис. різних слів сучасної української мови. Серед них чимало рідкісних слів: okazionalizmів, neologizmів, індивідуально-авторських слововживань на

зразок *аеродромчик, арифметик, безсердеччя, вертиголова, кулацюра, безвечірній, білоштанний, біло-спокійний, виразно-розмірений* та ін. Трапляються серед них діалектизми (*банякуватий, варгатий*), історизми (*амант, адамасковий*) і навіть слова-покручі – кальки з російських зразків (*авоська* (пор. із нормативними *сітка, сіточка*), *балбес* (норм. *бовдур, телепень*), *кукушка* (норм. *зозуля*), *нерозбериха* (норм. *безлад, плутанина*), *бандпосібник* (норм. *бандпомічник, бандпомагач*). Поява кальок та покручів у реєстрі ЧС зумовлена суцільною вибіркою слів для його реєстру, засадничим принципом формування реєстру цього словника. Тексти подано без жодних лакун, а подібні слова у творах відіграють роль стилістичних маркерів мови тих чи тих персонажів.

Приписаний кожному слову ранг, або ж його місце в ієрархії одиниць системи мови, реалізованої в текстах-джерелах ЧС, свідчить про ступінь активності їхнього вживання в дібраних для ЧС творах сучасної української художньої прози. Цілком очікувано такий ранговий список очолили слова службових частин мови, а також особові займенники. Ось які слова в ЧС виявилися найуживанішими, опинившись на перших 10 місцях (рангах) цього списку: 1 – *i<sub>1</sub> сп. (сполучник)* – 14276 (цифра зліва від слова – показник його рангу в частотному списку, а цифра справа – показник абсолютної частоти його вживання в опрацьованому текстовому масиві у 500 тис. слововживань); 2 – *в* – 12791; 3 – *не* – 11030; 4 – *на* – 10388; 5 – *з* – 8338; 6 – *він* – 8200; 7 – *a<sub>1</sub> сп. (сполучник)* – 6251; 8 – *до* – 4821; 9 – *я* – 4783 і 10 – *i<sub>2</sub> част. (частка)* – 4666. Найвищі ранги серед дієслів мають лексема *бути* у функції зв'язки в складеному дієслівному присудку (14-ий ранг, абсолютна частота вживання 2874) і у функції повнозначного слова (22-ий ранг, абсолютна частота вживання 2258), а також *сказати* (43-ій ранг, абсолютна частота вживання 1337) і *знати* (44-ий ранг, абсолютна частота вживання 1323). Серед іменників найактивнішими в досліджених текстах виявилися слова *рука* (33-ій ранг, абсолютна частота вживання 1587) та *око* (41-ий ранг, абсолютна частота вживання 1407).

Ось, наприклад, яку інформацію можна почерпнути зі статті до слова *думка* в ЧС. Загалом форми словозмінної парадигми цього іменника жіночого роду зустрілися в опрацьованих текстах 445 разів, відповідно 82 рази в прямій та 363 рази в авторській мові. Окремі форми цього слова, за свідченням ЧС, побутують в аналізованих текстах із такими показниками абсолютної частоти вживання, тобто вживання і в прямій, і в авторській мові: *думка* – 63; *думки род одн.* – 42; *думку* – 71; *думкою* – 19; *думці місц.* – 26; *думки наз. мн.* – 61; *думок* – 44; *думкам* – 3; *думки зн. мн.* – 47; *думками* – 37; *думках* – 32. Відносна частота найактивнішої в цій парадигмі форми називного відмінка однини *думка* в опрацьованому масиві художніх прозових текстів становить 0,000126, або 63/500000, де, нагадаємо, 500 000

слововживань – довжина текстової вибірки, в якій встановлено частотні характеристики слів реєстру ЧС.

Обраховані частотні характеристики слів створюють надійну основу для визначення ядра, найбільш активних і значущих слів для текстів української художньої прози 40–70-х років минулого століття, дають можливість визначити обсяг, склад і закономірності реалізації в цих текстах українського лексикону. «Частотний словник сучасної української художньої прози» подає «портрет» цього функціонально-стильового різновиду української літературної мови з погляду кількісних та якісних характеристик їхнього лексичного наповнення.

