

номічних відносин переважно сфери торгівлі та виробництва. Активне вживання виробничої лексики в староукраїнський період було зумовлено інтенсивним розвитком і диференціацією різних видів робіт, торговельних відносин, виникненням нових професій, особливо міських, укладанням майнових стосунків.

Інтенсивне формування української економічної термінології припадає на другу половину XIX ст. і пов'язане як з інтенсивним розвитком української економічної науки, так і з певними суспільно-політичними процесами того часу. «Розвиток промисловості, посилення класової диференціації суспільства і класової боротьби, піднесення революційного руху, поширення марксистсько-ленінських ідей на Україні з кінця XIX ст., нові наукові відкриття і загальне зростання культури народу зумовили значне поповнення лексичного складу української літературної мови» [1, с. 157]. «В історії формування української термінологічної лексики в галузі економічних наук останні передреволюційні десятиліття мали велике значення. Основи цієї термінології заклалися саме в ті роки...» [1, с. 206]. У цей час з'являються наукові й науково-публіцистичні праці економічної тематики, написані новою українською літературною мовою. Економічні розвідки, матеріали про народне господарство України від 70-х років до першої Світової війни на східно-українських землях подавала земська статистика; україномовні економічні праці, незважаючи на всілякі заборони, з'являлися на сторінках журналу «Киевская старина», виходили друком дослідження М. Зібер, С. Подолинського, П. Чубинського, М. Яснопольського та ін. Праці українських дослідників В. Барвінського, М. Грушевського, С. Дністрянського, І. Заневича, Б. Навроцького, О. Терлецького, І. Франка та багатьох інших друкувалися в періодичній пресі, яка на той час, як слушно зауважував М. А. Жовтобрюх [1, с. 205], була джерелом формування економічної термінології української мови.

Отже, формування основи термінології торгівлі української мови тісно пов'язане з історією розвитку загальноекономічної термінології, у складі якої виникла значна кількість лексем, використаних для номінації понять фахової мови торгівлі. У досліджуваній підмові широко вживаються традиційні, питомо українські терміни, тобто простежується тенденція до самозбереження національної мови в умовах термінологічної експансії, наприклад, *гуртовий продаж, продаж на гурт (оптовий продаж); крамниця (магазин), крам (товар для збуту); постачання (поставка)*.

Література:

1. Жовтобрюх М. А. Мова української періодичної преси (кін. XIX – поч. XXст.) / М. А. Жовтобрюх. – К. : Наук. думка, 1970. – 302 с.
2. Панько Т. І. Від терміна до системи : Становлення марксистсько-ленінської політ економічної термінології у східно-слов'янських мовах / Т. І. Панько. – Л. : Вища шк., 1979. – 148 с.
3. Панько Т. І. Українське термінознавство / Т. І. Панько, І. М. Кочан, Г. П. Мацюк. – Львів : Світ, 1994. – 216 с.
4. Тер-Минасова С. Г. Язык и межкультурная коммуникация / С. Г. Тер-Минасова. – М., 2000. – 624 с.
5. УТіС 1997: Українська термінологія і сучасність: Матеріали II Всеукр. наук. конф. – К., 1997. – 239 с.
6. Malinowski B. The problem of meaning in primitive languages // The Meaning of Meaning / B. Malinowski. – London, 1923. – P. 451–511.

УДК 811:311.21

С. К. Романюк,

Рівненський державний гуманітарний університет, м. Рівне

ЗАСТОСУВАННЯ СТАТИСТИЧНИХ МЕТОДІВ У ЛІНГВІСТИЧНИХ ДОСЛІДЖЕННЯХ

У статті розглядаються загальні положення статистичної лінгвістики. Описуються лінгвостатистичні методи: критерій Стьюдента і критерій хі-квадрат. Наведено конкретні приклади застосування цих методів.

Ключові слова: статистична лінгвістика, статистичні методи, критерій Стьюдента, критерій хі-квадрат.

В статье рассматриваются общие положения статистической лингвистики. Описываются лингвостатистические методы: критерий Стьюдента и критерий хи-квадрат. Приводятся конкретные примеры применения этих методов.

