

исключаем образования, полученные путем сокращения целого слова до одного или двух слогов, поскольку таковые не отвечают постулированным параметрам аббревиатур.

## ЛИТЕРАТУРА

1. Ганич Д. І., Олійник І. С. Словник лінгвістичних термінів / Д. І. Ганич, І. С. Олійник. – К., 1985.
2. Мацько О. М. Аббревіатури як згорнені мовні формули в дипломатичних текстах // Мовознавство. – 2000. - № 1. – С. 31-36.
3. Лингвистический энциклопедический словарь / Гл. ред. В. Н. Ярцева. – М.: Советская энциклопедия, 1990. – 688 с.
4. Булаховський Л. А. Основи мовознавства // Соціальна природа мови. 2-а лекція: Заоч. пед. вуз. – Харків, 1928-1929.
5. Селіванова О. О. Лінгвістична енциклопедія / О. О. Селіванова. – Полтава: Довкілля-К, 2010. – 844 с.
6. Теркулов В. И. Еще раз об основной единице языка // Вісник Луганського національного університету ім. Т.Г. Шевченка. – Луганськ, 2006. – №11 (106). – С. 127-136.

УДК 81-139

*Файфер Н.В.*  
(Київ, Україна)

## ВИКОРИСТАННЯ СТАТИСТИЧНИХ МЕТОДІВ У ЛІНГВІСТИЦІ

*Стаття присвячена різноманітним шляхам використання статистичних методів у лінгвістичних дослідженнях. Застосування статистики забезпечує високу вірогідність і надійність результатів аналізу, а також дозволяє глибше виявити сутність досліджуваних явищ. Використання статистичних методів у лінгвістичних дослідженнях зумовлене властивими мові кількісними ознаками, внутрішньою взаємозалежністю, яка існує між якісними та кількісними особливостями мовної структури, підпорядкованістю частоти мовних одиниць у мовленні певним статистичним закономірностям, можливістю отримати об'єктивні дані.*

**Ключові слова:** статистичний метод, статистика, вірогідність, частота, статистична закономірність.

*Статья посвящена различным способам использования статистических методов в лингвистических исследованиях. Применение статистики обеспечивает высокую достоверность и надёжность результатов анализа, а также позволяет выявить сущность исследуемых явлений. Использование статистических методов в лингвистических исследованиях обусловлено свойственными языку количественными признаками, внутренней взаимозависимостью, которая существует между качественными*

© Файфер Н.В., 2011

и количественными особенностями языковой структуры, подчинённостью частоты языковых единиц в речи определённым статистическим закономерностям, возможностью получить объективные данные.

**Ключевые слова:** статистический метод, статистика, достоверность, частота, статистическая закономерность.

*The article focuses on various ways of using statistical methods in linguistic investigations. The usage of statistics does not only ensure a high level of credibility and reliability of the results of the analysis, but also allows to reveal the essentiality of the investigated phenomena. The using of statistical methods in linguistic research investigations is predetermined by the quantitative characteristics peculiar to the language, by the inner interdependence which exists between the qualitative and quantitative peculiarities of the language structure, by the subordination of the frequency of language units to certain statistical regularities, by the possibility of getting objective data.*

**Key words:** statistical method, statistics, credibility, frequency, statistical regularity.

У цій статті зупинимося на важливості застосування статистичних методів у лінгвістиці, а також розглянемо можливі способи їх використання. Слід зазначити, що сфера дії статистичних законів надзвичайно широка: всі складні системи підкоряються перш за все законам статистичним. Однією з таких систем є звичайно мова, яка дійсно являє собою надзвичайно складну систему дискретних одиниць, які, як і всі дискретні одиниці будь-якої природи, можуть мати кількісні характеристики. Ці кількісні характеристики властиві одиницям усіх рівнів мовної системи, причому кількісні характеристики одиниць нижчого рівня можуть призводити до якісних відмінностей на вищому рівні [1 : 3]. Перед собою ставимо такі завдання: розглянути існуючі статистичні методи та прийоми; визначити значення статистики для досягнення об'єктивності та вірогідності результатів лінгвістичних досліджень.

Застосування статистичних методів дослідження [2 : 109] забезпечує надійність результатів [3 : 4; 4 : 21], оскільки статистика: 1) спирається на точні кількісні дані, які можна перевірити; 2) визначає, в яких випадках різниця між зіставлюваними частотами є істотною, а в яких – випадковою, зумовленою закономірними коливаннями частоти в різних вибірках; 3) визначає кількісні параметри матеріалу достатні для вірогідних висновків [5 : 27–28].

