

УДК 811.161.2'324

СТАТИСТИЧНІ ХАРАКТЕРИСТИКИ ПРИВАТНОЇ КОРЕСПОНДЕНЦІЇ: ПАРАМЕТРИ, ПОВ'ЯЗАНІ З ДОВЖИНОЮ СЛОВА

Ольга ГУМЕНЧИК

*Львівський національний університет імені Івана Франка,
кафедра загального мовознавства,
вул. Університетська, 1, ауд. 343, 79000 Львів, Україна
тел.: +380 (032) 2394756*

Досліджено питання стильової належності приватної кореспонденції за допомогою таких статистичних критеріїв як частка чотирискладових слів та середнє квадратичне відхилення довжини слова у складах.

Ключові слова: епістолярний жанр, приватний лист, статистичні методи, частка чотирискладових слів, середнє квадратичне відхилення довжини слова у складах.

В українському мовознавстві тривають дослідження функціонально-стильової різноманітності сучасної української мови. Поряд із цим вивчають структурну граматику української мови, проводять статистичні дослідження закономірностей функціонування системи мови (різних її рівнів) у функціональних стилях мовлення, а також аналіз фонемної, морфемної, словотворчої, семантичної структури слова в слов'янських, зокрема в українській мові.

Стилістикою української мови займалися П. Житецький, О. Курило, Л. Булаховський, О. Матвієнко, О. Фінкель, І. Білодід, І. Чередниченко, Г. Їжкевич, З. Франко. Проблемою мовленнєвих жанрів цікавився М. Бахтін [2, с. 428–472].

Над проблемою статусу епістолярного мовлення працюють вже досить давно, однак цілковитої єдності поглядів на його місце у системі функціональних стилів досі не існує. Думки вчених розходяться. Деякі лінгвісти, зокрема І. Чередниченко, С. Єрмоленко, О. Гвоздев та інші, не виділяють епістолярний стиль. У монографії «Нариси з загальної стилістики сучасної української мови» І. Чередниченко кваліфікує його як різновид або писемну форму *розмовно-побутового* стилю. Листування між літературними персонажами, що трапляється в художніх творах, він зараховує до художніх текстів.

Прихильники іншого погляду (В. Наєр, Т. Радзівєвська, С. Гіндін та інші) пропонують визначити епістолярне мовлення як «міжстильове явище, здатне існувати в межах чи на основі того чи іншого функціонального стилю, або поза їх межами» [5, с. 45]. О. Єфімова, Л. Мацько, К. Ленець визначають *епістолярне мовлення* як окремий функціональний стиль. Л. Кеца усі листи поділяє на *приватні й ділові* [3, с. 94–99]. *Приватні*, своєю чергою, охоплюють *листи побутові*, написані *розмовним стилем*, і *листи письменника*, написані *художнім стилем*, а все це разом утворює *епістолярний стиль*. Інші науковці роблять спробу розгляду *приватної кореспонденції* крізь призму жанру: *епістолярного* (Т. Зоріна), *лінгвістичного* (В. Ляпунова), *літературного* (О. Єліна) [6, с. 241–248].

Однак, більшість учених сходяться на думці, що *епістолярне мовлення* належить до міжстильових жанрів. Адже *епістолярій* досить точно відбиває розмовно-

побутове мовлення, містить у собі художнє та елементи офіційно-ділового мовлення. Листування – явище «багатостильове», знаходить своє відбиття в елементах усіх інших функціональних стилів. Тому доцільніше виділити *епістолярний жанр*, що сформувався на основі різностильових елементів, вибудованих за відповідними структурно-композиційними моделями [6, с. 241–248].

Оскільки *епістолярне мовлення* є міжстильовим жанром, яке характеризується змішуванням стилістично різнорідних елементів, оформлених за певними структурно-композиційними моделями, у цьому дослідженні зроблено спробу за допомогою одного і з статистичних методів, який, за даними попередніх досліджень, має стилерозрізнявальну потужність, а саме *частки чотирискладових слів та середнього квадратичного відхилення довжини слова у складах*, з'ясувати жанрову належність *приватних листів*.

