

Перебийніс В. І.

Статистичні методи для лінгвістів

Посібник

Видання друге, виправлене й доповнене

Вінниця
Нова Книга
2013

ББК 81.2 Англ
УДК 908(430)
П 27

Рецензенти:

Олексієнко Л. А. — канд. філол. наук, доцент
Дарчук Н. П. — канд. філол. наук, доцент

Відповідальний редактор: д-р філол. наук професор *Левицький В. В.*

Перебийніс В. І.

П 27 Статистичні методи для лінгвістів : посібник / Перебийніс В. І.
— Вид. 2, випр. і допов. — Вінниця : Нова Книга, 2013 — 176 с.
ISBN 978-966-382-475-8

У посібнику просто і доступно для лінгвістів, які не мають спеціальної математичної підготовки, викладено найнеобхідніші методи та процедури, прийняті у статистичних дослідженнях.

Для аспірантів, викладачів і всіх науковців, які мають намір використати статистичні методи у своїх дослідженнях.

ББК 81.2 Англ
УДК 908(430)

ISBN 978-966-382-475-8

© Перебийніс В. І., 2013
© Нова Книга, 2013

Зміст

Вступ	5
Література	10
Розділ 1. Передумови правильної організації статистичного дослідження	11
§ 1. Що рахувати?	11
§ 2. Для чого рахувати?	12
§ 3. Як рахувати?	14
§ 4. Правила організації вибірки	15
§ 5. Організація механічної вибірки	19
§ 6. Організація випадкової вибірки	22
§ 7. Організація типової (зональної) вибірки	25
§ 8. Визначення необхідного обсягу вибірки	26
Контрольні питання	32
Задачі	32
Література до розділу 1	33
Розділ 2. Обчислення основних статистичних характеристик вибірки	35
§ 1. Визначення абсолютних частот	35
§ 2. Обчислення середньої частоти	37
§ 3. Побудова інтервального варіаційного ряду розподілу	42
§ 4. Обчислення середнього квадратичного відхилення	49
§ 5. Міра коливання середньої частоти	52
§ 6. Обчислення коефіцієнта варіації	55
§ 7. Визначення відносної неточності дослідження	62
§ 8. Побудова інтервального ряду за методом 3σ	65
Контрольні питання	66
Задачі	66
Література до розділу 2	68
Розділ 3. Зіставлення статистичних характеристик різних виборок	69
§ 1. Зіставлення виборок за смугою коливання середньої	70
§ 2. Перевірка на статистичну однорідність	73

§ 3. Критерій Стьюдента	76
§4. Визначення істотності розходження процентних показників	82
§ 5. Об'єднання виборок	85
§ 6. Зіставлення варіаційних рядів	90
Контрольні питання	94
Задачі	94
Література до розділу 3	97
Розділ 4. Встановлення залежностей між різними явищами	
у вибірці	98
§ 1. Побудова таблиці спряжених ознак	98
§ 2 Рангова кореляція.....	107
§ 3. Обчислення коефіцієнта кореляції	110
Контрольні питання	117
Задачі	117
Література до розділу 4	118
Розділ 5. Непараметричні критерії відмінностей (розходжень)	
§ 1. Метод експертних оцінок	119
§ 2. Обчислення коефіцієнта подібності	122
§ 3. Критерій знаків	124
§ 4. Серійний критерій	127
§ 5. Критерій Вілкоксона.....	130
Контрольні питання	132
Задачі	132
Література до розділу 5	133
Післямова.....	134
Відповіді до задач	135
Додатки	147

Вступ

У нашому світі діють закони двох типів: детермінативні, або причинні, і статистичні. У випадку дії детермінативних законів співвідношення між причиною і наслідком точно визначене і тому однозначне, передбачити результат дії такого закону завжди легко. Багато законів фізики, хімії, біології є законами детермінативними: всі речовини з питомою вагою більше одиниці тонуть у воді, питома вага хімічного елемента визначається будовою його атома, після зими настає весна, все живе рано чи пізно вмирає і т.п.