Актуальність дослідження зумовлена спрямуванням сучасних лінгвістичних студій на вивчення системних і функціональних характеристик лексичних і граматичних одиниць, у тому числі й на базі статистичних методів.

Без сумніву, мова і мовлення мають не тільки якісні, але й кількісні ознаки. Саме тому математична лінгвістика, квантитативна лінгвістика, статистика мовлення і лінгвостатистика займаються кількісним аналізом мови й мовлення [6 : 11–12]. Використання статистичних методів у лінгвістиці присвятили свої праці багато дослідників [7 : 93; 8 : 1; 9 : 122 та ін.]. Використання математичних методів під час наукових досліджень дозволяє глибше виявити сутність досліджуваних явищ [10 : 5; 11 : 98].

Розгляд взаємодії мовних одиниць з погляду їх функціонування у мовленні є завжди актуальним. В останні роки дослідження мовних одиниць, їх властивостей та відношень між ними здійснюють за допомогою статистичних методів і прийомів [12; 13; 14; 15; 16;

17; та ін.], які дозволяють отримати надійні, вірогідні результати аналізу. Використання статистичних методів у лінгвістичних дослідженнях зумовлене властивими мові кількісними ознаками, внутрішньою взаємозалежністю, яка існує між якісними та кількісними особливостями мовної структури, підпорядкованістю частоти мовних одиниць у мовленні певним статистичним законамірностям, можливість отримати об'єктивні дані, незалежні від суб'єктивного сприйняття дослідника [6 : 5].

Мета, завдання та матеріал, що підлягає аналізу, зумовлюють використання певних **методів дослідження**. Так, встановлення лексичної близькості порівнюваних текстів або газет (лексична близькість вимірюється кількістю лексичних одиниць, спільних для кожної пари текстів/газет), визначення наявності/відсутності взаємозв'язку між функціональними особливостями досліджуваних одиниць та ін. ґрунтується на використанні *статистичних методів*, зокрема *відносної похибки, критерію Стьюдента, критерію Стьюдента для відсоткових показників, рангової кореляції* тощо.

За допомогою статистики можна визначити кількісне співвідношення досліджуваної одиниці в аналізованих текстах, а також лексичну спільність порівнюваних парно текстів/газет щодо використання зафіксованих одиниць. У межах зазначеної проблематики важливе місце посідає вивчення особливостей і закономірностей функціонування мовних одиниць у текстах різних стильових різновидів [15 : 17; 14 : 246]. Функціональні властивості мовної одиниці визначаються особливостями її вживання в певному тексті: її частотою, позицією, сполучуваністю, яка залежить від характеру текста, функціонального стилю і змінюється від тексту до тексту [5 : 30]. Саме шляхом виявлення розходжень між частотами мовної одиниці в різних стилях та жанрах встановлюються як закономірності функціонування одиниці в цих стилях, так і особливості зіставлених стилів. Частоти одиниць, істотно відмінні в різних стилях, називаються статистичними параметрами стилів.

Використання статистики можна назвати сучасним підходом до вивчення різних лінгвістичних явищ [18 : 5]. Одне з провідних завдань вбачають у статистичному описі різних мовних рівнів – лексичного, зокрема фразеології, граматичного, а також фонемно-графемного рівня [19 : 3]. Так, Г. Карлгрен наголошує на важливості використання статистики під час дослідження фонематичної і синтаксичної структур [20 : 111], В. І. Перейбініс – на необхідності кількісної характеристики стилів [4 : 21]. В. В. Левицький зауважує, що необхідність використання інформаційно-пошукових систем, частотних словників і т. ін. стимулює розвиток такого напрямку як «лексикостатистика» [21 : 3].

Частота одиниць мови є показником їх системних і функціональних властивостей. Системні властивості мовної одиниці (наприклад, синтаксична й лексична сполучуваність: моделі синтаксичних конструкцій, притаманних одиниці, групи слів, з якими вона сполучається, які є типовими для неї; семантика: кількість значень, лексична група, в яку ця одиниця входить тощо) можуть впливати на її частоту, отже, частота відображає ту чи іншу властивість мовної одиниці чи певну сукупність її властивостей [5 : 28–29]. Виявити взаємозв'язок між системною властивістю одиниці мови та її вживаністю можна лише за допомогою статистики.

Так, наприклад, необхідно дослідити певні одиниці в текстах кількох газет. Перш за все, визначаємо кількість, і частоту (абсолютну та відносну) аналізованих одиниць в одержаному нами масиві.