Ключевые слова: статистическая лингвистика, статистические методы, критерий Стьюдента, критерий хи-квадрат.

The article deals with the general problems of statistical linguistics. The statistical methods, such as Student's test and chi-square test, involve the use of different formulas to identify the rules of distribution of linguistic units in speech, to measure the links between language elements and to establish the relationship between qualitative and quantitative characteristics of speech. The examples of these methods application are resulted.

Keywords: statistical linguistics, statistical methods, Student's test, chi-square test.

Постановка проблеми. Мова є надзвичайно складною системою дискретних одиниць, які можуть мати кількісні характеристики. Ці кількісні характеристики властиві одиницям усіх рівнів мовної системи. Будова мови, її функціонування в мовленні, співвідношення мови й мислення, мови й суспільства підкоряються дії статистичних законів [5, с. 3, 6–7].

Статистичні методи у мовознавстві допомагають правильно організувати лінгвістичні спостереження, отримати об'єктивні дані, незалежні від суб'єктивного сприйняття дослідника, забезпечити надійність, точність, достовірність висновків.

З огляду на це залучення статистичних методів є **актуальним** під час опрацювання великих масивів даних у лінгвістичних дослідженнях.

Слід зауважити, що квантитативні методи рідко використовуються в лінгвістичних працях. Серед сучасних досліджень, у яких застосовуються лінгвостатистичні методи можна назвати роботи [1; 2; 4; 6; 7].

Метою статті є аналіз основних понять та інструментів математичної статистики, застосування основних методів та прийомів статистичного дослідження мовного матеріалу та формування статистичних вибірок у лінгвістиці.

Виклад основного матеріалу. Оскільки мова – це ймовірнісна, а не жорстко детермінована система, то для її пізнання розроблено квантитативні методи, пов'язані з дослідженням частотних, ймовірнісних та інших нелогічних характеристик [3; 5].

Головним завданням статистичної лінгвістики є застосування математичних методів для розкриття закономірностей функціонування одиниць мови у мовленні, а також у встановленні закономірностей будови тексту.

Розрізняють кількісні й статистичні методи. Кількісні методи зводяться до простого підрахунку частоти вживання мовних одиниць. Статистичні методи передбачають використання різних формул для виявлення правил розподілу мовних

одиниць у мовленні, для виміру зв'язків між мовними елементами, для встановлення тенденцій у розвитку та функціонуванні мови та для встановлення залежності між якісними та кількісними характеристиками мови.

Коли у роботі дослідника застосовується статистичний аналіз, то у більшості випадків лінгвісту потрібно вирішити два основних завдання – встановити подібність (зв'язок, близькість) або відмінність між отриманими кількісними сукупностями. З цією метою застосовується порівняно невелика кількість методів: критерій Стьюдента, критерій хі-квадрат і коефіцієнти спряженості [3, с. 132–133].

Розпочинаючи лінгвостатистичний експеримент мовознавець повинен чітко визначити, що рахувати, тобто, що є об'єктом дослідження; для чого рахувати, інакше кажучи, з якою метою виконується аналіз; і як рахувати чи, які методи застосовуються для розрахунків.

Це означає, що дослідник, по-перше, визначає ті лінгвістичні одиниці, форми та категорії, які будуть статистично обстежуватися в експерименті. Підрахунок можуть підлягати одиниці будь-якого рівня – фонемі або звуку, літери, сполуки звуків, фонем або літер, склади, морфемі, слова, словосполучення, синтаксичні конструкції, тексти тощо. По-друге, мовознавець визначає мету і завдання, які стоять перед статистичним дослідженням, тобто істотні чи неістотні розходження між вибірками. І останнє завдання лінгвіста передбачає визначення обсягу досліджуваного матеріалу для отримання статистично вірогідних результатів, знання правил організації вибірки, обчислення статистичних характеристик, зіставлення статистичних параметрів різних сукупностей, застосування методів визначення зв'язку досліджуваних одиниць у сукупності тощо.

