

Методи у дослідженні фоносемантичних явищ

Р. І. Мельничук

Чернівецький національний університет імені Юрія Федьковича, м. Чернівці, Україна

Paper received 17.05.2016; Accepted for publication 29.05.2016.

Анотація. Статтю присвячено висвітленню методологічної бази, яка використовується для дослідження фоносемантичних явищ. Розглянуто тенденції використання статистичних та кількісних методів у вище вказаних лінгвістичних дослідженнях. Запропоновано ефективні статистичні прийоми з метою розкриття явищ звуконаслідування та визначення характеру мотивації мовної одиниці. Комплексне застосування методів сприяє виявленню таких обставин функціонування німецької мови, які іншим способом не досліджуються; створює можливості для глибокої всесторонньої інтерпретації причин і результатів мовної еволюції.

Ключові слова: кількісні і статистичні методи дослідження мови, *хі-квадрат*, *квантитативний метод*, *кореляційний аналіз*.

Вступ. Для будь-якого сучасного лінгвістичного дослідження актуальним залишається загальнонаукове і філософське твердження про органічну єдність якісного та кількісного визначення предметів і явищ: «пізнати – означає виміряти» [2, с. 20]. У наукознавстві підкреслюється, що саме дія закону переходу кількості у якість служить важливою умовою виявлення закономірностей існування об'єктів та явищ, у тому числі мовних. Тому погоджуємося із думкою В. К. Журавльова, що комплексне застосування різних методів, у тому числі статистичних, допоможе отримати нову інформацію про лінгвістичні об'єкти, конкретизувати факти, знайдені на інших шляхах історико-лінгвістичного пізнання [1, с. 50].

Короткий огляд публікацій за темою. Серед науковців, які зробили свій внесок у розвиток квантитативної лінгвістики, варто назвати як вітчизняних учених (Н. Г. Єсипенко, В. І. Кушнерик, В. В. Левицький, О. Д. Огуй, І. Перебийніс, Б. Тошович), так і зарубіжних (Д. Алтман, К. Беста, К. Мертен).

Мета статі – проаналізувати кількісні та квантитативні методи дослідження фоносемантичних явищ і продемонструвати, які формули можуть стати у нагоді при дослідженні звуко-символічної та звуко-наслідувальної систем.

Виклад основного матеріалу дослідження. Якщо природничі науки порівняно давно «заговорили» мовою математики, то гуманітарні звернулися до неї значно пізніше: тільки у 40-50-их роках ХХ-го століття, при цьому першою з них була саме лінгвістика. І. А. Бодуен де Куртене, підводячи підсумки розвитку мовознавства ХІХ століття, наголошував на тому, що потрібно все частіше застосовувати у мовознавстві кількісне математичне мислення і таким чином більше наблизити його до точних наук [5, с. 345]. Дане твердження науковця виявилось беззаперечним, тому у наші дні виявлено основне: «для мови безсумнівним є квантитативний фактор» [3, с. 14].

Використання квантитативних методів при наукових дослідженнях дозволяє глибоко проникати у їх сутність. Оскільки деякі елементи мови і мовлення зумовлюються певною частотою і величиною, то мовленнєвий процес, відповідно, досліджується за допомогою методів математичної статистики. Учені-лінгвісти не могли не помітити статистичних характеристик елементів мовлення, тому зародження лінгвістичної статистики із використанням різних лінгвістичних методів і прийомів було неминучим.

Порівняльно-історичний метод (компаративний, лінгвогенетичний), який виник на початку ХІХ ст. – це сукупність прийомів і процедур історико-генетичного дослідження мовних сімей і груп, а також окремих мов для встановлення закономірностей їх розвитку [5, с. 342]. Його основоположниками є німецькі вчені Ф. Бопп і Я. Грімм, датський мовознавець Р. Раск і російський вчений О. Х. Востоков. Цей метод ґрунтується на наукових прийомах відтворення (реконструкції) мовних фактів шляхом планомірного порівняння відповідних пізніших фактів, відомих безпосередньо за їх уживанням у мовленні. Порівняльно-історичний метод дозволяє аналізувати еволюційні зміни та закономірності розвитку фонологічної системи німецької мови шляхом порівняння її становлення та розвитку на різних етапах формування (тобто, у динаміці) та встановити спільні фоносемантичні константи і відмінності на матеріалі досліджуваної нами мови, а також уможливиле виявлення та порівняння зміни звуків німецької мови, які мають строго закономірний характер, оскільки корені та флексії є стійкими впродовж тисячоліть, що дає змогу реконструювати архетипи.