Мовознавці, підкреслюючи складність системи мови і ймовірнісний характер мовлення, вважали за необхідне використання статистичних методів у лінгвістиці як для розкриття закономірностей будови і функціонування мови, так і для оцінювання достовірності теоретичних висновків.

Статистичний метод дає змогу скласти повну кількісну характеристику тексту загалом, що має важливе теоретичне значення саме по собі, а також важливе значення для пізнання внутрішньої структури мовного стилю.

Статистичні дослідження в Україні, в яких застосовують статистичні методи, проводять на *фонетичному, морфологічному, синтаксичному та лексичному* рівнях. У монографії «Статистичні параметри стилів» [7, с. 260] узагальнено досвід світової і вітчизняної лінгвістичної статистики, за допомогою чітко окресленого статистичного апарата встановлено закономірності функціонування елементів різних рівнів мовної структури: *фонологічного, морфологічного, лексичного і синтаксичного* у трьох групах стилів сучасної української літературної мови: *художньо-белетристичному (поезія, проза, драматургія), суспільно-політичному та науково-технічному*, використовуючи такі характеристики, як: *частоту фонем і груп фонем; частоту голосних, приголосних, пропусків між словами*, а також *груп приголосних: м'яких, сонорних, частоту різних типів складів, середню довжину складу; середню частоту афіксів; середню частоту флексій; частоту особових і неособових форм дієслова; префікси; словоформи та сполучники, середньої довжини речення, середню довжину слова*, і характер розподілу *довжини слова*. Автори дійшли висновку, що ці одиниці і особливості їх розподілу в текстах є статистичними параметрами стилів, які розмежовують не лише стилі й групи слів, але й твори авторів у межах одного стилю або жанру.

Отже, аналіз частот різних явищ на всіх рівнях мовної структури показав, що на кожному мовному рівні – *фонемному, морфемному, лексичному, синтаксичному* – встановлюються такі статистичні характеристики, які суттєво відрізняють один стиль (і функціональний, і авторський) від інших. Тому, кожний функціональний стиль можна описати через набір його статистичних параметрів, що дає в руки стиліста об'єктивні критерії для аналізу й опису кожного стилю. Дослідження також показало, що статистичні параметри стилів, які встановлюються на різних рівнях мовної структури, мають неоднакову стилерозрізнявальну потужність для різних груп стилів.

Подібні статистичні дослідження функціональних стилів проводили, зокрема, в університеті місто Грац (Австрія) де займалися дослідженням рівнів мови, а саме – синтаксичного і лексичного, та їх одиниць: речення і слова. Тут досліджували такі

критерії, як *середню довжину слова, довжину речення*, та займалися питанням виділення функціональних стилів і жанрів на основі мов слов'янських текстів (хорватської, сербської, словацької, словенської, російської). Дослідники (P. Grzybek, G. Antić, E. Kelih, E. Stadlober [12, с. 150–154; [17, с. 198–216; [14, с. 53–64; [15, с. 498–505]) дійшли висновку, що довжина слова та довжина речення є потужними показниками для розмежування текстових видів та сильними чинниками для текстової класифікації.

Оскільки вищеперелічені критерії слугують для розмежування функціональних стилів (жанрів в межах стилю), то у цьому дослідженні, яке є частиною австро-українського проекту «Створення збалансованого текстового банку даних української мови» [16, с. 125–132], було зроблено спробу розмежування приватного, відкритого листа, блогів у межах епістолярію за допомогою частки чотирискладових слів та середнього квадратичного відхилення довжини слова у складах у текстах приватних листів письменників.

Для дослідження було відібрано за прикладом хронологічної однорідності та гендерної паритетності приватні листи письменників XIX – початку XX ст. – О. Кобилянської [4], О. Теліги [11], В. Стуса [8, 9, 10], Б. Антоненка-Давидовича [1]. За вимогами проекту джерельна база роботи становила 40 одиниць в електронному вигляді і становила структурну, хронологічну вибірку. Мінімальна довжина підвибірki у листах не менше 250 слововживань, що є достатньою для одержання точних результатів. Величина повного обсягу проаналізованих текстів становить 31 641 слововживань.