Одним із важливих понять лінгвістичної статистики є поняття генеральної сукупності. Генеральною сукупністю називають однорідний масив (або корпус) деяких одиниць, які належить обстежити.

Наприклад, якщо ми досліджуємо особливості стилю Стівена Кінга, то всі його твори становлять генеральну сукупність. Якщо дослідженню підлягає, скажімо, рекламний дискурс журналу *New Yorker* 1925–1945 рр., то генеральну сукупність становлять собою всі рекламні тексти всіх журналів *New Yorker*, виданих за цей період. При дослідженні структури слова в англійській мові генеральною сукупністю буде словниковий склад цієї мови (реєстр найбільшого словника, доповнений даними різного роду інших словників).

З генеральної сукупності виділяють вибірку, тобто сукупність отриманих для аналізу елементів. Одиниці вибірки – це лінгвістичні одиниці, наприклад, аналізовані слова, речення, фрагменти текстів. Вибірку потрібно організувати так, щоб вона достатньо об'єктивно, надійно і в повній мірі представляла властивості всієї генеральної сукупності.

Кількість відібраного матеріалу може бути дуже великою. Завдання полягає в тому, щоб визначити той мінімум, який був би необхідним і достатнім для одержання надійних результатів. Статистика дійшла висновку, що частота досліджуваного явища залежить від обчислення відносної похибки: чим більша частота, тим менше можна взяти матеріалу для одержання результатів такої ж надійності.

Загальний обсяг вибірки можна знайти за таблицею залежності кількості прикладів від заданої відносної похибки (δ) [5, с. 152]. Зазвичай у статистичних дослідженнях прийнятою вважається похибка у 5% (у випадку найчастотніших одиниць) [5, с. 28]. За даними вищезгаданої таблиці, щоб досягнути 5% похибки, необхідно мати 1538 прикладів. Для низькочастотних одиниць у лінгвістиці допустима похибка 20-30%, тобто 96-43 приклади.

У лінгвістичних дослідженнях типовим завданням є зіставлення двох або більше варіаційних рядів, щоб визначити, чи існує між ними відмінність. Найчастіше для цього використовується критерій хі-квадрат, однак два ряди можуть бути зіставлені за допомогою критерія Стьюдента.

Цей метод полягає в тому, що порівнюються середні арифметичні двох сукупностей і за допомогою стандартної похибки t визначається достовірність різниці між середніми арифметичними.

Формула для визначення достовірності різниці між середнім арифметичним [3, с. 75]:

$$t = \frac{d}{S_d}, (1)$$

де d – різниця між середніми арифметичними двох сукупностей, а S_d – середня похибка цієї різниці. S_d знаходиться за формулою [3, с. 76]:

$$S_d = \sqrt{\frac{\sum (x_1 - \bar{x})^2 + \sum (y_1 - \bar{y})^2}{(n_1 - 1)(n_2 - 1)}} \times \frac{(n_1 + n_2)}{(n_1 \times n_2)}, (2)$$

де Σ – сума, x_1 і y_1 – абсолютні частоти двох вибірок, \bar{x} і \bar{y} – середні арифметичні першої та другої сукупностей, n_1 і n_2 – кількість підвибірок у вибірках.

Розглянемо приклад.

Приклад 1. Припустимо, що досліджувалися складні речення у двох прозових творах різних авторів (див. табл. 1).

Таблиця 1

Розрахунок даних для визначення S_d

x_i	$x_1 - \bar{x}$	$(x_1 - \bar{x})^2$	y_i	$y_1 - \bar{y}$	$(y_1 - \bar{y})^2$
29	-1,8	3,24	20	-1	1
25	-5,8	33,64	16	-5	25
38	7,2	51,84	27	6	36
30	0,8	0,64	21	0	0
32	1,2	1,44	18	-3	9
			24	3	9
		90,8			80

У першому випадку отримані числа: 29, 25, 38, 30, 32; у другому – 20, 16, 27, 21, 18, 24. Середнє арифметичне для першої сукупності $\bar{x} = 30,8$; середнє арифметичне для другої сукупності $\bar{y} = 21$; кількість першої підвибірки $n_1 = 5$ кількість другої підвибірки $n_2 = 6$.