Зіставний метод дозволяє виявити спільні і специфічні ознаки функціонування голосних і приголосних фонем німецької мови протягом останніх 30-ти десятиліть з метою виявлення причин подібностей та відмінностей у системі мови та в процесі еволюції.

Найдавніший і найпоширеніший *описовий метод*, базу якого у нашому випадку творять факти німецької мови, ґрунтується на спостереженні об'єкта аналізу, планомірній інвентаризації мовних одиниць та поясненні особливостей їх будови та функціонування у синхронії (тобто, на певному етапі розвитку мови) [3, с. 201]. Служить для вивчення фактів мови на рівні основних її ярусів з точки зору синхронічного аспекту, об'єднуючи усі методики і прийоми у дві групи: зовнішню та внутрішню інтерпретацію.

Прийоми внутрішньої інтерпретації вивчають мовні явища на основі стратифікації і систематизації синтагматичних і парадигматичних зв'язків. При цьому досліджувані одиниці (наприклад, морфеми) класифікуються за місцем і роллю у слові, участю у словотворенні, продуктивністю і частотністю. Внутрішня реконструкція мовного фонетичного процесу ґрунтується на гіпотезі регулярності дії фонетичних законів, суть яких передається наступною формулою:

$$L \left\{ \frac{a>b}{P} T \right\} \quad (1)$$

У мові L (langue) звук $[a]$ систематично переходить у звук $[b]$ у позиції P на етапі розвитку T [4, с. 42].

Мета *описово-аналітичного методу* у фоносемантичному дослідженні – це планомірна інвентаризація мовних одиниць у поетичних колиськових і маршових текстах та формування зв'язку фонологічних одиниць мови з асоціативно-символічним значенням, встановлення закономірностей функціонування фонем у сучасній німецькій мові та інтерпретація результатів відповідно до фонематичного статусу досліджуваних фонем на основі вербального експерименту в письмовій формі. У нашій роботі – це проведення психолінгвістичного експерименту, який ґрунтується на наукових прийомах *психолінгвістичного методу*. За допомогою даного методу опрацьовано та проаналізовано мовні факти, отримані від носіїв мови в результаті спеціально організованого експерименту з вербальними реакціями на відповідні мовленнєві стимули (у нашому випадку мова іде про ізольовані фонемі, дифтонги та фонестеми). Для кращого розуміння природи досліджуваних фактів мови слід насамперед досліджувати людську свідомість, яка є користувачем і вмістилищем всього того, що прийнято називати мовою. Тому при проведенні експерименту ми звертаємося до інтуїції дослідника та інформантів, на основі якої і формується гіпотеза наукового пошуку.

При аналізі лексики *методом компонентного аналізу* ми розглядали обмежені лексичні підсистеми, які виділяли на основі семантичного критерію. Обмежену лексику досліджували шляхом виділення відносно самостійних відрізків простору.

Метод апроксимативно-кількісного аналізу застосовували для виявлення частоти вживання фонологічних одиниць та розгляду динаміки їх функціонування та функціонального навантаження у німецькій мові.

Лінгвостатистичний метод – це сукупність статистичних (міжпарадигмальних) методик, прийомів і процедур, які використовуються для отримання кількісних даних про мовні явища з метою виявлення статистичних закономірностей функціонування одиниць мови та мовлення, а також для встановлення закономірностей побудови текстів [4, с. 89].

Слово «статистика» спочатку означало науку про стан держави (status – стан, лат.), яка займалася питаннями чисельності й армії, а також деякими господарськими аспектами. Як така, статистика сягає своїми коренями у сиву давнину; відомо, наприклад, що переписи населення проводилися задовго до початку нашої ери. Проте основи статистики в більш широкому значенні були закладені лише в XVII ст. в Англії і Голландії.

Слід зазначити, що перші серйозні спроби застосовувати квантитативні методи у мовознавстві були зроблені на початку минулого століття математиками Н. А. Морозовим і А. А. Марковим. Ще тоді вони вказували на те, що, досліджуючи певне явище, не можна обмежитись голослівними ствердженнями типу «більше», «менше», «рідше», «частіше», а потрібно оперувати числами, встановлювати рівень надійності отриманих результатів і вказувати на їх

можливі коливання [4, с. 56]. Звідси стає зрозуміло, що загальні квантитативні методи в мовознавстві спочатку вводилися в основному математиками, які у мовленні шукали і знаходили математичні закономірності.