Підрахунки частки чотирискладових слів та середнього квадратичного відхилення довжини слова у складах, вказаних листів, виконували за допомогою спеціально створених комп'ютерних програм у межах згаданого проекту за такими принципами:

1. Під час оброблення матеріалу не враховували: *ініціали, формули, графіки, малюнки, таблиці, діаграми, числа написані цифрами* (якщо такі були);

2. Для підрахунків брали лише *чотирискладові слова* у кожному тексті приватного листа (ці підрахунки робили автоматично, за допомогою спеціально написаної комп'ютерної програми). Під словом ми розуміли сукупність складів (на основі матеріалів досліджень австрійських мовознавців [12, с. 150–154]);

3. Для кожного тексту розраховували частку чотирискладових слів (кількість чотирискладових слів / загальну кількість слів) – P_4 і середнє квадратичне відхилення довжини слова у складах – m_2 .

4. Під час розрахунків ми користувалися такими формулами [13, с. 13–19]:

$$1) \bar{x} = \frac{1}{N} \sum x_i;$$

де \bar{x} – середня довжина слова;

x_i – довжина i -того слова;

N – кількість слів.

$$2) m_2 = \frac{1}{N} \sum x_i^2 - \bar{x}^2,$$

вихідною є формула – $m_2 = \frac{1}{N} \sum (x_i - \bar{x})^2$, де m_2 – середнє квадратичне відхилення довжини слів у складах;

$$3) d = \frac{m_2}{(m_1 - 1)},$$

де m_1 – середня довжина слова.

$$4) P_4 = N_4 / N,$$

де P_4 – частка чотирискладових слів; N_4 – число чотирискладових слів у тексті.

5. Частка чотирискладових слів (P_4) у кожній підвибірці була своя, і була вирахована за допомогою виведеної формули. Результати підрахунків представлені на рисунку 1 де порівняно приватні листи різних авторів з відкритими листами (результати досліджень у відкритих листах частки чотирискладових слів і середнього квадратичного відхилення довжини слова у складах взято із [13, с. 13–19]). Із графіка видно, що не всі приватні листи мають однакову частку вживання чотирискладових слів. Частка вживання чотирискладових слів найменша у В. Стуса і сягає позначки 0,10; в О. Кобилянській це позначка 0,11; в О. Телігі це позначка 0,12. Найбільша частка вживання чотирискладових слів прослідковується у листах Б. Антоненка-Давидовича, це позначка 0,09, ріст якої збільшується до 0,16, тому деякі листи письменника потрапляють в область відкритих листів. Такий показник свідчить про наближення приватних листів Б. Антоненка-Давидовича до офіційного стилю.