$$\text{Отже, } S_d = \sqrt{\frac{90,8 + 80}{(5-1)(6-1)}} \times \frac{5+6}{5 \times 6} = \sqrt{\frac{171}{20}} \times \frac{11}{30} = \sqrt{3,14} = 1,77$$

$$t = \frac{30,8 - 21}{1,77} = 5,54$$

Щоб визначити значущість t , нам потрібно знати величину, яку називають ступенем вільності, i , яка обчислюється за допомогою формули [3, с. 76]:

$$d_f = (n_1 - 1) + (n_2 - 1), \quad (3)$$

Отже, $d_f = (5 - 1) + (6 - 1) = 4 + 5 = 9$. За таблицею критичних значень t [5, с. 156] знаходимо, що при $d_f = 9$ і рівні значимості $P = 0,05$ t повинно дорівнювати 2,26, а при $P = 0,01$, $t = 3,25$. За нашими обчисленнями $t = 5,54$, що більше, ніж 3,25 і 2,26. Це свідчить, що між двома дослідженими сукупностями спостерігається суттєва відмінність.

За допомогою критерія χ^2 можна встановити подібність чи відмінність між двома і більше емпіричними розподілами або ступінь відповідності емпіричного розподілу теоретичному. Найбільш розповсюдженою формулою для обчислення критерія χ^2 є [3, с. 88]:

$$(4)$$

де O – емпіричні величини (що фактично спостерігаються), E – теоретично очікувані величини.

Продемонструємо обчислення суми χ^2 на прикладі.

Приклад 2. Припустимо, що ми маємо дані про частоту вживання семи слів $V_1 - V_7$ у творах двох авторів (див. табл. 2).

Таблиця 2

Слово	Автор		Всього
V_1	10	20	30
V_2	15	25	40
V_3	20	30	50
V_4	25	35	60
V_5	25	45	70
V_6	30	20	50
V_7	50	35	85
Всього	175	210	385

Емпіричні величини – це частоти, наведені у табл. 2. Щоб знайти теоретично очікувані величини, розбиваємо сумарні числа 175 і 210 пропорційно сумарним числам, представленим в останньому стовпчику таблиці (30, 40, 50 тощо). Наприклад, щоб знайти теоретично очікувану величину для першого поля в другому стовпчику, потрібно $\frac{30}{385} \times 175 = 13,6$. Величина, яка фактично спостерігається – 10.

Дані для розрахунку χ^2 знаходяться в наступній таблиці (див. табл. 3).

Таблиця 3

Дані для розрахунку χ^2

O	E	O - E	(O - E) ²	
10	13,6	-3,6	12,96	0,95
15	18,2	-3,2	10,24	0,56
20	22,7	-2,7	7,29	0,32
25	27,3	-2,3	5,29	0,19
25	31,8	-6,8	46,24	1,45
30	22,7	+7,3	53,29	2,35
50	38,6	+11,4	129,96	3,37
20	16,4	+3,6	12,96	0,79
25	21,8	+3,2	10,24	0,47
30	27,3	+2,7	7,29	0,27
35	32,7	+2,3	5,29	0,16
45	38,2	+6,8	46,24	1,21
20	27,3	-7,3	53,29	1,95
35	46,4	-11,4	129,96	2,80
£385	£385			£16,84

Отже, сума $\chi^2 = 16,84$. Оскільки в табл. 2 порівнювалась відповідність частот розподілення слів в творах двох авторів, отримана сума χ^2 повинна дати відповідь, чи є подібність між емпіричними рядами, чи розподілення частот слів у творах двох письменників суттєво відрізняється. Таким чином необхідно визначити значущість отриманої суми χ^2 . Для цього обчислюється ступінь вільності за формулою [3, с. 89]:

$$d_f = (r - 1)(c - 1), \quad (5)$$

де r – число горизонтальних рядків у таблиці, c – число вертикальних стовпчиків у таблиці. Для табл. 2

$$d_f = (7 - 1)(2 - 1) = 6 \times 1 = 6.$$

За таблицею критичних значень χ^2 [5, с. 155] знаходимо, що при $df = 6$ і рівні значимості $P = 0,05$, сума χ^2 повинна дорівнювати 12,6, а при $P = 0,01$ сума $\chi^2 = 16,8$. Відповідно, отримана в результаті обчислень сума $\chi^2 = 16,84$ є значущою. З цього можна зробити висновок, що між розподілом частот двох авторів спостерігаються значущі відмінності.