У випадку дії законів статистичних однозначне завбачення результату певного впливу на предмет неможливе. Так, той факт, що після зими йде весна, належить до законів детермінативних. Але характер зими або весни в кожному окремо взятому році однозначно передбачити неможливо, так само як і строки її настання. Характер пори року (тепле чи холодне літо, сухе чи дощове, вітряне чи спокійне тощо) залежить від величезної кількості факторів, завбачити які людина не в змозі, а про багатьох з них взагалі поки що не здогадується. Природно, що відсутність достатніх знань про ці фактори не дозволяє однозначно визначити і результат, що є їх рівнодіючою. Тому в таких випадках людина може лише з тим чи іншим ступенем вірогідності передбачити, яким буде літо цього року, визначивши можливі межі коливання температури, кількості опадів та ін.

Слід зазначити, що сфера дії статистичних законів надзвичайно широка: всі досить складні системи підкоряються перш за все законам статистичним. А такими системами є і живі організми, і їх поведінка, і економіка, і результати діяльності великих колективів, і розвиток науки, і мова.

Зупинімося докладніше на тих передумовах і причинах, які роблять можливим і необхідним застосування статистичних методів у мовознавстві, тобто на тих рисах мови, які дозволяють стверджувати, що мова і мовлення підкоряються статистичним законам.

Першою підставою для цього твердження є те, що мова являє собою надзвичайно складну систему дискретних одиниць, які, як і всякі 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. Цікаві докази наводить В. В. Іванов:

«...деякі основні кількісні характеристики мови, особливо істотні для лінгвіста, носять дуже простий характер. Такою є, зокрема, розбіжність між числом слів (від 10 000 до 100 000), числом морфем (кілька тисяч), числом складів (від кількох сотень до кількох тисяч) і числом фонем

(від 10 до 80). Згідно з висловленим вище припущенням ці співвідношення зв'язані зі структурою людської пам'яті.

Прості кількісні співвідношення між словами, складами і фонемами дозволяють дати класифікацію мов, яку можна використати і при вивченні їх історії. Так, у випадку, якщо слова в мові односкладні, середня довжина слова занадто мала для того, щоб було можливим членування слова на частини; тому в мовах з односкладовими словами слово не ділиться на морфемі. Разом з тим необхідність розрізнення кількох тисяч морфем (слів) при обмеженому інвентарі складів робить необхідним розрізнення складів за допомогою музичного наголосу. Тому мови з односкладовими словами-морфемами завжди є мовами з музичним наголосом (в'єтнамська, класична китайська, деякі центрально-африканські мови і т.п.). Якщо в історії мови слова в результаті звичайного скорочення довжини слова стають односкладовими, то вони перестають членуватися на морфемі, й у мові розвивається музичний наголос (так пояснюється наголос у тибетській і класичній китайській мовах). Іншою істотною закономірністю є зв'язок між кількістю фонем і середньою довжиною морфемі: так, в абхазькій мові, бзібська говірка якої має рекордну кількість фонем (81), коренева морфема зазвичай буває однофонемною. У гавайській та інших полінезійських мовах морфема в середньому складається з двох складів (4 фонем), односкладові морфемі становлять незначний відсоток загальної кількості морфем; разом з тим є нерозкладні морфемі з 20 фонем (11 складів), наприклад, назва риби *homohotonukunukuapia*.

Знання цих залежностей можна використати в порівняльно-історичному мовознавстві: так, якщо для певної мови (наприклад, пракартвельської) припускається однофонемність кореня, то це з необхідністю потребує припущення про розвинену систему фонем, що перевищує середню норму (40 фонем). Необхідність кількісного дослідження явно чи підсвідомо на практиці визнається всіма лінгвістами, хоча вони іноді заявляють протилежне» [Иванов, 1963, С. 176–177].

