

УДК 371.81'33

ИСПОЛЬЗОВАНИЕ МЕТОДОВ СТАТИСТИКИ ДЛЯ ЭФФЕКТИВНОГО ИЗУЧЕНИЯ ЯЗЫКА ПЕДАГОГИКИ И ОБРАЗОВАНИЯ

Ключевые слова: язык профессиональной педагогической коммуникации, оптимизация научно-педагогического исследования, междисциплинарный подход.

Андриенко Е.А.

*ассистент кафедры английского языка,
аспирант кафедры английского языка
Педагогического института
Южного федерального университета*

Как отмечал В.А. Звегинцев, язык не существует вне тех, кто думает и говорит на нем. Он отмечал, что язык коренится в глубинах индивидуального сознания: оттуда он берет свою силу, чтобы воплотиться в звуках человеческой речи. По его мнению, индивидуальное сознание представляет собой только один из элементов коллективного сознания, диктующего свои законы индивидууму, а развитие языков представляет собой только один из видов развития общества. В этой связи им делается общий вывод о методах изучения языка, который сводится к тому, что, только изучая социальную роль языка, можно составить себе представление о том, что такое язык [2].

Несомненно, язык, в самом широком его понимании, представляет собой универсальное средство общения, обслуживающее различные сферы деятельности человека. В более узком значении язык определяется как совокупность категорий и понятий какой-либо отдельно взятой науки или научного направления, например – язык педагогики и образования. С помощью языка раскрывается предмет науки, описывается та часть объективной реальности, которая изучается данной наукой. Причем эмпирические и теоретические знания, полученные в результате изучения данного отрезка объективной реальности, каждая наука излагает на своем собственном языке.

В этой связи незнание языка приводит к многочисленным трудностям в понимании и осмыслении предмета изложения, ракурсов и подходов к рассматриваемой проблеме. Так, например, помимо общих категорий и понятий, являющихся универсальными для всех отраслей педагогики и составляющих лишь малую часть

ее языка, любая педагогическая теория излагается на своем собственном языке, который имеет свою функциональную специфику. В частности, язык педагогики включает в себя научный, профессиональный, обыденный языки; он развивается под влиянием других отраслей знания, усиливающегося межнаучного и межкультурного диалога в едином поликультурном образовательном пространстве. Все это затрудняет профессиональную педагогическую коммуникацию и актуализирует исследование ее языка.

В течение последнего столетия языкознание как наука о языке всегда приводилось как пример науки, развивавшейся стремительно и очень быстро достигшей методической зрелости. Этот процесс происходит наряду с интенсивным развитием так называемых «точных» наук, таких как математика, статистика, методы которых стремятся проникнуть практически в каждую сферу человеческого познания. В связи с этим возникает вопрос о возможной целесообразности применения методов математики и статистики для эффективного изучения языка вообще и языка педагогики и образования в частности.

Для того чтобы ответить на данный вопрос, в первую очередь, необходимо определить те области лингвистических исследований, в которых данные методы находят свое наиболее эффективное применение. Важно и выявление как той меры, в какой они согласуются со спецификой языкового материала, так и того потенциала, который позволяет подсказать ответы на вопросы, которые ставит перед собой наука о языке.

Приступая к рассмотрению возможностей применения методов статистики для изучения явлений языка, необходимо, пожалуй, пояснить, что же

понимается под методом. Так, В.А. Звегинцев определяет метод в языкознании как некоторую совокупность исследовательских приемов, объединенных каким-либо общим принципом и применяемых в лингвистическом исследовании для выполнения частных исследовательских задач [2].

Необходимо отметить, что применение математического (или, точнее, статистического) критерия для решения языковых исследовательских задач является отнюдь не новым для науки о языке и в той или иной мере уже давно используется учеными как в лингвистических, так и в междисциплинарных изысканиях. Так, например, продуктивность грамматических элементов или даже критерии родственных отношений между языками в известной степени строятся на относительных статистических признаках. Но если первоначально применение статистических методов в лингвистических исследованиях носило скорее стихийный характер, то в последнее время они стали применяться для решения конкретных исследовательских задач. Так, в последнее время довольно активно развиваются исследования по созданию частотных словарей и справочников, фиксирующих количественные признаки лексических компонентов языка и его отдельных подязыков (см., например, частотные словари на информационном ресурсе <http://homepages.tversu.ru>).