Висновки. Існування тісного взаємозв'язку між кількісними та якісними характеристиками мови передбачає застосування математичного апарату статистичної лінгвістики, яка займає вагомe місце у вирішенні проблем мовознавства. За допомогою статистичних методів можна дослідити всі рівні мови. Застосування того чи іншого статистичного метода підкоряється перш за все виконанню завдань, які поставлені в дослідженні і залежать від умов експерименту.

Перспективи дослідження. Подальші дослідження пов'язані з проведенням лінгвостатистичного експерименту на матеріалі дискурсу американської комерційної реклами.

Література:

1. Григорійчук Т. Кількісна характеристика синонімів / Т. Григорійчук // Наук. вісн. Чернів. ун-ту. Германська філологія. – Вип. 168. – Чернівці, 2003. – С. 32–45.

2. Жуковська В. В. Лексико семантичні та прагматичні характеристики англійського дієслова у художньому тексті : лінгвостатистичний аспект (на матеріалі популярних політичних романів) : дис. ... канд. філол. наук : 10.02.04 / Вікторія Вікторівна Жуковська. – К., 2006. – 307 арк.
3. Левицкий В. В. Квантитативные методы в лингвистике / В. В. Левицкий. – Черновцы : Рута, 2004. – 190 с.
4. Левицкий В. В. Статистическое изучение лексической семантики / В. В. Левицкий. – К. : УМК ВО, 1989. – 155 с.
5. Перебийніс В. І. Статистичні методи для лінгвістів : навч. посіб. / В. І. Перебийніс. – Вінниця : Нова книга, 2001. – 268 с.
6. Перебийніс В. С. Частотні словники та їх використання / В. С. Перебийніс, М. П. Муравицька, Н. П. Дарчук. – К. : Наукова Думка, 1985. – 204 с.
7. Altmann G. Einführung in die quantitative Lexikologie / G. Altmann, D. Baugheri, H. Goebel, R. Köhler, C. Prün. – Göttingen : Peust / Gutschmidt Verlag, 2002.

УДК 37.018.43=03.111=161.2

Б. І. Шуневич, І. М. Дробіт,

Львівський державний університет безпеки життєдіяльності, м. Львів

КОМП'ЮТЕРНІ СЛОВНИКИ ПОЖЕЖНО-ТЕХНІЧНИХ ТЕРМІНІВ: ЛЕКСИЧНІ МАТЕРІАЛИ, ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ

У статті подано досвід укладання студентами напряму підготовки «Філологія» кваліфікації «Перекладач» Львівського державного університету безпеки життєдіяльності традиційних і комп'ютерних словників з галузі пожежної та техногенної безпеки під час практичних занять з дисципліни «Комп'ютерна лексикографія» спільно з викладачами відповідних технічних кафедр. Описано лексичну базу традиційних словників та програмне забезпечення для укладання комп'ютерних словників з цієї тематики на прикладі «Короткого українсько-англійського словника зі сфери надзвичайних ситуацій» та «Англійсько-українського словника пожежно-технічних термінів».

Ключові слова: комп'ютерний словник, комп'ютерна лексикографія, термін, переклад.

В статті представлено досвід складання студентами напрямку підготовки «Філологія» кваліфікації «Перекладач» Львівського державного університету безпеки життєдіяльності традиційних і комп'ютерних словників з галузі пожежної та техногенної безпеки під час практичних занять з дисципліни «Комп'ютерна лексикографія» спільно з викладачами відповідних технічних кафедр. Описано лексичну базу традиційних словників та програмне забезпечення для складання комп'ютерних словників з цієї тематики на прикладі «Короткого українсько-англійського словника галузі надзвичайних ситуацій» та «Англо-українського словника пожежно-технічних термінів».