Дійсно, наявність у мові кількісних характеристик визнається в явному чи в неявному виді всіма мовознавцями. Свідченням цьому є хоча б використання при описі мови таких кількісних понять, як «більше», «менше», «багато», «часто», ін. Звичайно, такі приблизні оцінки не мають ані доказовості, ані наукової ваги. Саме тому відомий російський математик В.Я.Буняковський ще в 1847 році говорив:

«...хай дозволено мені буде додати кілька слів про інше використання аналізу ймовірностей...

Нові використання стосуються граматичних та етимологічних досліджень будь-якої мови, а також порівняльної філології. Коли говориться про одну мову, то перш за все припустимо, що маємо її вичерпний арифметичний опис, чи, скажімо, її статистику, тобто числові дані про повний підсумок слів цієї мови, розподіл цих слів за частинами мови, за кількістю літер, за початковими літерами, за закінченнями тощо. Сюди ж будуть входити відомості про загальні правила, про винятки різного роду, про слова, без сумніву запозичені з інших мов, і т. п. Ось числові матеріали, точний розбір яких потребує, певна річ, міркувань математичних.

Маючи подібні статистичні дані для двох чи кількох мов, можна порівнювати їх у різних відношеннях, і результати одержать певну авторитетність, яку на своє виправдання не завжди можуть забезпечити філологи при сучасному стані науки» [Буняковський, 1847, С. 48].

Таким чином, кількісні характеристики, властиві одиницям мови і всій мові як надзвичайно складній системі дискретних елементів, можна вважати підставою для твердження про підпорядкованість мови дії статистичних законів.

Але можна й заперечити, що кількісні характеристики, про які говорилося вище, ще не доказують статистичність мови, оскільки далеко не всякі кількісні характеристики є характеристиками статистичними.

Як уже говорилося, підставою для твердження про підпорядкованість певного явища статистичним законам може служити наявність безлічі різнонаправлених факторів, що впливають на дане явище і сукупно зумовлюють його розвиток. Розглянемо, які сукупності факторів є доказом підпорядкованості мови статистичним законам.

Як і будь-яка система, мова має свій спосіб організації, або, іншими словами, свою структуру. Всі елементи структурно організовані. Це визначає їх взаємозв'язок: усі елементи мови зв'язані цілою системою відношень, і кожний елемент, тобто кожна одиниця певного мовного рівня, визначається не тільки сіткою відношень, що зв'язують одиниці даного рівня, але й відношеннями з одиницями нижчих рівнів, як ми бачили у прикладі, наведеному В. В. Івановим. Кількість відношень, якими зв'язані одиниці кожного рівня, надзвичайно велика і росте з кількістю цих одиниць. Так, якщо вивчати відношення між фонемами у двофонемних утвореннях, то кількість їх дорівнює кількості цих утворень, тобто вимірюється числом K^2 , де K — кількість фонем у даній мові. Для мови, в якій 40 фонем, це число дорівнює $1600 = 40^2$. Правда, далеко не всі сполуки фонем для кожної мови припустимі, можливі [Перебийніс, 1970, стор 80]. Існує багато пар фонем, у яких відношення між фонемами те ж саме — нульове. Але навіть якщо одна третина цих сполук

реалізується мовою, то й у цьому випадку кількість відношень залишається занадто великою, щоб можна було врахувати дію кожного з них при побудові одиниць вищого рівня (морфем) або навіть при утворенні складів чи просто двофонемних сполук. Якщо ж розглядати морфемі, склади чи сполучення фонем, які складаються з більшої кількості фонем, то число їх буде стрімко зростати. Природно, що дослідник не може навіть на фонемному рівні врахувати дію кожного цього відношення і передбачити його результати. Він може лише припустити їх з певною ймовірністю, пізнати яку допоможуть статистичні методи.

Аналогічна картина спостерігається й у мовленні. Її будова залежить від багатьох факторів, кожний з яких певним чином впливає на мовленнєвий ланцюжок: закони мови (закони будови одиниць мови, використовуваних у мовленні), закони сполучуваності одиниць мови в мовленнєвому ланцюжку, закони жанру, тема висловлювання і його цільове призначення, смаки автора, його психофізичний стан у момент мовлення та ін. Дія всіх цих факторів настільки переплітається, що виділити і визначити результати впливу на мовлення кожного з них практично неможливо. Але якщо сукупність діючих факторів залишається незмінною, то результати будуть приблизно однаковими, тобто будова мовлення буде характеризуватися певними рисами, розкрити які допоможе статистика.

