

Р. Г. Пиотровский. Информационные измерения языка. — Л., «Наука», 1968. 116 стр.

Со времени опубликования статьи К. Шеннона «Предсказание и энтропия печатного английского текста» (1951) появился ряд работ, в которых использовался теоретико-вероятностный подход к исследованию лингвистического текста для выяснения ряда частных задач математического, лингвистического и психологического порядка.

Обращение лингвистов к методам теории информации с ее математическим аппаратом диктовалось необходимостью найти новые приемы исследования таких единиц и связей языка и текста, которые не могут быть вскрыты ни путем простого перечисления, ни описания необходимых фактов, ни логических рассуждений по поводу этих фактов.

Поиски общих количественных оценок информационной эффективности имеют большое прикладное значение. С учетом этих характеристик строятся программы оптимального кодирования и автоматической переработки речевой информации. Однако информационные измерения языка и текста представляют интерес и с точки зрения лингвистической теории. Они нужны как для сравнения структур исследуемых языков, так и для измерения и описания языковых категорий.

Использование информационных измерений в лингвистике препятствовало, во-первых, отсутствию методики этих измерений применительно к языку, а во-вторых, то, что не были сформулированы те лингвистические задачи, которые могут быть решены с помощью информационных измерений. В этом смысле рецензируемая работа вполне своевременна и необходима.

Автор вводит читателя в общий круг теоретических проблем, связанных с информационными измерениями языка и речи, знакомит с технологией этих измерений (глава I). В книге представлены результаты информационных исследований по различным параметрам некоторых индоевропейских и тюркских языков (глава II, § 19, 18). Данные о количестве информации и избыточности различных языков могут быть использованы для типологических сопоставлений, а также для решения задач автоматического анализа и синтеза речи. Кроме того, в работе намечаются пути использования шенноповского метода угадывания для решения методических и психологических задач (глава II, § 22).

Систематическое, последовательное детальное описание методов информационных измерений позволяет повторить этот эксперимент для исследования нового лингвистического материала. Поэтому книга Р. Г. Пиотровского может быть использована и как учебное пособие для

студентов и аспирантов, изучающих математическое языкознание.

Приступая к информационным исследованиям языка, Р. Г. Пиотровский дает теоретическое обоснование рациональности использования модифицированного метода Шеннона по угадыванию информантом букв незнакомого текста для измерения статистической информации, содержащейся в речи. При этом автор исходит из положения, что язык это код, в котором заданы ограничения на сочетаемость единиц этого кода и вероятность их появления в речи.

В I главе книги «Эксперимент по угадыванию букв неизвестного текста» описывается процедура эксперимента по полной и сокращенной программе, дается методика расчета различных оценок энтропии (информации), показывается целесообразность учета достоверных продолжений и методы лингво-математической корректировки результатов угадывания¹. Автор описывает приемы лингвистического, психологического и математического порядка, с помощью которых можно улучшить результаты угадывания и приблизить их к результатам идеального предсказания. Здесь же приводится объяснение необходимости подбора экспериментального материала из разных источников с соблюдением определенных жанрово-стилистических пропорций для получения общеязыковых информационных характеристик, которые могут оказаться весьма полезными при расчете памяти специализированных ЭВМ, а также при определении целесообразности использования того или иного типа универсальных электронных вычислительных машин. Такие информационные оценки необходимы и при составлении программ автоматического анализа и синтеза, ориентированных на любой текст, написанный на данном языке, или на какой-либо стилиевой разновидности.

Особое внимание уделяет автор проверке достоверности получаемых измерений и количественных оценок.

В работе рассматривается несколько методов оценки достоверности результатов, в том числе и два новых лингво-психологических эксперимента: коллективное угадывание букв неизвестного текста и угадывание по методу Колмогорова.

В главе II «Информационные характеристики текста» приводятся данные об оценках предельной информации и избыточности трех стилей и общеязыковой

¹ О достоверных продолжениях при угадывании текста см. в работе А. А. Пиотровской, Р. Г. Пиотровского, К. А. Разживина «Энтропия русского языка», ВЯ, 1962, 6.