Ключевые слова: компьютерный словарь, компьютерная лексикография, термин, перевод.

Modern terminology, traditional and computer lexicography among other set a practical aims to meet the needs of society by ordering, organizing and standardizing terminology, compiling dictionaries. The paper describes experience of compiling conventional and computer dictionaries by students of Lviv State University of Life Safety. The students' major subject is English translation. Students compile unidirectional bilingual dictionaries during practical lessons on Computer Lexicography under the supervision of lecturers and instructors from the Department of Foreign Languages and Technical Translation taking into consideration advice provided by lecturers and instructors from relevant technical departments of the University.

The article describes lexical database and software for compiling computer «Concise Ukrainian-English Dictionary of Emergencies» and «English-Ukrainian Dictionary of Fire and Technical Terms.» The dictionaries were compiled grounding on Ukrainian, English and Russian dictionaries, glossaries, reference books, and terminology standards. PolyDic v.1.0 was chosen as software tool to compile electronic dictionaries. It was created by experts of the Technical Committee for Scientific and Technical Terminology Standardization at Lviv Polytechnic National University. The software tool consists of two programs: PolyDic Editor – for typing, formatting and editing dictionary databases, and PolyDic Viewer – for viewing electronic dictionaries.

We consider it advisable to involve in compiling computer dictionaries not only linguists and specialists in the industry for which the dictionary is written, but also students. Such type of work motivates them to use computer dictionaries, including the ones which they compiled, in further professional activities of a translator, and to research terms.

Keywords: computer dictionary, computer lexicography, term, translation.

Фундаментальні дослідження, розроблення і впровадження новітніх технологій у практичну діяльність підрозділів державної служби України з надзвичайних ситуацій мають важливе наукове і практичне значення. Спілкування наукової спільноти у галузі пожежної та техногенної безпеки відбувається різними мовами, оскільки надзвичайні ситуації інколи мають не локальний, а міжнародний, або й глобальний характер (наприклад, ліквідація наслідків аварій на атомних електростанціях, масштабних стихійних лих, тощо). Для цього необхідне добре знання міжмовних відповідників. Поширення наукових здобутків, практичних навиків серед іноземних партнерів у справі забезпечення пожежної та техногенної безпеки, спеціалістів сфери цивільного захисту та фахівців суміжних наукових дисциплін також потребує відповідного мовного забезпечення.

Як слушно зауважив директор Інституту української мови НАН України П. Гриценко, «центральною проблемою сучасного термінознавства залишається упорядкування термінології різних галузей знань із постійною увагою до точних визначень наукових понять та зв'язків між ними, що має знаходити відтворення у відповідних термінопозначеннях» [2, с. 6].

Це спонукало авторський колектив Львівського державного університету безпеки життєдіяльності (ЛДУ БЖД) до укладання низки галузевих словників, а також дослідження української та іншомовної термінології цієї тематики. Зокрема, укладено іншомовно-українські пожежно-технічні словники-мінімуми термінів обсягом, відповідно, 3000, 1600, 2000 і 5000 термінів [1; 4; 5; 7], а також ілюстрований словник-довідник «Цивільний захист. Частина перша: Пожежно-рятувальна справа» обсягом понад 3000 термінів [3]. Після публікування словників, для збереження за собою авторського права на ці словники, ведеться робота над укладанням на їхній основі однойменних комп'ютерних словників.

Мета нашої статті – поділитися досвідом укладання студентами напряму підготовки «Філологія» кваліфікації «Перекладач» традиційних і комп'ютерних словників з галузі пожежної та техногенної безпеки під час практичних занять з дисципліни «Комп'ютерна лексикографія» спільно з викладачами відповідних технічних кафедр, описати лексичну базу традиційних словників та програмне забезпечення для укладання комп'ютерних словників з цієї тематики на прикладі «Короткого українсько-англійського словника зі сфери надзвичайних ситуацій» та «Англійсько-українського словника пожежно-технічних термінів».