Нарешті, на будову і мови, і мовлення впливає будова людського мозку: об'єм пам'яті, швидкість запам'ятовування і відтворення того, що запам'яталося, характер психофізіологічних реакцій людського організму та ін. Усі ці фактори діють сукупно, власне, тому і роботу мозку і його характеристики також вивчають статистично: вони неоднакові у різних людей, але є певна середина, навколо якої розподіляються відхилення у різних людей. Ця середина називається нормою, а вона встановлюється тільки статистично. Мова розвивається і функціонує у суспільстві, на неї впливає маса індивідумів, а також маса так званих екстралінгвістичних факторів: різні групи і класи з їх установками відносно мови як суспільного явища і суспільства в цілому, історичні умови існування народ-носія даної мови, його контакти з іншими народами, ступінь розвитку суспільного життя, культури, літератури та ін. Результати впливу цих факторів на розвиток мови не можна завбачити не тільки тому, що їх багато і вони різноманітні, але й тому, що взаємодія їх зі структурними особливостями мови надзвичайно складна і ще не пізнана. Всі спроби окремих людей або груп чи класів впливати на мову за своїм бажанням і розсудом у більшості випадків кінчалися невдачею: мова сприймає невелику кількість нав'язуваних їй правил і одиниць, очевидно, тільки ті, які узгоджуються з її структурою. Тому взаємодія мови і суспільства

підкоряється статистичному закону (або законам) і вивчати її треба статистичними методами.

Існують такі закони будови й функціонування мови, які можна пізнати тільки з допомогою статистики [Перебийніс, 2003]. Одним з них є закон, відкритий Дж. Дьюї на початку ХХ ст. і названий нами законом переваги. Він полягає в тому, що і мова, і мовлення віддають перевагу невеликій кількості одиниць, які часто використовуються і становлять ядро будь-якої мовної чи мовленнєвої підсистеми, тоді як переважна кількість одиниць є низькочастотними. Так із 516 різних двофонемних інтервокальних сполук українських приголосних усього 5 (1 % інвентаря) становлять 16,7 % вживань цих сполук у словнику, а всього 51 найуживаніше українське слово становить 33 % у 500-тисячному масиві художніх прозових текстів. [ЧС, 1981, Т. 1, С. 16]

Таким чином, і будова мови, і її функціонування в мовленні, і співвідношення мови й мислення, мови й суспільства підкоряються дії статистичних законів, тому вивчати їх необхідно статистичними методами.

Робота багатьох колективів лінгвістів над укладанням частотних словників (ЧС) базується на визнанні факту підпорядкованості мови й мовлення статистичним законам. Статистичне обстеження великого обсягу матеріалу має своєю метою встановлення деяких характеристик мови (визначення валентності, тобто кола сполучуваності кожної одиниці) і мовлення (ланцюжка зв'язків, реалізовані в мовленні, питома вага кожного зв'язку і ланцюжки зв'язків у текстах різних функціональних стилів). Робота над ЧС потребує опрацювання матеріалу, яке передбачає володіння статистичними методами. Мета даного посібника полягає в тому, щоб надати мовознавцям мінімум знань про математичну статистику, необхідний для використання їх як при укладанні ЧС, так і при інших дослідженнях, завдання яких включає пізнання статистичних закономірностей будови і функціонування мови.

Оскільки більшість мовознавців не мають спеціальної математичної підготовки, ми намагалися викласти відомості зі статистики якомога простіше й доступніше, особливу увагу звертаючи на техніку статистичних підрахунків. Можливо, математикам деякі міркування в даному посібнику здаватимуться недостатньо точними та надмірно деталізованими. Ми схильні погодитися з ними, але, однак, такий виклад матеріалу та деяке його спрощення диктується необхідністю: треба роз'яснити математично не підготовленому читачеві основні положення статистики і методи та прийоми статистичного опрацювання матеріалу, причому викласти їх так, щоб читач не тільки зрозумів їх, але й зробив їх своїм інструментом аналізу. Тому в посібнику надзвичайно детально описуються етапи організації статистичного експерименту, прийоми підрахунків та ін., що математиків давно знайоме і не потребує спеціального роз'яснення.

