

рипов, А. А. Дарбеева, М. Я. Лешика) на материале конкретных языков приводились доказательства в пользу существования залогов как самостоятельной грамматической категории. Вместе с тем обсуждение настоящей проблемы выявило и наличие разногласий, связанных с проблемой залогов (И. Е. Маманов, Т. З. Козырева, А. А. Юлдашев); например, по вопросу о взаимоотношении категории переходности и непереходности и категории залога; по вопросу об отнесении к залогу производных образований с залоговыми аффиксами, лишенными значения данного залога или обозначающими лишь один из его оттенков; по вопросу о принципах выделения, с одной стороны, страдательного и возвратного залогов, а с другой — поучительного залога.

Наиболее плодотворной оказалась дискуссия по вопросу о природе залоговых форм. В докладах и в выступлениях содержались новые наблюдения, которые будут учтены нашими лингвистами в их дальнейших исследованиях вопросов залога.

*Н. З. Гаджиева*

### СОВЕЩАНИЕ ПО СТАТИСТИКЕ РЕЧИ

С 1 по 4 октября 1957 г. в Ленинграде происходило совещание, посвященное статистике речи, созванное Секцией речи Комиссии по акустике АН СССР и Ленинградским университетом. В совещании приняли участие также представители МГУ и МГПИИЯ, НИИ Министерства радиотехнической промышленности, Институтов физиологии и языкознания АН СССР, Лаборатории электро моделирования ВИНТИИ АН СССР, Военно-воздушной академии им. А. Ф. Можайского и ряда других организаций.

Совещание открылось вступительным словом Л. Р. З и н д е р а. В центре внимания совещания было два основных круга вопросов: 1) применение статистических исследований устной и письменной речи к разработке проблем, выдвигаемых теми разделами современной техники, которые связаны так или иначе с хранением, обработкой и передачей информации; 2) соотношение структурных и статистических методов в языкознании.

Доклад «Значение статистических исследований речи для техники» сделал Л. А. В а р ш а в с к и й. Докладчик указал основные направления в исследовании устной речи, развитие которых существенно для техники телефонной (проводной и непереводной) связи. К таким направлениям относится прежде всего исследование (в том числе статистическое) физических характеристик звуковых и электрических сигналов, служащих для передачи речи. В этой связи представляется целесообразным развитие общей теории сигналов, как это подчеркнул в своем докладе «Энергетические характеристики интервалов корреляции электрических сигналов, в частности, речевых сигналов» Н. А. Ж е л е з н о в.

Другой круг проблем, указанный Л. А. Варшавским, касается восприятия устной речи, передаваемой по каналам связи. Качество канала измеряется его артикуляцией, под которой в технике связи понимают процент правильно воспринятых звуковых единиц, переданных по каналу. Для определения артикуляции используются специальные звукосочетания, которые и передаются по каналу. Уже при составлении таблиц таких звукосочетаний необходимо учитывать статистические соотношения. Но наибольший интерес представляет вопрос о связях между артикуляциями звуковых единиц разного порядка: отдельных звуков, слогов, морфем, слов и т. п. При этом решающую роль играет изучение влияния, которое оказывает на восприятие данного звука восприятие окружающих его звуков. Это влияние обусловлено вероятностной зависимостью (корреляцией), имеющей место между соседними звуками.

Изучению подобной корреляции между соседними элементами речи (т. е. следующими друг за другом звуками, словами и т. п.) был посвящен доклад Л. Р. З и н д е р а «О лингвистической вероятности». Как отмечалось в докладе, каждый элемент речи несет определенную информацию (в ряде случаев весьма большую) о непосредственно следующем за ним элементе. Лингвистические вероятности нулевого порядка (т. е. абсолютные вероятности появления в речи тех или иных элементов) не совпадают, как правило, с лингвистическими вероятностями первого порядка (т. е. условными вероятностями появления одних элементов после других). Лингвистические вероятности как нулевого, так и высших порядков подразделяются на лексические (вероятности появления тех или иных лексем), грамматические (вероятности появления тех или иных грамматических форм) и звуковые (вероятности появления тех или иных звуковых единиц). Лингвистические вероятности оказывают значительное влияние на восприятие и, в конечном счете, понимание устной речи (в большой степени это относится к грамматическим и в меньшей к звуковым вероятностям). Доклад Л. Р. Зиндера сопровождался демонстрацией трех таблиц звуковых вероятностей первого порядка для русского языка. Были показаны составленные на основе статистического обследования

текстов общим объемом около 90 000 фонем отдельные таблицы для сочетаний звуков внутри слова, на стыке слов и сводная таблица.