Как отмечает В.А. Звегинцев, в последние годы случаи применения математических методов к языковому материалу значительно умножились и стали складываться более или менее определенные направления: стилостатистика, лингвостатистика. В рамках стилостатистики осуществляется определение и дается характеристика стилистических особенностей отде-

льных произведений или авторов посредством количественных отношений используемых языковых элементов. В основе статистического подхода к исследованию стилистических явлений лежит понимание литературного стиля как индивидуального способа владения средствами языка. Недостаток данного рода исследований заключается в том, что все внимание сосредотачивается на количественной стороне, в то время как качество языковых элементов, их смысловая и эмоционально-экспрессивная сторона остаются без внимания. Таким образом, художественное произведение выступает в виде механической совокупности, специфика построения которой находит свое выражение лишь через числовые отношения ее элементов. Тем не менее представители данного направления исследований, не отвергая методы, традиционной стилистики, сознательно применяют математические методы, аргументируя свой выбор объективностью полученных таким образом данных, что, по мнению представителей стилостатистики, окупает все недостатки математических методов.

Несколько по-иному строятся подсчеты относительной частоты употребления слов. В этом случае учитывается не только общая сумма слов, но и частота употребления отдельных слов. Углубляя вышеописанный метод, стилостатистика в качестве стилиевой характеристики стала прибегать к критерию стабильности относительной частоты наиболее употребительных слов. На основе проведенных исследований было высказано предположение, что то, что наблюдается в отношении некоторых (наиболее употребительных) слов, может быть применено и по отношению ко всем использованным в произведении сло-

вам. Отсюда следует, что стиль автора можно характеризовать определенным соотношением изменчивости средней частоты употребления слова к общей для данного языка частоте его употребления. Это соотношение и рассматривается в качестве объективной количественной характеристики стиля автора. Аналогичным образом исследуются и иные формальные элементы структуры языка.

Таким образом, можно отметить, что методы стилостатистики, безусловно, развиваются. Однако они дают лишь беспристрастные, поверхностные, сугубо внешние «объективные» результаты.

Следующим направлением применения математических методов для изучения языковых явлений является лингвостатистика. В Большой Советской энциклопедии лингвостатистика (статистическая лингвистика) определяется как дисциплина, изучающая количественные закономерности естественного языка, проявляющиеся в текстах. Согласно определению, представленному в энциклопедии «Русский язык», лингвостатистика в широком смысле представляет собой область применения статистических методов в языкознании. Она опирается на математическую статистику подсчетов и измерений при изучении языка и речи. В узком смысле данное направление лингвистических исследований представляет собой изучение некоторых математических проблем, связанных с лингвистическим материалом, главным образом с типами статистических распределений языковых единиц в тексте [5].

Таким образом, лингвистическая статистика – это раздел языкознания, занимающийся статистическими методами изучения количественных закономерностей в языке и речи. Пер-

вичным материалом в лингвистической статистике является текст, рассматриваемый как последовательность лингвистических единиц заданного уровня: букв или фонем, морфов или морфем, словоформ или лексем, словосочетаний, предложений. На этом материале изучаются количественные характеристики лингвистических форм – их употребительность, совместная встречаемость, законы распределения в тексте, их физические размеры. На основе полученных данных описываются свойства текста, формулируются гипотезы о механизмах его образования и об устройстве системы языка. Основные понятия и категории в лингвистической статистике заимствуются у математической статистики. Такими понятиями являются понятия генеральной совокупности и выборки, частоты и вероятности, вероятностные распределения и статистические оценки. Однако применение этих понятий к лингвистическому материалу имеет ряд особенностей. В частности, в языкознании могут быть рассмотрены два принципиально разных вида генеральной совокупности: либо совокупность текстов (корпус текстов) одинакового жанра, заданного списка авторов или заданного временного интервала, либо совокупность единиц, принадлежащих одному лингвистическому уровню: фонем, морфем, слов или предложений.