У кожному розділі посібника даються приклади, як у тексті розділу, так і в кінці його (разом з питаннями для самоперевірки). Перша задача пояснюється дуже детально, наступна — побіжно, в ній звертається увага тільки на моменти, нові по відношенню до попередньої задачі. Розв'язання задач, що наводяться в кінці розділу, подається в додатку «Відповіді й роз'яснення». Тут для найпростіших задач, що є просто повторенням тих, які були роз'яснені в тексті, даються відповіді, а складніші одержують або виклад ходу обчислень (у вигляді таблиць чи роз'яснень), або пояснення логічного ходу думок при розв'язанні задачі.

Рекомендується не просто читати посібник, а опрацювати його, повторюючи всі обчислення слідом за викладом у книзі або ж виконуючи їх самостійно з наступною перевіркою за книгою.

В додатках до посібника наведено необхідні статистичні таблиці, методи роботи з якими пояснюються у відповідних розділах посібника, а при таблиці наявна вказівка на параграфи й розділи, в яких містяться необхідні роз'яснення.

У список літератури входять не лише посібники зі статистики, а й мовознавчі праці, де описуються статистичні дослідження, які можуть слугувати взірцями використання статистичних методів, описуваних у нашому посібнику.

Автор буде вдячний за конструктивну критику, яка дозволила б у подальшому зробити посібник доступнішим для використання широкими колами мовознавців.

Література

Буняковский В. Я. О возможности введения определительных мер доверия к результатам некоторых наук наблюдательных и преимущественно статистики / В. Я. Буняковский // Современник. — 1847. — Т. 3, раздел 2. — С 48.

Иванов В. В. Некоторые проблемы современной лингвистики / В. В. Иванов // Народы Азии и Африки. — 1963. — № 4.

Перебийніс В. С. Кількісні та якісні характеристики системи фонем сучасної української літературної мови / В. С. Перебийніс. — К. : Наукова думка, 1970.

Перебийніс В. І. Що дає статистика мовознавцям? / В. І. Перебийніс // Вісник Київського Лінгвістичного Університету. — 2003. — Т. 6, № 5. — С. 27–32. — (Серія: Філологія).

Частотний словник сучасної української художньої прози : у 2 т. — К. : Наукова думка, 1981.

Розділ 1

Передумови правильної організації статистичного дослідження

Розпочинаючи статистичне дослідження матеріалу, мовознавець повинен чітко уявляти відповіді принаймні на три основних питання:

- 1) Що рахувати?
- 2) Для чого рахувати?
- 3) Як рахувати?

Коротко зупинімося на проблемах, пов'язаних з відповіддю на кожне з цих питань.

§ 1. Що рахувати?

Відповідь на це питання передбачає визначення тих мовних одиниць, форм і категорій, які будуть статистично обстежуватися в експерименті.

Оскільки всі рівні мовної системи підвладні дії статистичних законів, підрахункам можуть піддаватися одиниці будь-якого рівня — фонемі або звуки, літери, сполуки звуків, фонем або літер, склади, морфемі, слова, словосполучення, синтаксичні конструкції тощо.

Першою і, мабуть, єдиною вимогою до вибору одиниць підрахунків є вимога їх точного, по можливості формального, структурного визначення, що спирається на ознаки і критерії, які максимально звільняли б визначення від суб'єктивізму та неперевірюваності.

Неважко довести, що недотримання цієї вимоги веде до одержання різних результатів на одному й тому ж матеріалі, до непорівнюваності результатів, одержаних різними дослідниками, і взагалі до дискредитації статистичних методів.