Доклад Е. В. Падучевой «Статистическое исследование структуры слога (в связи с применением методов теории информации)» был посвящен сравнительному анализу сочетаний фонем внутри слога и на стыке слов. По мнению докладчицы, лишь сочетания фонем внутри слога могут составить предмет фонологического изучения; сочетания на стыке слогов — скорее следствие словообразования. Естественно ожидать, что на стыке слогов будут ослабляться ограничения, накладываемые на сочетаемость фонем внутри слога. В то же время, как показала докладчица на примере испанского языка, простое деление сочетаний фонем на возможные и невозможные оказывается недостаточным; к более удовлетворительным результатам (для русского языка) приводят вероятностные соображения.

Из перечисленных в докладе Л. А. Варшавского проблем важнейшими являются, по-видимому, проблемы компрессии, связанные с наиболее эффективным использованием каналов связи. Одним из способов повышения пропускной способности канала служит уменьшение времени, затрачиваемого на передачу одного звука; встает вопрос, до каких пределов можно сокращать длительность звука. Как показал в своих докладах «Статистика длительности глухих согласных и их восприятие» и «Статистика характерных участков звучания гласных звуков русского языка» М. Ф. Деркач, акустический состав звуков неоднороден во времени. Поэтому, если из временного промежутка, в течение которого произносится некоторый звук, выделить меньший интервал и воспроизвести звук только в этом интервале, то восприятие может нарушиться. Участники совещания имели возможность прослушать магнитофонную запись слога *та*, искусственно полученного из слога *са* посредством «отрезания» начала звука *с* (в то же время, если в слоге *ас* отрезать конец звука *с*, то нарушения восприятия не произойдет).

Более радикальным путем повышения пропускной способности канала связи явилось бы выделение минимальной информации, необходимой для различения единиц устной речи (например, фонем) и передача по каналу только этой минимальной информации. (При этом становится очевидной ведущая роль исследований по выяснению семантической нагрузки речевых единиц, прежде всего фонологических исследований.) Имеется в виду помещение на передающем конце линии связи анализирующего устройства, способного различать фонемы и превращающего их в последовательности дискретных сигналов, а на приемном конце — синтезирующего устройства, способного превращать воспринимаемые сигналы в звуки. Принципиально более сложным является создание различающего устройства. Одним из возможных вариантов такого устройства явилось бы устройство, способное определять фонетическую принадлежность звука по его спектрограмме.

Новый метод, отличающийся от обычно применяемого формантного анализа, предложила Л. А. Чистович в докладе «Применение статистических методов к определению фонетической принадлежности индивидуального гласного звука». Л. А. Чистович отказалась от поисков в спектрограмме инвариантов, которые позволили бы точно определить фонетическую принадлежность звука; ее идея состоит в том, что ответ на вопрос о фонетической принадлежности звука может быть дан не в категорической форме, а с определенной вероятностью. Рассматриваются огибающие спектрограмм разных звуков, произнесенных разными лицами. Для каждого звука производится усреднение этих огибающих (разных у разных лиц); так получается основная кривая для данного звука. Эта основная кривая рассматривается как сигнал, а реальная огибающая (для конкретного произносителя) как сигнал, искаженный некоторым шумом, внесенным произносителем. Чтобы определить фонетическую принадлежность конкретного произнесенного звука, вычерчивается огибающая его спектрограммы и определяются отклонения ее от основных кривых. При этом вычисляются вероятности того, что данная конкретная огибающая получилась искажением той или иной основной кривой или, что то же самое, того, что данный конкретный звук есть звук *а*, *о*, *у* и т. д. Метод Л. А. Чистович может быть без изменения применен и к определению фонетической принадлежности звуков связной речи с учетом влияния на их восприятие восприятия соседних звуков.

В докладе «Применение статистических методов в экспериментально-фонетическом и психологическом изучении речи» В. А. Артемов рассказал о работах, проводимых в Лаборатории экспериментальной фонетики МГПИИЯ.

Для техники телеграфной связи (проводной и непроводной) большое значение имеет статистическое исследование письменной речи, в частности вычисление энтропии распределения сочетаний букв. Как сообщили в своем докладе «Статистика трехбуквенных сочетаний русского печатного текста» В. А. Гармаш и Д. С. Лебедев, в Лаборатории по разработке научных проблем проводной связи АН СССР был произведен подсчет частот трехбуквенных сочетаний, встречающихся в отрывке из романа Л. Н. Толстого «Война и мир» объемом в 30 000 букв (пропуск между словами также считается «буквой» и обозначается дефисом «-»). В докладе были приведены частоты некоторых сочетаний, причем наиболее частыми оказались сочетания «и-»

(частота  $82 \cdot 10^{-4}$ ) и «-не» (частота  $71 \cdot 10^{-4}$ ). Вычисление соответствующей энтропии показало, что подходящим кодированием трехбуквенных сочетаний объем текста может быть сокращен в среднем в  $\frac{5}{3}$  раза.