Потреба у використанні статистики у мовознавстві ставала все більш неминучою, і в 1949 року французький лінгвіст Марсель Коен заявив, що тривале нехтування числом у лінгвістиці приведе до гальмування її розвитку. Інший сучасний французький мовознавець П'єр Гіро відносить мовознавство до типічних статистичних наук, оскільки «мова, по ходу справи, є статистичним явищем», тобто таким, яке охоплює обчислювальні елементи, які проявляють статистичні тенденції, а тому вони можуть бути визначеними й інтерпретованими у числах. На його думку, частота мовних знаків володіє функціональним значенням і, як висновок, «статистична мовна функція є не менш важливішою і не менш об'єктивнішою, ніж семантична функція, і стільки ж потрібно для повного і правильного розуміння мовних фактів» [1, с. 60].

Слід зупинитися на тому, що у сукупності математичних прийомів умовно розрізняють «кількісні» і «некількісні» методи. За допомогою останніх (теорія множин, теорія алгоритмів, математична логіка) можна вивчати, перш за все, систему мови (langue). Даний розділ науки отримав назву «комбінаторна лінгвістика». Використовуючи кількісні методи (теорія ймовірностей і математична статистика), можна досліджувати текст (parole). Саме цей другий напрямок і є квантитативною лінгвістикою.

Говорячи про роль статистичних методів у дослідженні мови, один із радянських лінгвістів Б. Н. Головін у передумові до своєї книги «Мова і статистика» (М., 1970) написав: «Викладачі-філологи вищих навчальних закладів потребують зараз не математизації своєї мови і наукових уявлень, а допомоги математики, яку вони відчують тоді, коли зможуть застосовувати деякі найбільш доступні інструменти лінгвостатистики» [1, с. 59].

Відповідно до ступеня математизації дослідження у галузі квантитативної лінгвістики можна умовно поділити на три рівні.

Верхній рівень – це рівень використання апарату математичної лінгвістики (теорія множин, теорія алгоритмів, математична логіка). Застосування даного рівня потребує спеціальної математичної освіти. Праці, що сюди належать, характеризуються пошуками «моделей» і «законів», використанням важких формул, які займають значний об'єм площі наукової публікації і є недостатньо зрозумілими лінгвісту. Результати, отримані у роботах такого роду, зводяться до обговорення наступного: якому теоретичному розподілу відповідає отримана емпірична величина. Хоча такі праці містять суттєві відкриття і мають безсумнівну евристичну цінність, вони вважаються нецікавими і нудними для лінгвіста.

Середній рівень характеризується використанням таких статистичних показників як середнє арифметичне, середнє квадратне відхилення і пропонує оперування наступними поняттями: біноміальне, нормальне, коефіцієнт варіації, коефіцієнт асиметрії,

критерій хі-квадрат, коефіцієнт кореляції. Для їх оволодіння не обов'язково мати спеціальну математичну освіту, проте слід оперувати поглибленими знаннями лінгвістичних методів.

Нижчий рівень можна було б обмежити умінням збирати елементарні дані про частоту поширення мовних факторів і представляти їх у вигляді відсоткового відношення. Проте це не є достатнім для отримання достовірних результатів [4, с. 60].

Одним із найефективніших статистичних прийомів у дослідженні фоносемантичних явищ вважаємо використання критерію середнього рівня хі-квадрат (χ^2). Оскільки мова розглядається не тільки в якісному, але й у кількісному відношенні, то у лінгвістиці існує багато понять, пов'язаних із мірою (наприклад, тенденція до аналітичності). Тому при описі мови ми іноді використовуємо ряд термінів, опосередковано пов'язаних із кількістю, а саме статистичний прийом, за допомогою якого вимірюється ступінь відповідності фактичних даних і теоретично очікуваних. Представимо формулу для обчислення критерію χ^2 :

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

де O – фактично очікувана величина, E – теоретично очікувана величина [2, с. 78].

Ми використовуємо такі допоміжні статистичні поняття при користуванні, як рівень значущості "P" і кількість ступенів вільності.

Так, рівень значущості дозволяє судити про ступінь достовірності отриманих у результаті розрахунків величин. Коли явище встановлено з рівнем значущості $P = 0,05$ (якщо величина суми хі-квадрат менша ніж 3,84 при $df = 1$ і $P < 0,05$), це означає, що випадкове відхилення від отриманих результатів складає 0,05 або, іншими словами, можливість помилки дорівнює 5%. Отже, певним рівням значущості відповідають певні значення ймовірності. При ймовірності 99,9% рівень значущості $P = 0,001$, при ймовірності 99% рівень ймовірності $P = 0,01$ (якщо величина суми хі-квадрат менша ніж 6,63 при $P = 0,01$).