Лингвистическая статистика как научное направление возникла в связи со стремлением дополнить совокупность структурных характеристик лингвистических единиц характеристикой их употребительности. Это основано на предположении, что любая лингвистическая единица обладает априорно присущей ей вероятностью быть употребленной в тексте заданного класса. Таким образом, лингвостатистика

получает возможность оперировать не только количественными, но и качественными признаками. Например, в ходе статистико-морфологического исследования словоупотребления текста группируются по признаку их принадлежности к той или иной части речи. Основные понятия лингвистической статистики и методические указания по применению математических методов можно найти у К.Б. Бектаева и Р.Г. Пиотровского [4].

В основе лингвостатистики лежит предположение, что некоторые численные характеристики и функциональные зависимости между ними, полученные для ограниченной совокупности текстов, характеризуют язык в целом или его функциональные стили (публицистический, научный, научно-популярный).

Так, например, функциональный стиль языка педагогической коммуникации характеризуется высокой частотностью употребления таких глаголов, как *base, become, ask, believe, can, teach*.

Например: *Teachers would **begin** by placing example items in the categories, and then **ask** questions such as: What belongs together? **Can** you find something that is similar to this? Other ways to explore our **teaching** include self-observation, observation of other teachers, and talking about **teaching** we observe. In the pages that follow I elaborate on these three ways to explore **teaching** because I have found them to be quite conducive to seeing our **teaching** differently. In both ProW and CTBI, part of the rationale for their adoption is **based** on the result of research into the processes rather than the products of language use and language learning, usually with reference to more proficient target language users.*

В связи с рассмотрением функционального стиля научно-педагогичес-

кого общения стоит отметить тот факт, что в функциональной стилистике стиль рассматривается как содержательная форма, причем содержание превалирует над стилистической формой его представления в речи. Таким образом, формирование того или иного функционального стиля в языке связано с тем, что общество коллективно осознает специфику содержательной стороны каждой сферы общения, связанной с определенным видом деятельности. Более того, оно коллективно разрабатывает в практике особые комплексы форм и функций, служащих для более полного выражения актуального в данной сфере содержания.

До недавнего времени изучение и разграничение функциональных стилей проходило преимущественно с опорой на анализ их качественной стороны. Такой подход представляется вполне допустимым, учитывая тот факт, что проведение достоверных количественных исследований было затруднено в связи с невозможностью обработки больших текстовых массивов и проведения широкомасштабных лингвистических исследований.

Опора на лингвостатистические методики, активно развивающиеся сегодня на основе информационных и компьютерных технологий, и фактические данные этой науки позволяет сегодня исследователю изучать речевой стиль в единстве его качественной и количественной сторон. Она рассматривает стиль как систему статистического характера, где многообразно реализуется диалектический закон перехода количества в качество и различия между стилями проявляются не только в наличии разных качественных признаков, но и в различных количественных проявлениях качественно общих признаков. Тем самым определяется возможность более глубокого и

диалектического понимания природы стиля.

Отметим, что наиболее изученной в прикладном и теоретическом плане числовой характеристикой стиля является относительная частота употребления. Данная характеристика, как правило, изучается на материале различных лингвистических единиц (букв, фонем, слогов, слов, синтаксических конструкций), их классов (гласных, согласных, частей речи) и сочетаний (например, последовательностей из букв). При этом основной акцент делается на исследовании их функциональной зависимости, приближенно описывающей связь между частотой слова и его номером (рангом) в последовательности по убыванию частот (закон Ципфа–Мандельброта). Статистическая лингвистика изучает также зависимости между частотой и длиной слова (в числе слогов), числом его значений и возрастом. Накопленные данные используются для выявления особенностей стиля отдельных авторов, функциональных стилей, жанров и подъязыков.

Успех каждого лингвистического исследования зависит от организации статистического наблюдения, которая предусматривает выбор лингвистического признака и установление единицы совокупности, с одной стороны, и определение способа наблюдения – с другой; причем каждый количественный или качественный признак, применяемый для выделения единицы совокупности, должен иметь лингвистический смысл и отвечать задачам данного языковедческого исследования. Однако каким бы ни было основание для группировки – естественным или искусственным, определенным или неопределенным, конечное решение должно быть всегда строго фиксированным. Каждый лингвистический

объект должен быть признан либо обладающим, либо не обладающим данным качественным признаком [4].