Роль статистических методов в языковедении не определяется исключительно потребностями технических приложений. Доклад «Соотношение структурных и статистических методов в языковедении» сделал И. И. Ревзин, подчеркнувший, что было бы большой ошибкой недооценивать значение вероятностных и статистических соображений для развития самой лингвистической теории (хотя бы потому, что структура языка, по мнению докладчика, в значительной степени обусловлена избыточностью языка как кода; эта избыточность вызвана в свою очередь тем, что язык по необходимости должен быть помехоустойчивым). Как указал И. И. Ревзин, обычные лингвистические методы должны дополняться статистическими.

В докладе Ю. К. Лекомцева «Порядковые и функциональные отношения вьетнамского глагола» было показано, что во вьетнамском языке после выделения некоторых полнзначных глаголов служебные глаголы могут быть опознаны на основе чисто статистических подсчетов их сочетаемости с полнзначными.

Статистические методы с успехом применимы и для получения новых лингвистических результатов. Так, в докладе И. А. Мельчук «Применение статистики к вопросу о категории рода во французском и испанском языках» было доказано, что во французском языке грамматическая категория рода формально выражена окончанием существительного. При этом автор доклада выдвинул следующий статистический критерий выраженности категории рода: категория рода в данном языке считается выраженной, если существуют правила не слишком большого объема (не превосходящие, скажем, объема аналогичных правил для испанского языка, выраженность рода в котором не подвергается обычно сомнению), позволяющие узнавать род по окончанию существительного и охватывающие не менее 94% существительных рассматриваемого языка (соответствующие правила для испанского охватывают свыше 98%).

Как отмечалось на совещании, статистика может быть использована и в текстологии, для определения автора того или иного текста.

В докладе «Вероятностное определение лингвистического времени (в связи с проблемой применения статистических методов в сравнительно-историческом языковедении)» Вяч. Вс. Иванов продемонстрировал возможность применения статистических методов к внутренней реконструкции. Анализируя состояние языка в данный момент, можно установить направление лингвистического времени. Так, часто встречающиеся в текстах языковые единицы имеют тенденцию встречаться в последующем состоянии языка, если они обладают малой степенью изолированности в системе; если же эти часто встречающиеся единицы имеют высокую степень изолированности, то они характерны для предыдущих состояний. Частота встречаемости и степень изолированности определяются статистически. Другой способ реконструкции и предсказания связан со статистическим исследованием стилей языка. Родство языков наиболее достоверно определяется в терминах изоморфизма реконструируемых систем. Что же касается часто применяемых для установления степени родства языков подсчетов совпадений, то, как указал докладчик, здесь необходим тщательный вероятностный анализ (так как вероятность случайных совпадений может быть велика и, таким образом, наличие совпадений ни о чем не говорит).

И. И. Ревзин подчеркнул в своем докладе необходимость развития специальной отрасли языковедения — лингвистической статистики. Уже сейчас полученные в лингвистической статистике соотношения между такими величинами, как частота слова, его ранг (т. е. номер по порядку в частотном словаре), его длина и т. п., проливают свет на теоретико-информационную природу языка как кода. Об этом же говорил в докладе «Некоторые вопросы статистического обследования лексических групп» Р. Г. Пиотровский.

В докладе И. И. Ревзина был выяснен также двусторонний характер соотношения структурных и статистических методов. Не только статистика помогает лучше разобраться в структуре языка, но, в свою очередь, единицы, число которых подсчитывается, нуждаются в точном структурном определении. Так, по мнению автора доклада, недостатком глоттохронологии является отсутствие точного определения родственных слов. (В. В. Иванов в своем докладе указал на недостаточную мотивированность самого выбора базисного словаря.) Ясно, что статистическое исследование структуры слога невозможно без строгого определения понятия «слог» (в первой части доклада Е. В. Падучевой было дано такое определение испанского слога, которое, по-видимому, позволяет однозначно разбивать слово на слоги). Роль структурных категорий обнаруживается и при проведении статистических работ, имеющих непосредственное практическое приложение. Для создания оптимальных правил машинного перевода необходимо статистическое обследование языков отдельных наук. В докладе «О статистическом словаре русских математических текстов» И. А. Мельчук, Т. Н. Мошнина, А. Л. Шумилкина, З. М. Волочка и И. Н. Шелмова в сообщении о результатах статистического обследования языка математической литературы;

при этом возникла необходимость четкого определения таких понятий, как «синтагма», «тип синтагмы», «связь слов в предложении» и т. п.