При статистичній обробці матеріалу прийнято вважати, що 5 %-й рівень значущості ($P = 0,05$) є мінімальним для того, щоб розглядати отримані величини як достовірні. Природно, що з підвищенням вимог до достовірності результатів підвищується й рівень значущості. Вказану величину визначають за спеціальними статистичними таблицями [2, с. 87]. Для її визначення необхідне знання величини числа ступенів вільності. Не заглиблюючись у більш детальний аналіз цього поняття, обмежимося розглядом зв'язку кількості ступенів вільності з рівнем значущості, оскільки це має безпосереднє відношення до статистичної інтерпретації даних експериментів.

Існуючий зв'язок кількості ступенів вільності і рівня значущості можна сформулювати так: зі збільшенням кількості об'єктів, які вивчаються, знижуються вимоги до величини, рівень значущості якої визначається. Інакше кажучи, зафіксованому зв'язку властива обернено пропорційна залежність.

Необхідно підкреслити, що сам хі-квадрат є лише показником наявності чи відсутності зв'язку, а як міру

зв'язку використовують коефіцієнт взаємної спряженості K , або, як його ще називають, "коефіцієнт Чупрова", який обчислюється за формулою:

$$K = \sqrt{\frac{\chi^2}{N \cdot df}} \quad (3),$$

де K — коефіцієнт взаємної спряженості, N — сума частот у вихідній таблиці, χ^2 – величина хі-квадрата, df – кількість ступенів вільності. Величини коефіцієнта K коливаються від 0 до 1. Зазначимо, що при виході на одиницю коефіцієнтів, коли величини N та df ідентичні для всіх таблиць, як умовну міру зв'язку можна допустити величину 1.

Для перевірки лінгвістичних гіпотез, коли специфіка досліджуваного матеріалу не дозволяє отримати достатню кількість вибірок, будують чотирипільну таблицю спряжених ознак. Представимо схему таблиці спряжених ознак:

Схема чотирипільної таблиці

A	b	a+b
C	d	c+d
a+c	b+d	a+b+c+d

При цьому числа у таблиці повинні бути розташовані так, щоб $a \geq b$, $a \geq d$. Сума χ^2 у даному випадку обчислюється за формулою:

$$\chi^2 = \frac{(ad-bc)^2 N}{(a+c)(b+d)(a+b)(c+d)} \quad (4),$$

де a, b, c, d – емпіричні величини у таблиці спряжених ознак, N – загальне число спостережень.

Для чотирипільних таблиць коефіцієнт взаємної спряженості K обчислюється за наступною формулою:

$$K = \sqrt{\frac{\chi^2}{N}} \quad (5),$$

де N — загальне число спостережень, сума частот у вихідній таблиці, χ^2 – величина хі-квадрата.

Для визначення залежностей між різними одиницями і встановлення причинного зв'язку використовуємо *метод кореляційного аналізу*, метою якого є виявлення спільних тенденцій, тобто, прямого та зворотнього кореляційного зв'язків досліджуваних нами фоносемантичних явищ. Мова іде про пряму кореляцію, коли при збільшенні частоти одного явища збільшується частота інших параметрів, а про зворотню – коли при збільшенні частоти одного явища частота інших параметрів навпаки зменшується.

Для обчислення рангового коефіцієнту, який є найпростішим видом кореляційного аналізу, В. І. Перейбініс подає наступну формулу:

$$\rho = 1 - \frac{6 \sum d^2}{n^3 - n} = \frac{6 \sum d^2}{n(n^2 - n)} \quad (6),$$

де ρ – критичне значення, n – кількість зіставляваних рядів [2, с. 90].

В. В. Левицький стверджував, що величина коефіцієнта кореляції вказує на ступінь зв'язку між явищами, а його знак демонструє характер зв'язку. Тому учений подає наступну найпоширенішу формулу для обчислення коефіцієнта кореляції:

$$r = \frac{\sum(x_1 - \bar{x})(y - \bar{y})}{\sqrt{\sum(x_1 - \bar{x})^2 \sum(y - \bar{y})^2}} \quad (7)$$

де r – коефіцієнт лінійної кореляції, \sum – сума величин, отриманих у результаті певної операції, x_1 – значення першої ознаки, y_1 – значення другої ознаки, \bar{x} – середня (теоретична) величина значень першої ознаки, \bar{y} – середня (теоретична) величина для іншої ознаки [2, с. 123].