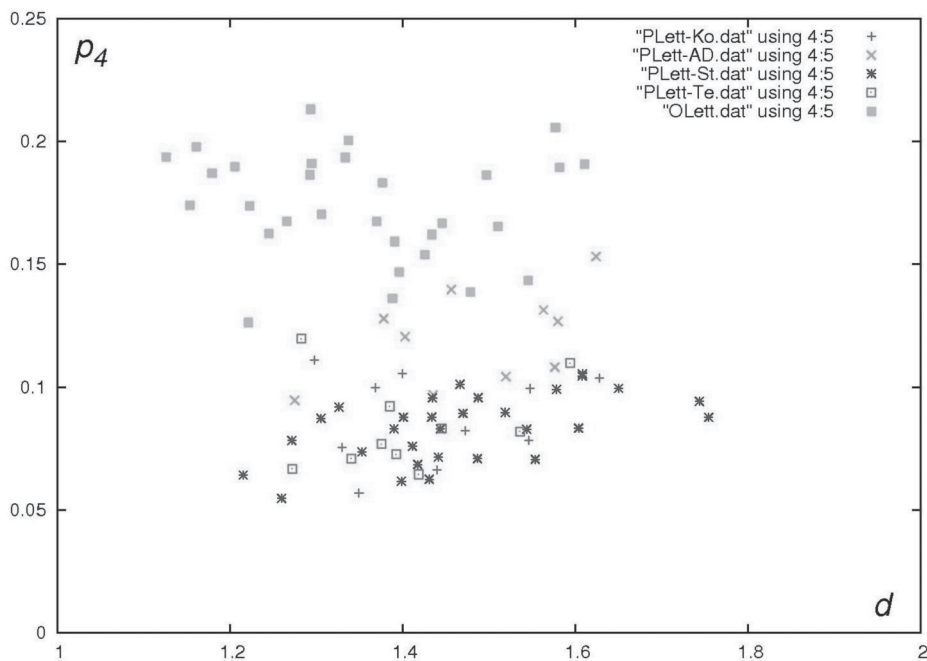


Рис. 1. (Антоненко-Давидович – AD, Кобилянська – Ko, Теліга – Te, Стус – St, відкриті листи – OLet)

Отримані результати дослідження частки чотирискладових слів у приватних листах можна порівняти із результатами досліджень частки чотирискладових слів у відкритих листах, блогів, спортивних репортажах, які були проведені у межах вищезгаданого проекту (див. рисунок 2). Адже відкритий лист, який належить до епістолярію за результатами досліджень проекту, відходить до публіцистичного стилю, одиницями якого є блоги, що ще раз підтверджує належність епістолярію до міжстильових жанрів.

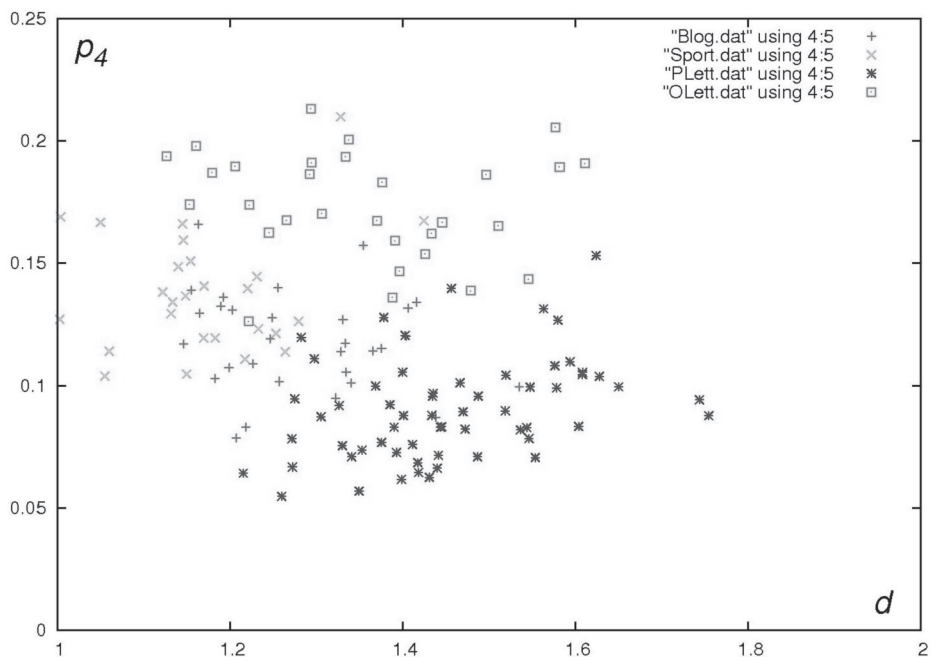


Рис. 2. (Blog – блоги; Sport – спортивні репортажі; Plett – приватні листи; Olett – відкриті листи)

На графіку видно відмежування приватних листів від блогів, спортивних репортажів та відкритих листів, хоча деякі приватні епістоли переходять межу блогів і відкритих епістол. Найнижча позначка вживання *чотирискладових слів* у приватних листах становить 0,06 і зростає до 0,16, знову ж таки переходячи в область блогів та відкритих листів, які сягають найвищої позначки 0,21. Найближче до приватних листів розташовані блоги, вони перетинаються на межі позначки 0,08; далі приватні листи перетинаються з спортивними репортажами на межі позначки 0,10; і найдалі розташовані відкриті листи, з якими перетинається незначна частка приватних на позначці 0,12, не переходячи позначки 0,16.