Совещание в Ленинграде, бесспорно, имело принципиальное значение, не ограниченное кругом вопросов, указанных в его названии. На совещании отчетливо выявились два обстоятельства:

1. Проникновение математических, в частности статистических, методов в языковедение, несомненно, плодотворно. Эти методы могут играть очень важную, но все же подчиненную роль при решении лингвистических проблем. Полностью формализовать реальный язык в виде некоей математической системы, по-видимому, никогда не удастся, однако можно ставить вопрос о тех или иных формализованных приближениях к реальному языку, причем расхождение между реальным языком и таким приближением должно оцениваться статистически.

2. Лингвистические исследования начинают приобретать все большее и большее практическое значение, не укладывающееся, как раньше, в рамки составления школьных грамматик и орфографических правил. Это не означает, что языковедение утрачивает свой теоретический профиль. Наоборот, с развитием техники оказывается, что наиболее тонкие теоретические построения наиболее важны для приложений. Положение дел в лингвистике можно сравнить в этом отношении с положением в математике, теоретические отрасли которой (такие, как математическая логика) приобрели в последнее время особое, прикладное значение.

Большим достоинством совещания явилось разнообразие представленных на нем специальностей, от радиотехники до физиологии. Совещание показало необходимость и дальнейшей координации деятельности представителей разных наук в области прикладной лингвистики.

В. А. Успенский

## О ДИАЛЕКТОЛОГИЧЕСКОМ АТЛАСЕ БОЛГАРСКОГО ЯЗЫКА

Летом 1956 г. Институт болгарского языка Болгарской Академии наук (БАН) совместно с Институтом славяноведения АН СССР начал работу по составлению атласа болгарских говоров, распространенных на территории Болгарии. (Атлас болгарских говоров на территории СССР уже завершен и находится в печати.)

Предварительно проф. С. Стойковым была составлена «Программа для сбора материала по болгарскому диалектологическому атласу», содержащая 379 вопросов (122 — по фонетике, 141 — по морфологии, 6 — по синтаксису и 110 — по лексике). Эта программа была подробно обсуждена осенью 1955 г. в Софии и весной 1956 г. в Москве.

Первая экспедиция под руководством проф. С. Стойкова и проф. С. Б. Бернштейна работала в составе 28 человек в августе — сентябре 1956 г. Ею было обследовано 70 сел в Бургасской, Малкотырновской, Грудовской, Елховской, Поморийской и Поповградской околиях, — из них в 28 селах собран полный, а в 16 — частичный материал по «Программе» атласа.

В 1957 г. объем работы над атласом значительно расширился. В экспедиции, которая проводилась с 1 августа по 10 сентября, принял участие 51 человек: 9 сотрудников АН СССР, сотрудники БАН, преподаватели, аспиранты и студенты Софийского гос. университета. В ней участвовали также научные работники из ГДР, Румынии и Чехословакии. Было обследовано еще 93 села с исконным населением в Грудовской, Елховской, Тополовградской, Ямбольской, Харманлийской, Свиленградской, Хасковской и Ивайловградской околиях. Кроме того, были собраны сведения в ряде сел со смешанным населением.

В результате двух экспедиций был собран ценный материал, подтвердивший в основном известные в науке факты о юго-восточных болгарских говорах (см., например, I. j. Miletič, *Das Ostbulgarische*, Wien, 1903), но внесший в них известные коррективы и дополнения. В юго-западной части Елховской околии проходит западная граница так называемого грудовского, «загорского» о-говора, однако участники экспедиции обнаружили здесь ряд «рупешких» особенностей, характерных для самого восточного из исследованных диалектов — страуджанского. Большое разнообразие обнаружено в употреблении частицы будущего времени; все же достаточно четко вырисовываются три диалектные группы: страуджанская (*ше, шь*; членная форма муж. рода — *шт*), грудовская (*ке, ки, кь*; членная форма муж. рода — *о*) и более западная группа от Тополовградской околии до Родонов (*эь, жь, жи*; членная форма муж. рода — *ь*). Следует отметить, что о-говоры имели в прошлом глубокий тыл на юге, главным образом в Адрианопольской области. Ряд особенностей, зафиксированных в изученных говорах, свидетельствует о их связи с западноболгарскими и македонскими диалектами.