Для того, щоб з'ясувати, чи істотне розходження відносної частоти аналізованих одиниць у текстах порівнюваних попарно газет, можна використати критерій Стьюдента для процентних показників за формулою:

$$t = \frac{|p_1 - p_2|}{Sed\%}, \quad (1)$$

де  $t$  – коефіцієнт критерію Стьюдента,  $p_1$  – процентний показник (відносна частота у відсотках) для першого масиву одиниць,  $p_2$  – для другого масиву досліджуваних одиниць.  $Sed\%$ , у свою чергу, обчислюється за формулою:

$$Sed\% = \sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}, \quad (2)$$

де  $n_1$  – обсяг першої вибірки, на основі якого обчислювалися проценти, а  $n_2$  – обсяг другої вибірки. Символ  $P$  означає середній для двох порівнюваних виборок процентний показник, який визначаємо за формулою:

$$P = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2} \quad (3)$$

$$Q = 100\% - P \quad (4)$$

Необхідно зазначити, що частота аналізованих одиниць характеризує закономірність вживання досліджуваних одиниць у вибірці, яка, в свою чергу, репрезентує генеральну сукупність. Сталість вживання є важливою статистичною характеристикою, яку можна використати для зіставлення однакових одиниць у різних вибірках.

Щоб з'ясувати, чи є істотна різниця між частотою певних одиниць у текстах порівнюваних газет, використовується формула відносної різниці:

$$\mathcal{E} = \frac{|x_1 - x_2|}{x_1}, \quad (5)$$

де  $\mathcal{E}$  – відносна різниця між порівнюваними частотами,  $x_1$  і  $x_2$  – зіставлявані відносні частоти в текстах аналізованих газет,  $|x_1 - x_2|$  – читається *по модулю* й означає, що  $x_1 > x_2$ .

Якщо помножити цей дріб на 100, одержимо відносну різницю у відсотках. Розходження між порівнюваними частотами вважається істотним, якщо  $\mathcal{E} > 20\%$ .

Щоб перевірити порівнювані газети на ступінь лексичної близькості між ними, використаємо формулу:

$$R = \frac{c}{(V_a + V_b) - c}, \quad (6)$$

де  $R$  – ступінь лексичної близькості між порівнюваними попарно газетами,  $c$  – кількість однакових досліджуваних одиниць, які зафіксовано в текстах обох газет,  $V_a$  – кількість аналізованих одиниць у текстах однієї з порівнюваних газет, а  $V_b$  – кількість ана-

лізованих одиниць, зафіксованих у текстах іншої газети. Наявність лексичної близькості фіксуємо, якщо  $R > 20\%$ .

Аналіз лексичних одиниць та особливостей їх використання у будь-якому стилі не можна вважати повним без дослідження моделей їх граматичної сполучуваності. Так, методом рангової кореляції можна перевірити, чи однаково розподіляються частоти моделей граматичної сполучуваності в певних групах досліджуваних одиниць за формулою:

$$p = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}, \quad (7)$$

де  $n$  – це кількість зафіксованих моделей граматичної сполучуваності досліджуваних одиниць. За допомогою зазначеного вище методу можна також прослідкувати, чи є взаємозалежність між поширеністю моделі та її частотою.

Функціонування зафіксованих моделей граматичної сполучуваності аналізованих одиниць можна розглядати з погляду їх вживаності та поширеності (поширеність моделі представлена кількістю досліджуваних одиниць, які реалізують аналізовану модель у газетах/текстах; вживаність моделі – це частота лексичних одиниць, яким притаманна ця модель).

Таким чином, користь статистики безперечна й багатогранна, оскільки саме статистичні методи забезпечують високу вірогідність результатів дослідження; уточнюють висновки та умовиводи; виявляють хибні уявлення і спростовують їх; розкривають взаємозв'язки системних і функціональних властивостей мовних одиниць, а також їх властивості, які не можуть бути виявлені іншими методами.

Дослідження особливостей функціонування мовних одиниць у різних текстах, здійснене за допомогою статистичних методів, має подальші перспективи. Перспективними вважаються розвідки, присвячені порівнянню різних функціональних стилів (наукового, художнього тощо) у використанні різних мовних одиниць.