Показники *середнього квадратичного відхилення довжини слова у складах та частки чотирискладових слів* у блогах, спортивних репортажах, відкритих листах є значно вищими, ніж у приватних листах, хоча у деяких приватних епістолах цей показник значно зростає. Це свідчить про офіційний, діловий характер блогів, спортивних репортажів, відкритих листів та їхню належність до публіцистичного та офіційного стилів. У приватних листах наявне вживання дещо складніших синтаксичних конструкцій, що притаманно приватним напівофіційним епістолам (це листи Б. Антоненка-Давидовича) на відміну від решти приватних листів, які мають розмовно-побутовий, художній характер і значно простішу синтаксичну будову.

Отже, *частка чотирискладових слів та середнє квадратичне відхилення довжини слова у складах* можуть слугувати статистичним параметром для розмежування чотирьох типів текстів – приватного, відкритого листів, блогів та спортивних репортажів. Так само, як за даними підрахунків дослідників австрійського університету

довжина слова та довжина речення виявляються основними критеріями для розмежування функціональних стилів, текстових типів та авторських стилів у слов'янських мовах.

Оскільки епістолярій – це міжстильове явище, якому притаманні риси різних стилів, то, відповідно, можна розрізнати приватні листи, які мають ознаки розмовного, художнього та елементи офіційного стилю. Про це свідчать також результати досліджень австрійських мовознавців, які вид приватного листа зараховують до розмовного та художнього жанрів. Із результатів нашого дослідження випливає, що серед приватних листів, які належать до епістолярного жанру, можна виділити приватні листи з рисами розмовно-побутового, художнього та частково офіційного стилю. Це підтверджує праця В. Перебийніс «Статистичні параметри стилів» [7, с. 260]. Беручи до уваги, що частка чотирискладових слів є досить малою, то логічно, що у тексті мають бути слова, у складі яких не менше 8–9 фонем. Якщо взяти результати досліджень довжини слова, то саме у белетристичному стилі, до якого частково можна віднести приватний лист, частота 8–9 фонемних слів становить 14–16%, в офіційно-діловому частота 8-фонемних слів становить – 8,2%, а 9-фонемних – 7,4%.

Досліджена у роботі частка чотирискладових слів та середнє квадратичне відхилення довжини слова у складах є статистичними параметрами виокремлення та розмежування текстів навіть у межах епістолярію, публіцистичного та офіційного стилів. Мала частка чотирискладових слів (від 0,06), а подекуди її зростання (до 0,16) у приватних листах свідчить про їх розмовний, художній та частково офіційний характер, про вживання у текстах простих речень (довжина яких не перевищує 8–9 фонемовживань), синтаксичних конструкцій, подекуди складних, коротких слів, про їх емоційність, експресивність та інтимно-побутовий характер. На відміну від блогів, спортивних репортажів та відкритих листів, яким властива висока частка чотирискладових слів (від 0,08 до 0,21), офіційний, діловий, публіцистичний характер, про вживання у цих текстах складних синтаксичних конструкцій, слів, що відрізняє їх певною мірою від приватного листа.

Дослідження статистичних параметрів природної мови набуває великого самостійного теоретичного і практичного значення, оскільки висвітлює найбільш глибокі її закономірності. Їх пізнання дає змогу оптимізувати програми комп'ютерного опрацювання природної мови.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Антоненко-Давидович Б. Знаний і незнаний. Нащадки прадідів / Б. Антоненко-Давидович. – «КМ Academija». – К., 1998.
2. Бахтин М. М. Проблема речевых жанров / М. М. Бахтин // Литературно-критические статьи. – М. : Худ. лит.-ра, 1986. – С. 428–472.
3. Кеца А. И. Место епистолярного стиля в системе функциональных стилей / А. И. Кеца // Известия АН АЗССР. Серия литературы. – 1971. – № 3–4. – С. 94–99.
4. Кобилянська О. Листи / О. Кобилянська // Твори : у 5 т. – К. : Держлітвидав, 1963.
5. Наер В. Л. К проблеме жанра в системе функционально-стилистической дифференциации языка / В. Л. Наер // Стилистические аспекты устной и письменной коммуникации. – М. : ИГП ИИЯ, 1987. – С. 39–47.