Статистическое наблюдение предусматривает как сплошное (если совокупность обозрима и необходимо учесть все употребления интересующих нас языковых объектов), так и выборочное (статистическое описание языка писателя, отдельного художественного произведения) обследование генеральной совокупности. В большинстве случаев генеральная совокупность настолько велика, что применить сплошное обследование не представляется возможным даже при условии применения вычислительной техники. Поэтому в таком случае подвергается исследованию лишь часть единиц генеральной совокупности. Такое наблюдение может быть осуществлено с помощью либо повторной, либо бесповторной выборки. И в том и в другом случае осуществляется перенос результатов наблюдения над частотной выборкой на всю генеральную совокупность, что может быть осуществлено лишь в том случае, если средняя величина признака и его относительная частота в выборочном наблюдении достаточно хорошо воспроизводят среднюю величину и относительную частоту признака в генеральной совокупности.

Статистика предполагает следующие приемы выборочного наблюдения: 1) случайный отбор, при котором выбор отдельных единиц осуществляется либо по жребию, либо путем использования таблиц случайных чисел. При этом каждая единица совокупности имеет равную возможность попасть в выборку, что обеспечивает достаточно близость средней выборочной величины к средней генеральной величине. Однако ввиду его громоздкости данный вид отбора довольно редко используется в лингвистике;

2) механический отбор, при котором единицы совокупности выбираются в определенном, формально установленном порядке; 3) серийный отбор, при котором в противоположность рассмотренным выше видам выборки, где отбор каждой единицы проводится в индивидуальном порядке, предполагается отбор сериями. Эти серии отбираются в случайном порядке, чаще бесповторным способом. Отобрав таким образом серии, исследователь проводит внутри них сплошное наблюдение; 4) типический отбор, который предполагает предварительную разбивку генеральной совокупности по определенному признаку на однородные тематические группы, из которых затем случайным порядком выбираются интересующие нас лексические или грамматические единицы. При сопоставлении частотных словарей типическая выборка сочетается с серийным отбором. Количество серий, извлекаемых из каждой тематической группы, определяется удельным весом этой группы в генеральной совокупности [там же].

Как отмечается в работах В.И. Тузлюковой, мозаичность, многоплановость и сложность проблем педагогической коммуникации обуславливает необходимость их решения за счет привлечения новых перспективных идей и значительного научного потенциала лингвистики и смежных наук. Интенсифицируется необходимость проведения разноаспектных исследований языка профессиональной педагогической коммуникации как языка, репрезентирующего разнообразные профессиональные пространства в единстве их уникальных связей и отношений [6]. Такие исследования, по мнению ученого, являются значимыми в условиях современного развития общества и отражают происходящие сегодня

интенсивные процессы интеграции и дифференциации научных и профессиональных знаний, структурную сложность сфер профессиональной коммуникации и новейшие требования к используемым в их рамках языковым средствам, речевым моделям и формам. В частности, использование методов лингвостатистики позволит установить связи между наблюдаемыми характеристиками лексических компонентов языка современной педагогики и образования и осуществить выбор наиболее существенных из них.

Литература

1. Большая Советская энциклопедия. 3-е изд. М.: Изд-во «Советская энциклопедия», 1973.
2. *Звегинцев, В.А.* Очерки по общему языкознанию / В.А. Звегинцев [Электронный ресурс]. Режим доступа: <http://www.classes.ru>.
3. Методы анализа ЕЯ описания предметных областей [Электронный ресурс]. Режим доступа: <http://www.philippovich.ru>.
4. *Пиотровский, Р.Г.* Математическая лингвистика / Р.Г. Пиотровский, К.Б. Бектаев, А.А. Пиотровская. М., 1977.
5. Русский язык: Энциклопедия. М., 1997.
6. *Тузлукова, В.И.* Предисловие к сборнику «Язык профессиональной коммуникации: функции, среды, технологии» / В.И. Тузлукова // Язык профессиональной коммуникации: функции, среды, технологии: сб. докладов международной науч.-практ. конф. / под ред. В.Р. Саркисянца, В.И. Тузлуковой. Ростов н/Д, 2008.