выборки для пяти европейских (русского, польского, английского, французского и румынского) и двух тюркских (азербайджанского и казахского) языков. Для сравнения демонстрируются информационные характеристики двух стилей адыгейского, армянского и чешского языков, полученные в результате использования других методических приемов. Эти сопоставления показывают, что структура развитых европейских языков оказывается более избыточной по сравнению с информационной насыщенностью тюркских языков. Использование теоретико-информационных оценок может иметь большие перспективы при установлении степени родства языков и различия в их морфологическом строе.

Если при выведении общих информационно-оценок текста и контроля этих оценок на достоверность применялись приемы, уже использовавшиеся в других работах, то введение в научный обиход приемов, связанных с оценкой количества информации отдельных лингвистических явлений, осуществлено Р. Г. Пиотровским впервые. К этим приемам относятся: а) определение лексической схемы текста; б) выведение информационных схем слова на буквенном, слоговом и морфемном уровнях; в) количественная оценка совокупности лингвистических ограничений в тексте, иначе — контекстной обусловленности; г) измерение флексивной и аналитической морфологии.

Дальнейшее развитие нового лингвоматематического метода находит применение в установлении соотношения лексических и грамматических ограничений в различных языках. Автор приходит к выводу, что грамматические связи в тексте устанавливаются быстрее, чем лексические.

В результате расчета грамматической информации, заключенной в префиксах, суффиксах, внутренних флексиях и окончаниях слова, Р. Г. Пиотровский получает данные, свидетельствующие о значительной избыточности грамматических аффиксов в развитых европейских языках. А если это так, то грамматические значения, присущие этим аффиксам, должны передаваться с помощью других средств, в первую очередь с помощью служебных слов» (стр. 92).

Рассматривая вопрос о соотношении флексивного способа и аналитической техники в передаче грамматических связей, автор анализирует информационно-статистический аппарат отношений, существующих между служебным и значительным словом, с одной стороны, и основой значительного слова и его окончанием, с другой.

Здесь выясняется, что служебное слово в значительно меньшей степени зависит от предшествующего слова, чем флексия. Однако вряд ли можно согласиться с предположением автора о высокой степени

обусловленности флексий за счет лексической основы, а не предшествующего контекста. Ведь основа знаменательного слова определяет графическую форму флексии (например, *-s* или *-es*, *-d* или *-ed* для английских существительных и глаголов), но выбор грамматической флексии для выражения необходимого грамматического времени, лица, числа, падежа и т. д., т. е. форм согласования и управления, определяется в первую очередь предыдущим контекстом.

Оценка доли лексико-грамматических ограничений в общей сумме контекстной обусловленности слова и, особенно, измерение грамматической информации (флексивной морфологии) по некоторым индоевропейским языкам выявила интересные и неожиданные явления. Так, если в отношении аналитического английского языка были количественно подтверждены ранее высказывавшиеся предположения о значительной роли контекста и большой избыточности флексивной морфологии, то для аналитического французского и румынского языков вес грамматической информации оказался примерно вдвое больше, чем для синтетического русского языка. Эта, обнаруженная информационным методом, особенность грамматического строя двух романских языков объясняется Р. Г. Пиотровским так:

«1. Глагольные парадигмы французского и румынского языков характеризуются значительным количеством флексивных графических форм (ср. французские *Conditionnel*, *Présent*, *Passé simple*, *Imparfait*, *Imparfait du Subjonctif*), превосходя в этом отношении не только английский, но и русский язык. Употребление французских глагольных окончаний не всегда предопределено формой существительного или местоимения, ср. фр. *je (tu) chantais, nous chantons* (*chantions*); рум. *eu cînt (cîntam, cîntai)*.

2. Романские служебные слова, играющие большую роль в передаче грамматической информации, часто сами имеют флексии (ср. формы разных видов артикля: фр. *le, l', la, les*; *un, une*; рум. *un, unui, unei*). Что же касается русских и английских служебных слов, то они чаще всего оказываются неизменяемыми» (стр. 98).

Следует, однако, отметить, что если задача оценки флексивной морфологии в работе решена полностью, то приемы информационного измерения и описания аналитических конструкций различных языков по-прежнему остаются неразработанными. В книге лишь намечаются перспективы дальнейшего углубленного развития информационного метода для исследования сложных вопросов анализа.