## ЛІТЕРАТУРА

1. Перебийніс В. І. Статистичні методи для лінгвістів : [навч. посіб.] / Валентина Ісидорівна Перебийніс. – Вінниця : Нова Книга, 2002. – 171, [1] с.
2. Никифорова І. В. Роль статистичного аналізу при дослідженні просодії звукової мови / І. В. Никифорова // Вісник Харківського національного університету ім. В. Н. Каразіна. – Харків : Константа, 2006. – № 741. – С. 105–109.
3. Перебийнос В. І. Роль моделей в розграниченні значень многозначного глагола (к вопросу о методах лексикологических исследований – на материале глаголов МАКЕ и ДО в английском языке) : автореф. дис. на соискание учен. степени канд. филол. наук : спец. 10.02.04 «Германские языки» / В. И. Перебийнос. – Москва, 1962. – 15 с.
4. Перебийніс В. С. Статистичні параметри стилів / Валентина Сидорівна Перебийніс. – К. : Наукова думка, 1967. – 260 с.
5. Перебийніс В. І. Що дає статистика мовознавцям? / В. І. Перебийніс // Вісник Київського лінгвістичного університету : зб. наук. праць. Серія : Філологія. – К. : Вид. центр

КНЛУ, 2003. – Т. 6, № 2. – С. 27–32.

6. Левицкий В. В. Квантитативные методы в лингвистике / Виктор Васильевич Левицкий. – Винница : Нова Книга, 2007. – 264 с.

7. Průcha J. Contextual Constraints and the Choice of Semantic Lexical Units / J. Průcha // Prague Studies in Mathematical Linguistics. – Prague, 1966. – № 1. – P. 93–104.

8. Woods A. Statistics in Language Studies / An. Woods, P. Fletcher, Ar. Hughes. – Cambridge : Cambridge University Press, 1996. – 322 p.

9. Biber D. Corpus Linguistics. Investigating Language Structure and Use / D. Biber, S. Conrad, R. Reppen // Applied Linguistics. – 2001. – Vol. 22, № 22. – P. 122–123.

10. Носенко И. А. Начала статистики для лингвистов : [учеб. пособ.] / Иван Андреевич Носенко. – М. : Высш. шк., 1981. – 157, [1] с.

11. Dillon G. L. Corpus, Creativity, Cliché : Where Statistics Meet Aesthetics / George L. Dillon // Journal of Literary Semantics : An International Review. – 2006. – Vol. 35, № 2. – P. 98.

12. Шевченко Т. Е. Функциональные характеристики словоизменительной парадигмы английского глагола (на материале современной художественной прозы): дис. ... канд. филол. наук : 10.02.04 / Шевченко Татьяна Егоровна. – К., 1997. – 201 с.

13. Варфоломеев А. П. Психосемантика слова и лингвостатистика текста : метод. рекомендации к спецкурсу / Варфоломеев А. П. – Калининград : Калинингр. ун-т, 2000. – 37 с.

14. Бойко Ю. Частота вживання підрядних речень як параметр авторського та функціонального стилів / Ю. Бойко // Науковий вісник Чернівецького університету : зб. наук. праць. – Чернівці : Рута, 2003. – С. 245–259.

15. Бенвенист Э. Общая лингвистика / Эмиль Бенвенист ; [пер. с фр. Ю. Н. Караулова, В. П. Мурат, И. В. Барышева, И. Н. Мельникова]. – М. : Едиториал УРСС, 2002. – 448 с.

16. Нагорна С. С. Стилеметричні характеристики дієслова у сучасних англословних наукових текстах: дис. ... канд. філол. наук : 10.02.04 / Нагорна Світлана Сергіївна. – К., 2004. – 262 с.

17. Жуковська В. В. Лексико-семантичні та граматичні характеристики англійського дієслова у художньому тексті: лінгвостатистичний аспект (на матеріалі популярних політичних романів) : дис. ... канд. філол. наук : 10.02.04 / Жуковська Вікторія Вікторівна. – К., 2006. – 307 с.

18. Abraham S. On 'Modernism' in Linguistics / S. Abraham // Linguistics : An International Review. – 1967. – № 37. – P. 5–11.

19. Статистика речи / ред. кол. П. М. Алексеев, В. М. Калинин, Р. Г. Пиотровский. – Л. : Наука, 1968. – 259, [1] с.

20. Karlgren H. Statistical Context Analysis / Hans Karlgren // Statistique et analyse linguistique. – Paris, 1966. – P. 111–112.

21. Левицкий В. В. Статистическое изучение лексической семантики : учеб. пособ. / Виктор Васильевич Левицкий. – Черновцы : ЧГУ, 1989. – 156 с.