1986 р. Валентина Ісидорівна Перебийніс перейшла на роботу до Київського інституту іноземних мов (згодом – Київського національного лінгвістичного університету), де створила й очолила лабораторію комп'ютерної лінгвістики, але не поривала зв'язків з відділом структурно-математичної лінгвістики. 1998 р. на базі ЧС з прямим абеткуванням одиниць реєстру, або абеткуванням за початком слів колектив співробітників відділу структурно-математичної лінгвістики Інституту мовознавства ім. О. О. Потебні НАН України на чолі з Н. Ф. Клименко, учнів і колег В. І. Перебийніс, уклав за допомогою комп'ютера й видав однотомний «Обернений частотний словник сучасної української художньої прози» (Київ, 1998), який становить один із можливих граматичних словників сучасної української мови. У цьому частотному словнику, на відміну від попереднього двотомника, подано розгорнуту граматичну інформацію про особливості вживання та сполучуваності слів у прямій та авторській мові творів української художньої прози другої половини ХХ ст.

Методика статистичного опрацювання текстів та формат статті «Частотного словника сучасної української художньої прози» були використані в комп'ютерних частотних словниках публіцистичного та наукового стилів української мови, які співробітники відділу структурно-математичної лінгвістики Інституту мовознавства ім. О. О. Потебні НАН України разом із працівниками лабораторії комп'ютерної лінгвістики кафедри української мови та прикладної лінгвістики Київського національного університету імені Тараса Шевченка уклали в 90-х роках минулого століття. Цю роботу в лабораторії комп'ютерної лінгвістики продовжила учениця В. І. Перебийніс доктор філологічних наук, доцент Н. П. Дарчук, уклавши разом зі студентами, майбутніми бакалаврами та магістрами з прикладної лінгвістики, «Частотний словник сучасної української поетичної мови» на матеріалі творів Ю. Андруховича, М. Вінграновського, О. Забужко, І. Драча, І. Жиленко, С. Йовенко, Л. Костенко, І. Малковича, Д. Павличка, В. Стуса та інших українських письменників кінця ХХ – початку ХХІ ст. Ці словники, як і базовий для них «Частотний словник сучасної

української художньої прози», виставлені у вільному доступі на лінгвістичному порталі mova.info

Після науковця, педагога, організатора наукових досліджень у спадок наступним поколінням залишаються його праці і учні, які продовжують справу свого Вчителя. Валентина Ісидорівна Перебийніс прожила в науці щасливе життя. Вона реалізувала свої задуми і створила школу академічної структурної, прикладної та математичної лінгвістики в Україні, відкрила зокрема новий – статистичний – напрямок в українській лексикографії. Уже третє покоління її вірних учнів активно й плідно продовжує її справу в українському мовознавстві, розвиває ідеї статистичної лексикографії.

### Джерела

Обернений частотний словник сучасної української художньої прози. Київ: АртЕк, 1998.

ЧС – Частотний словник сучасної української художньої прози. Київ: Наук. думка, 1981. Т. 1-2.

Частотний словник сучасної української поетичної мови: URL: [www.mova.info](http://www.mova.info) (дата звернення 20.01.2024).

### Legend

ЧС – Frequent dictionary of Modern Ukrainian Artistic Prose. Kyiv: Nauk. dumka, 1981. Vol. 1-2.

Converse Frequent dictionary of Modern Ukrainian Artistic Prose. Kyiv: ArtEk, 1998.

Frequent dictionary of Modern Ukrainian Poetic Language: URL: [www.mova.info](http://www.mova.info) (date of request 20.01.2024).

*Ievgeniia Karpilovska*

***The founder of Ukrainian statistical lexicography: To the 100th anniversary of the birthday of Doctor of Philology, Professor Valentina Isidorivna (Sydorivna) Perebyinis (13.X.1923–21.12.2020)***

*In the article on the occasion of the 100th anniversary of the birthday of the Doctor of Philology, Professor Valentina Isidorivna (Sydorivna) Perebyinis, a new for Ukrainian lexicography type of dictionary – "Frequent Dictionary of Modern Ukrainian Artistic Prose" in 2 volumes (Kyiv, 1981) and the role of Professor Perebyinis in the foundation and development of the new field of Ukrainian lexicography – statistical lexicography is discussed. The principles of compiling frequency dictionary are demonstrated, as well as importance such*

*dictionaries for studying the regularities and peculiarities of the functioning of the Ukrainian lexicon in texts of different styles is emphasized.*

**Key words:** *Valentina Isydorivna (Sydorivna) Perebyinis, Ukrainian lexicography, statistical lexicography, frequent dictionary*