Информационно-статистическое исследование синтагматики языка завершается

построением усредненной пословной схемы текста. Модель текста свидетельствует о наличии сильных внутрисловных ограничений, которые проявляются гораздо отчетливее, чем контекстные связи между словами. Факт неравномерного распределения статистической информации в слове может иметь принципиальное значение при расчете памяти анализирующей или синтезирующей текст электронно-вычислительной машины.

Анализируя квантовый характер распределения статистической информации в схеме предложения, Р. Г. Пиотровский делает предположение, что такое распределение обусловлено особенностями работы головного мозга человека в ходе переработки им лингвистической информации. Если это действительно так, то информационные измерения текста помогут в определенной степени понять механизм функционирования головного мозга человека при переработке речевой информации.

III глава посвящена изучению информационно-статистических характеристик слова на буквенном, слоговом и морфемном уровнях в языке (внеконтекстные слова) и в речи (текстовые слова).

Здесь выясняется, что в тех языках (русском, английском, французском), где слог не совпадает с морфемой, слово имеет четко выраженную морфемную структуру, которая подавляет слоговое членение слова.

Из сказанного следует, что неоднократно высказывавшиеся предположения о том, что пиковые значения информации (энтропии) могут рассматриваться в качестве объективных показателей границ слога², не подтверждаются, а результаты работы по статистике и комбинаторике букв, буквосочетаний и слогов могут быть использованы лишь в качестве дополнительных данных для изучения языка на уровне знаков (морфем, слов...), так как сами по себе они не дают достаточных сведений о структуре конкретного языка.

Особый интерес с методической точки зрения представляют разделы книги, в которых излагаются результаты информационных лингво-статистических ограничений, накладываемых контекстом на сочетаемость букв, слогов и морфем в слове (внутрисловная обусловленность), а также на лексико-грамматические связи между словами. Было количественно подтверждено, что внутрисловные связи проявляются гораздо энергичнее, чем связи между словами в тексте. К сожалению, эти разделы изложены слишком лаконично.

² См.: Е. В. Падучева, Статистическое исследование структуры слога, сб. «Вопросы статистики речи», Л., 1958; А. А. Пиотровская, Р. Г. Пиотровский, К. А. Разживин, указ. соч.

Оценивая книгу Р. Г. Пиотровского, следует прежде всего выяснить ценности самого метода исследования лингвистического материала.

Известно, что информационно-статистический прием угадывания букв незнакомого текста информантом позволяет изучить распределение синтаксической (вероятностно-статистической) информации в тексте. Эксперимент не ставит целью непосредственное измерение прагматического и семантического содержания лингвистических символов, хотя совершенно очевидно, что все три аспекта языка не просто находятся в тесном взаимодействии, но определенным образом коррелированы и взаимообусловлены. Однако методика измерения семантической и прагматической информации в «чистом виде» в настоящее время еще не ясна ввиду отсутствия необходимой теории и математического аппарата. Поэтому наиболее реалистичными остаются исследования языка на уровне синтактики.

О недостатках и неясностях в технологии эксперимента по угадыванию весьма обстоятельно говорит сам автор работы, который рассматривает результаты проведенных исследований лишь как один из шагов на пути проб и ошибок, ведущих к познанию тайн функционирования языка в речи.

Однако, как уже было сказано, информационно-статистический метод вскрывает ряд принципиально важных характеристик семиотической природы языка и речи, позволяет определить иерархию единиц, образующих письменный текст, количественно оценить степень близости различных языков в комплексе их грамматики, фразеологии, лексики, фонетики и орфографии. Работа содержит ряд интересных типологических сопоставлений, позволяющих выяснить некоторые особенности исторического развития языков (процессы замены флективной техники аппаратом аналитического строя языка).

Информационное измерение пограничных сигналов в слове, а также неравномерное распределение статистической информации в слове может быть использовано при решении инженерно-лингвистических задач, связанных с кодированием и декодированием текстов, их автоматическим анализом и синтезом.

Монография Р. Г. Пиотровского имеет методический и лингво-психологический аспекты.

При решении задач психологического и методического характера объектом исследования становится сам информант и его отношение к предъявляемому тексту.

