

5. FOCUS-ONLINE [Electronic resource]. – Access mode: <http://www.focus.de/nachrichtenarchiv/artikel-2012.html>
6. SPIEGEL-ONLINE [Electronic resource]. – Access mode: <http://www.spiegel.de/nachrichtenarchiv/artikel-2011.html>
7. SPIEGEL-ONLINE [Electronic resource]. – Access mode: <http://www.spiegel.de/nachrichtenarchiv/artikel-2012.html>
8. SPIEGEL-ONLINE [Electronic resource]. – Access mode: <http://www.spiegel.de/nachrichtenarchiv/artikel-2013.html>

Горченко О.А.

аспірант,

Чернівецький національний університет

імені Юрія Федьковича

ЛІНГВОСТАТИСТИКА ЯК МЕТОД АНАЛІЗУ ТЕКСТУ

Для пізнання мови є необхідними квантативні методи, які базуються на імовірнісних, гра дуальних та частотних характеристиках [3, с. 15]. В.І. Перебийніс зауважує, «...мова і мовлення підкоряються статистичним законам» [6, с. 3]. Застосування математичних прийомів для систематизації і узагальненні лінгвістичного матеріалу, в сукупності з досконалою інтерпретацією результатів дослідження відкриває шлях до потаємної будови мови багатьом лінгвістам [6, с. 150]. За словами В.В. Левицького: «...Сьогодні навряд існують сумніви в тому, що мова володіє і якісними, і кількісними ознаками, і що різні властивості мови можуть вивчатися різними методами. Є такі дослідження, де відповіді на поставлені питання можна отримати тільки за допомогою квантативних методів» [4, с. 3, 9].

За допомогою квантативних методів у цей час досліджуються всі мовні підсистеми, тобто всі рівні мови й мовлення. Є спроби яким-небудь чином систематизувати лінгвостатистичні дослідження, тобто виділити ті галузі мовознавства, де квантативні методи успішно застосовуються або можуть бути ефективно використані [1].

Дослідження фонетичного, лексичного складу мови, ідіостилів, семантика мови, темпи розвитку і зміни мови, порівняльний аналіз мов успішно проведено саме завдяки методам лінгвостатистики [4].

За допомогою критерію χ^2 оцінюють незалежність у розподілі об'єктів сукупності за градаціями досліджуваної ознаки, відповідність між фактичним і теоретично очікуваним розподілами, а також однорідність розподілів [2, с. 31]. Цей критерій розраховується за такою формулою:

$$\chi^2 = \sum \frac{(O-E)^2}{E}, \quad (1)$$

де Σ – сума одержаних величин, О – емпіричні, а Е – теоретично очікувані частоти.

Варто зауважити, що сума χ^2 є тим більшою, чим більше емпіричні величини відхиляються від теоретичних.

Наведемо приклад розрахунку χ^2 -тесту для іменників, дієслів, прикметників і прислівників британських письменників XIX століття.

Таблиця 1

Величини χ^2 для вживання частотності іменників, дієслів, прикметників і прислівників у англійському дискурсі (на прикладі XIX століття)

Автор	Дієслово	Іменник	Прикметник	Прислівник
Charles Dickens «A tale of two cities»	23,16	23,13	13	24,8
Charlotte Bronte «Jane Eyre»	<u>1,92</u>	<u>0,67</u>	17,29	46,16
Theodore Dreiser «Stoic»	4,35	<u>1,86</u>	6,59	11,79
Jane Austen «Emma»	<u>3,11</u>	65,85	<u>1,85</u>	399,07

Отже, можна зазначити, що підкреслені величини не є значущими для підведення підсумків на визначення частотності у межах вибірки.

Оскільки сума хі-квадрат дозволяє встановити лише ознаки, а не міру, то в дослідженнях доцільно використовувати коефіцієнт взаємної спряженості за О. Чупровим [5].

$$K = \sqrt{\frac{\chi^2}{N\sqrt{(r-1)(c-1)}}}, \quad (2)$$

де **K** – коефіцієнт взаємної спряженості,

N – загальна кількість спостережень,

r – кількість горизонтальних рядків таблиці,

c – кількість вертикальних рядків.

Величина коефіцієнта коливається від 0, при відсутності зв'язку між величинами до +1, при чіткому зв'язку між ними від $0 \leq K \leq 1$. За розрахунком коефіцієнта **K** ми отримали такі результати:

Таблиця 2

Коефіцієнт К для вживання частотності іменників, дієслів, прикметників і прислівників у британському дискурсі (на прикладі XIX століття)

Автор	Дієслово	Іменник	Прикметник	Прислівник
Charles Dickens «A tale of two cities»	0,007	0,007	0,005	0,007
Charlotte Bronte «Jane Eyre»	0,002	0,001	0,006	0,009
Theodore Dreiser «Stoic»	0,003	0,002	0,003	0,005
Jane Austen «Emma»	0,002	0,011	0,002	0,028

Можна стверджувати, що між величинами існує чіткий зв'язок, який вказує на доцільність даного дослідження.

З вищезгаданого, статистичні методи є ваговою складовою будь-якого дослідження, оскільки вказують на частотність відібраного матеріалу, що уможлиблює отримати точні, аргументовані і достовірні результати.

Список використаних джерел:

1. Варбанець П.Д., Якімова Н.А. ЛІНГВОСТАТИСТИКА. Методичний посібник для практичних занять / П.Д. Варбанець, Н.А. Якімова. – Одеса, 2014. – 130 с.
2. Король А.А. Висловлення-звинувачення у сучасному німецькомовному художньому дискурсі: дис. канд. філол. наук: 10.02.04 / А. А. Король. – Чернів. нац. ун-т ім. Ю. Федьковича. – Чернівці, 2007. – 237 с.
3. Кочерган М.П. Загальне мовознавство: підручник / М.П. Кочерган. – 2-ге вид., виправл. і доповн. – К.: Видавничий центр «Академія», 2006. – 464 с.
4. Левицкий В.В. Квантитативные методы в лингвистике / В.В. Левицкий – Черновцы: Рута, 2004. – 190 с.
5. Носенко И.А. Начала статистики для лингвистов / И.А. Носенко. – М.: Высшая школа, 1981. – 157 с.
6. Перебийніс В.І. Статистичні методи для лінгвістів: [навч. посіб.] / Валентина Ісидорівна Перебийніс. – Вінниця: Нова Книга, 2002. – 171 с.