Уявімо, що дослідник вирішив визначити частоту слів на означення, наприклад, поняття часу в певній сукупності текстів. Якщо він заздалегідь не визначить точно коло лексики, яку він співвіднесе з цим поняттям, підрахунки його втратять будь-який смисл: в одних текстах він буде включати в це поняття одні слова, в іншому — ці плюс ще якісь. Адже не секрет, що чим більше уваги приділяєш якомусь питанню, тим більше одиниць у тексті зв'язуєш з ним, тим більше сумнівів виникає відносно характеру та кола зв'язаних з ним одиниць. Особливо справедливе наше твердження відносно семантики.

Тому будь-який підрахунок семантичних груп повинен базуватися на заданому списку слів, що включаються в дану групу або ряд. Навіть

якщо список цей неповний, результати дослідження будуть мати цінність і об'єктивність, тому що вони можуть бути перевірені: будь-який інший дослідник одержить такі ж самі результати на даному матеріалі, зможе перевірити і себе, і попереднього дослідника.

Якщо ж рахунку підлягають одиниці граматичні, необхідно оговорити, яке саме розуміння автор статистичного дослідження вкладає в їх визначення. Відомо, що існує багато різних визначень і розумінь кожної мовної одиниці, від фонемі — до речення. Якщо один дослідник під фонемою розуміє, скажімо, тип звука, для нього російські [а], [л], [ъ] будуть трьома фонемами, оскільки і в артикуляційному, і в акустичному відношенні це три типи звуків. Інший мовознавець може вважати фонемою реляційну одиницю, що виконує певні функції і протиставляється іншим одиницям за диференційними ознаками. При такому розумінні фонемі три вказаних типи звука складають варіанти однієї фонемі [а].

Припустімо, кожний з них вирішив в'яснити частоту фонемі [а] в певному тексті. Ясно, що у першого дослідника частота [а] буде істотно нижча, ніж у другого. Якщо ні один, ні другий (або хоча б один з них) не визначив свого розуміння терміна «фонема», результати їх підрахунків не можна зіставляти, оскільки вони спричиняють неправильні висновки.

Таким чином, визначення одиниць та підрахунку може здійснюватися трьома способами:

а) пояснення розуміння даного терміна або категорії; такого пояснення буває достатньо, якщо методика виділення даних одиниць добре розроблена та широко використовується й однієї вказівки на розуміння терміна достатньо для співвіднесення його з відповідною методикою встановлення досліджуваних одиниць;

б) вказівка на критерії виділення даних одиниць, які були б досить процедурними і такими, що перевіряються, тобто забезпечували б одержання то- жоних інвентарів одиниць різними дослідниками;

в) задавання одиниць рахунку списком, якщо для їх виділення не розроблена досить точна методика, а сам дослідник не може знайти точних критеріїв їх виділення.

§ 2. Для чого рахувати?

Відповідь на це запитання передбачає визначення мети і задач, які стоять перед статистичним дослідженням.

Найтиповіша постановка задачі в статистиці — це визначення істотності чи неістотності розходжень між двома вибірками.

Формулюючи задачу в термінах математичної статистики, дві сукупності перевіряємо на нульову гіпотезу. Вони відповідають нульовій гіпотезі, якщо між ними немає істотних розходжень у частоті досліджуваних одиниць. Тоді вважається, що ці дві сукупності належать до однієї генеральної сукупності (поняття генеральної сукупності буде наведено дещо пізніше). У протилежному разі, тобто якщо між ними спостерігається статистично істотне розходження в частоті певних одиниць, нульова гіпотеза відкидається і робиться висновок про те, що вони належать до різних генеральних сукупностей.

При використанні статистичних методів у лінгвістиці перевірка на нульову гіпотезу здійснюється кожного разу, коли метою дослідження є встановлення статистичних закономірностей функціонування мовних одиниць у творах різних авторів, у даному жанрі або функціональному стилі, в даний період розвитку мови. При цьому завжди відбувається зіставлення статистичних даних і характеристик або двох (кількох) авторів, або двох жанрів, стилів, періодів та ін. Зараз виконано дуже багато подібних досліджень. [Див. Ермоленко, 1970]

Перевірка на нульову гіпотезу здійснюється і при дослідженні статистичних характеристик мови, а не мовлення: статистичні характеристики використання морфем для побудови слів, фонем — для побудови морфем та ін. При цьому порівнюються можуть або різні періоди розвитку однієї мови, або різні мови в один і той же період.