Дело в том, что в описываемом эксперименте испытуемый выступает одновременно в роли субъекта-наблюдателя и в функции вычислительного устройства, оценивающего вероятности отдельных букв на каждом шаге этого текста. Как отмечает автор, на основании прошлого

лингвистического опыта в сознании угадчика формируется степень убежденности по поводу того, какие лингвистические элементы встречаются чаще, а какие реже. Это состояние убежденности, а следовательно, и готовности к приему символов самого угадчика постоянно меняется по мере развертывания эксперимента. Сначала информант угадывает буквы текста, опираясь на свои знания букв, слогов, морфем и слов. Изменения состояния убежденности угадчика диктуется его лингвистическим опытом. Однако, отдавая одну-две фразы, а иногда всего лишь несколько слов, испытуемый начинает все чаще обращаться к своей собственной оценке тематики текста и своему жизненному и культурному опыту. Иными словами, эксперимент по угадыванию оценивает не только статистическую информацию текста, но также прагматическую информацию, связанную с индивидуальным восприятием этого текста угадчиком. Следовательно, метод угадывания может быть использован не только для информационно-измерений языка, но и для оценки индивидуальных особенностей человека, участвующего в эксперименте по угадыванию. Значение этой идеи для психологов трудно переоценить. Сюда относится предложенная автором мера степени владения языком, которой может служить сравнение оценок энтропии текста, получаемых от лиц, недостаточно владеющих языком, с оценками энтропии этого же текста, получаемых от носителей языка. Чем хуже владеет угадчик языком, на котором написан контрольный текст, тем выше количественные оценки энтропии и ниже избыточность этого текста для данного угадчика. На основе этих соображений был поставлен эксперимент в Ленинградском институте им. А. И. Герцена и в Минском институте иностранных языков. Сюда относятся также сопоставление величин энтропии или избыточности, получаемых при угадывании научно-делового (или вообще специального текста) специалистом в данной области знаний, с аналогичными ве-

личинами, получаемыми из эксперимента по угадыванию этого же текста неспециалистом, для оценки деловой квалификации³. Дальнейшее развитие этих идей (оценка степени трудности текстов и оценка способности принимать обоснованные решения о будущих событиях в условиях неопределенности) нашло свое отражение в последних работах автора и его сотрудников⁴.

Следует отметить, что тот прагматический «шум», от которого приходится абстрагироваться, чтобы приблизить результаты угадывания к идеальному предсказанию, здесь становится полезной информацией, отражающей индивидуальные особенности испытуемых. Это открывает новые возможности для объективной количественной оценки некоторых параметров таких субъективных (психических) процессов, как восприятие, память и мышление, для количественной оценки смыслового экспериментального материала, а также для оценки степени сформированности ряда навыков и умений.

В целом, книга Р. Г. Пиотровского, в которой разработана технология информационно-статистического измерения и описания лингвистического материала, представляет большой интерес для лингвистов, психологов, методистов и математиков. Эта работа представляет собой определенный вклад в методiku исследования языка, речи и проблем коммуникации.

Г. П. Богуславская, П. Б. Невельский

³ Такой эксперимент проведен в Харьковском университете.

⁴ «Проблемы языкознания. Доклады и сообщения советских ученых на X Международном конгрессе лингвистов», М., 1967; «Материалы III Всесоюзного съезда Общества психологов СССР», I, М., 1968; «XVI International congress of applied psychology», Amsterdam, 1968; «Conference on Psychology of human learning», Prague, 1969; «Actes du X Congrès international des linguistes», Bucarest (в печати).

R. Lafont. La phrase occitane. Essai d'analyse systématique. — Presses Universitaires de France, 1967. 522 стр.

Заголовком своим — «Окситанская фраза. Опыт систематического анализа»¹ — книга Р. Лафона привлечет внимание ограниченного количества читателей — главным образом, романистов и, в част-

¹ Термин «окситанская» (occitane) построен на основе слова *oc*, входящего в старое название провансальского языка *langue d'oc*.

ности, синтаксистов. Однако своим содержанием работа значительно шире узкой, с традиционной точки зрения, задачи, сформулированной в заглавии. Прежде всего, это — исследование, выполненное в духе школы Гийома и, стало быть, представляющее особый взгляд на понятие системы в языке, на соотношение языка и речи и вытекающих из тезиса о дихотомии «язык — речь» положений. Иными слова-