6. Павлик Н. В. Специфіка епістолярного жанру як міжстильового явища / Н. В. Павлик // Лінгвістика: зб. наук. пр.: Луганськ : Видавництво ЛНПУ ім. Т. Шевченка «Альма-матер», 2005. – С. 241–248.
7. Статистичні параметри стилів / за ред. В. С. Перебийніс. – К. : Наук. думка, 1967. – 260 с.
8. Стус В. Твори : у 4-х т. 6-ти кн. / В. Стус. – Т. 6 (додатковий), кн. 1. Листи до рідних. – Львів : Просвіта, 1997.
9. Стус В. Твори : у 4-х т. 6-ти кн. / В. Стус. – Т. 6 (додатковий), кн. 2. Листи до друзів та знайомих. – Львів : Просвіта, 1997.
10. Стус В. Твори : у 4-х т. 6-ти кн. – Львів : Просвіта 1997– Т. 6 (додатковий), кн. 4. Повісті та оповідання; незакінчені твори; сценарії; літературна критика; заяви, публіцистичні листи та звернення; з таборового зошита.
11. Теліга О. Листи. Спогади / О. Теліга / упор. Н. Миронець. – 2-ге вид, випр. – К. : Видавництво ім. О. Теліги, 2004. – 400 с.
12. Antić G. Word Length and Frequency Distributions / G. Antić, E. Stadlober, P. Grzybek, E. Kelih. C. Weihs and W. GAUL (Eds.), Classification – The Ubiquitous Challenge. Springer, Heidelberg, 2005. – P. 150–154.
13. Buk S. Word-length-related parameters of text genres in the Ukrainian language. A pilot study / S. Buk, Humenychuk, L. Mal'tseva, A. Rovenchak // Text and Language : Structures – Functions – Interrelations. Quantitative perspectives / Ed. by P. Grzybek, E. Kelih, J. Mačutek. – Wien : Praesens, 2010. – P. 13–19.
14. Grzybek P. Quantitative Text Typology : The Impact of Word Length / P. Grzybek, E. Stadlober, E. Kelih, and G. Antić. C. Weihs and W. GAUL (Eds.), Classification – The Ubiquitous Challenge. Springer, Heidelberg, 2005. – P. 53–64.
15. Kelih E. Classification of Author and / or Genre? The Impact of Word Length / E. Kelih, G. Antić, P. Grzybek, and E. Stadlober. C. Weihs and W. GAUL (Eds.), Classification – The Ubiquitous Challenge. Springer, Heidelberg, 2005. – P. 498–505.
16. Kelih E. Project Description : Designing and Constructing a Typologically Balanced Ukrainian TextDatabase / E. Kelih, S. Buk, P. Grzybek, A. Rovenchak // Methods of Text Analysis : Omnibus volume. – Chernivtsi : ČNU, 2009. – P. 125–132.
17. Kelih E. Quantitative Text Typology. The Impact of Sentence Length / E. Kelih, P. Grzybek, G. Antić, E. Stadlober. C. Weihs and W. GAUL (Eds.), Classification – The Ubiquitous Challenge. Springer, Heidelberg, 2005. – P. 198–216.

Стаття надійшла до редколегії 13.10.2010

Принята до друку 22.10.2010

**СТАТИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ
ЧАСТНОЙ КОРЕСПОНДЕНЦИИ:
ПАРАМЕТРЫ, СВЯЗАННЫЕ С ДЛИНОЙ СЛОВА**

Ольга ГУМЕНЧИК

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

Исследовано вопрос стилевой принадлежности частной корреспонденции с помощью таких статистических критериев как доля четырехсложных слов и среднего квадратического отклонения длины слова в слогах.

Ключевые слова: эпистолярный жанр, частное письмо, статистические методы, доля четырехсложных слов, среднее квадратическое отклонение длины слова в слогах.

**STATISTICAL CHARACTERISTICS
OF PRIVATE CORRESPONDENCE:
CRITERIA RELATED WORD LENGTH**

Olga HUMENCHYK

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

In the article, the style attribution of private correspondence is analyzed using statistical criteria such as four-syllabic words fraction and mean square deviation of the word length.

Key words: epistolary genre, private letters, statistical methods, fraction of four-syllabic words, mean square deviation of the word length.