Результати. Тенденція щодо застосування статистичних прийомів при дослідженні лінгвістичних явищ обґрунтовується наступними чинниками:

- об'єктивною властивістю мови і кількісних характеристик;
- внутрішньою залежністю, яка існує між якісними і кількісними характеристиками мовної структури;
- підвладністю частоти різних елементів мови у мовленнєвому потоці тим чи іншим статистичним законам.

Висновки. Важливість застосування статистичних методів у лінгвістиці беззаперечна, адже багато мовних закономірностей не можна розглядати лише, як абсолютні закони, їх слід трактувати як правила, які впливають із статистичної ймовірності. Слід сказати, що неможливо у мові обійтись без таких категорій, як мало та багато; несуттєво, суттєво; правило, виняток. Фактично вони є джерелом квантитативної лінгвістики, яка намагається надати поняттям, пов'язаним із кількісними показниками, точного змісту. Значить, мову необхідно розглядати за допомогою чітких статистичних методів для аналізу мовного матеріалу і його систематизації. Таким чином, імовірно-статистична методика дозволяє отримувати результати та на їх підставі робити змістовні лінгвістичні висновки, які дозволяють глибше проникати у природу досліджуваного явища і прослідкувати певні тенденції у мовних фактах щодо характеру функціонування мовних одиниць у мовленні.

ЛІТЕРАТУРА

1. Гиро П. Проблемы и методы статистической лингвистики // Общее языкознание / П. Гиро. – Минск, 1976. – С. 44-63.
2. Левицкий В.В. Квантитативные методы в лингвистике / В.В. Левицкий. – Черновцы: Рута, 2004. – 189 с.
3. Попова З.Д. Когнитивная лингвистика / З.Д. Попова, И.А. Стернин. – М.: АСТ: Восток-Запад, 2010. – 314 с.
4. Селіванова О.О. Лінгвістична енциклопедія / О.О. Селіванова. – Полтава: Довкілля-К, 2010. – 844 с.
5. Ченки А. Семантика в когнитивной лингвистике // А. Ченки // Фундаментальные направления современной американской лингвистики: Сб. обзоров / под ред. А.А. Кибрика и др. М.: Изд-во Московского университета, 1997. – 340-369 с.
6. Bußmann, Hadumod. Lexikon der Sprachwissenschaft / H. Bußmann, H. Lauffer. – Stuttgart: Alfred Kröner Verlag, 2008. – S. 342-343.

REFERENCES

1. Nyro, P. Problems and methods of statistical linguistics // Common linguistics / P. Nyro. – Minsk, 1976. – 44-63 p.
2. Levytskyy, V.V. Quantitative methods in linguistics / V.V. Levytskyy. – Chernivtsi: Ruta, 2004. – P. 189.
3. Popova, Z.D. Cognitive Linguistics / Z.D. Popova, I.A. Sternin. – M.: AST: East-West, 2010. – 314 p.
4. Selivanova, O.O. Linguistic encyclopedia / O.O. Selivanova. – Poltava, 2010. – 844 p.
5. Chenki, A. Semantics in cognitive linguistics // A. Chenki // Fundamental trends in contemporary American Linguistics: Sb. obzorov / ed. A.A. Kibrik et al. M.: Publishing house of the Moscow University, 1997. – 340-369 p.

Methods in phonosemantic research

Melnychuk R. I.

Annotation. The article deals with the description of methodological base that is used for research of the phonosemantic phenomena. Statistical and quantitative methods of language research are considered. Effective statistical methods are offered with the aim to find out the onomatopoeia phenomena and to determine a character of language unit motivation. Complex application of methods helps to explore specific circumstances functioning of German that by another way is not investigated and creates the possibilities for deep all-round interpretation of reasons and results in language evolution.

Keywords: quantitative and statistical methods of language research, criterion χ^2 , correlation analysis, phonosemantic phenomena.

Методы в исследовании фоносемантических явлений

Р. И. Мельничук

Аннотация. Статья посвящена описанию методологической базы, которая используется для исследования фоносемантических явлений. Рассмотрены тенденции использования статистических и количественных методов у лингвистических исследованиях. Предложено эффективные статистические приемы с целью раскрытия явления звукоподражания и определения характера мотивации языковой единицы. Комплексное применение методов способствует выявлению таких обстоятельств функционирования немецкого языка, которые иным способом не исследуются; создает возможности для глубокой всесторонней интерпретации причин и результатов языковой эволюции.

Ключевые слова: количественные и статистические методы изучения языка, хи-квадрат, квантитативный метод, корреляционный анализ.