Нарешті, ця ж гіпотеза перевіряється й у психолінгвістичних дослідженнях, при вивченні одноманітності сприйняття тих чи інших одиниць мови чи відрізків мовлення різними носіями мови або групами людей.

Особливою сферою застосування статистичних методів є порівняльно-історичне мовознавство: статистичне дослідження питомої ваги у мовній системі тієї чи іншої фонетичної зміни або фонетичного закону (а також і інших явищ та категорій) додало б більшої доказовості та питомої ваги висновкам у цій галузі науки про мову.

При укладанні частотних словників мається на меті як встановлення статистичних характеристик мови (коло сполучуваності одиниць нижчого рівня при утворенні одиниць вищого рівня), так і знаходження статистичних характеристик мовлення в цілому та його функціонально-стилістичних і авторських різновидів. Таким чином, у цьому випадку нульова гіпотеза перевіряється при зіставленні статистичних характеристик одиниць мови між собою, а також при зіставленні статистичних характеристик однієї і тієї ж одиниці (або групи одиниць) у різновидах мовлення. Існує інша відповідь на запитання «для чого рахувати». Мається на увазі теоретичне і практичне значення результатів статистичних досліджень для мовознавства.

Теоретичне значення застосування статистичних методів та їх результатів для мовознавства важко переоцінити. Статистичні методи не тільки додають більшої ваги і авторитетності, доказовості мовознавчим висновкам, вони здатні розкрити такі закономірності будови мови та мовлення, які без них розкрити неможливо, перевірити і часто відкинути такі загальноприйняті твердження, існування яких можливе було лише внаслідок недостатнього проникнення в глибину структури мови.

Як приклад пошлемося не тільки на наведені вище висловлювання В. В. Іванова та В. Я. Буняковського, але і на праці В. А. Богородицького [Богородицкий, 1913, С. 39–40], А. М. Пешковського [Пешковский, 1925], М. М. Петерсона [Петерсон, 1951, С. 23–28], В. Ф. Чистякова і Б. К. Крамаренка [Чистяков, 1929], а також сучасних лінгвістів (див. список літератури).

У плані прикладному статистичні методи мають велике значення для відбору учбового матеріалу і методики викладання іноземної та рідної мови, раціональної побудови стенографії, встановлення авторства непідписаних рукописів і т.д., і т.п. Особливе значення мають ці методи для вирішення цілої низки проблем, так чи інакше пов'язаних з автоматичним опрацюванням лінгвістичних даних за допомогою ЕОМ [Использование ЭВМ..., 1990; Морфологический анализ научного текста ЭВМ, 1989].

§ 3. Як рахувати?

Відповідь на це запитання передбачає розгляд та роз'яснення основних положень математичної статистики, що стосуються правил організації вибірки, визначення обсягу досліджуваного матеріалу для одержання статистично вірогідних результатів, обчислення статистичних характеристик, зіставлення статистичних параметрів різних сукупностей, статистичних методів визначення зв'язку досліджуваних одиниць у сукупності і т.п.

Щоб навчитися рахувати, треба також оволодіти технікою статистичного опрацювання матеріалу, мати уяву про основні поняття і терміни математичної статистики.

Власне кажучи, відповіддю на це запитання повинен бути цей посібник. Але слід мати на увазі, що в ньому викладаються найнеобхідніші, початкові дані з математичної статистики, які далеко не вичерпують ані методів статистичного дослідження, ані можливих підходів до техніки обчислень.

Отже, переходимо до роз'яснення основних понять математичної статистики і правил організації статистичного дослідження, знання яких конче потрібне для успішного застосування статистичних методів.