Оглавление

Введение

Глава 1. История применения математических методов в лингвистике

1.1. Становление структурной лингвистики на рубеже XIX – ХХ веков

1.2. Применение математических методов в лингвистике во второй половине ХХ века

Глава 2. Отдельные примеры использования математики в лингвистике

2.1. Машинный перевод

2.2.Статистические методы в изучении языка

2.3. Изучение языка методами формальной логики

2.4. Перспективы применения математических методов в лингвистике

Заключение

Литература

Приложение 1. Ronald Schleifer. Ferdinand de Saussure

Приложение 2. Фердинанд де Соссюр (перевод)

Введение

В ХХ веке наметилась продолжающаяся и поныне тенденция к взаимодействию и взаимопроникновению различных областей знаний. Постепенно стираются грани между отдельными науками; появляется всё больше отраслей умственной деятельности, находящихся «на стыке» гуманитарного, технического и естественнонаучного знания.

Другая очевидная особенность современности – стремление к изучению структур и составляющих их элементов. Поэтому всё большее место как в научной теории, так и на практике уделяется математике. Соприкасаясь, с одной стороны, с логикой и философией, с другой стороны, со статистикой (а, следовательно, и с общественными науками), математика всё глубже проникает в те сферы, которые на протяжении долгого времени было принято считать чисто «гуманитарными», расширяя их эвристический потенциал (ответ на вопрос «сколько» часто помоагет ответить и на вопросы «что» и «как). Исключением не стало и языкознание.

Цель моей курсовой работы – кратко осветить связь математики с такой отраслью языкознания, как лингвистика. Начиная с 50-х годов прошлого века, математика применяется в лингвистике при создании теоретического аппарата для описания строения языков (как естественных, так и искусственных). Однако следует сказать, что она не сразу нашла себе подобное практическое применение. Первоначально математические методы в лингвистике стали использоваться для того, чтобы уточнить основные понятия языкознания, однако с развитием компьютерной техники подобная теоретическая посылка стала находить применение на практике. Разрешение таких задач, как машинный перевод, машинный поиск информации, автоматическая обработка текста требовало принципиально нового подхода к языку. Перед лингвистами назрел вопрос: как научиться представлять языковые закономерности в том виде, в котором их можно подавать непосредственно на технику. Популярным в наше время термином «математическая лингвистика» называют любые лингвистические исследования, в которых применяются точные методы (а понятие точных методов в науке всегда тесно связано с математикой). Некоторые учёные прошлых лет, считают, что само выражение нельзя возводить в ранг термина, так как оно обозначает не какую-то особую «лингвистику», а лишь новое направление, ориентированное на усовершенствование, повышение точности и надёжности методов исследования языка. В лингвистике используются как количественные (алгебраические), так и неколичественные методы, что сближает её с математической логикой, а, следовательно, и с философией, и даже с психологией. Ещё Шлегель отмечал взаимодействие языка и сознания, а видный лингвист начала ХХ века Фердинанд де Соссюр (о его влиянии на становление математических методов в лингвистике расскажу позже) связывал структуру языка с его принадлежностью к народу. Современный исследователь Л. Перловский [43] идёт дальше, отождествляя количественные характеристики языка (например, число родов, падежей) с особенностями национального менталитета (об этом в разделе 2.2, «Статистические методы в лингвистике»).

Взаимодействие математики и языкознания – тема многогранная, и в своей работе я остановлюсь не на всех, а, в первую очередь, на её прикладных аспектах.

**Глава I. История применения математических методов в лингвистике**

**1.1 Становление структурной лингвистики на рубеже XIX – ХХ веков**

Математическое описание языка основано на представлении о языке как о механизме, восходящем к известному швейцарскому лингвисту начала ХХ века Фердинанду де Соссюру.

Начальное звено его концепции – теория языка как системы, cостоящей из трёх частей (собственно язык – *langue*, речь – *parole*, и речевую деятельность – *langage*), в которой каждое слово (член системы) рассматривается не само по себе, а в связи с другими членами. Как впоследствии отметил другой видный лингвист, датчанин Луи Ельмслев, Соссюр «первый требовал структурного подхода к языку, то есть научного описания языка путём регистрации соотношений между единицами» [22, c. 54].

Понимая язык как иерархическую структуру, Соссюр первым поставил проблему ценности, значимости языковых единиц. Отдельные явления и события (скажем, история происхождения отдельных индоевропейских слов) должны изучаться не сами по себе, а в системе, в которой они соотнесены с подобными же составляющими.

Структурной единицей языка Соссюр считал слово, «знак», в котором соединялись звучание и смысл. Ни один из этих элементов не существует друг без друга: поэтому носителю языка понятны различные оттенки значения многозначного слова как отдельного элемента в структурном целом, в языке.

Таким образом, в теории Ф. де Соссюра можно увидеть взаимодействие лингвистики, с одной стороны, с социологией и социальной психологией (следует отметить, что в это же время развиваются феноменология Гуссерля, психоанализ Фрейда, теория относительности Эйнштейна, происходят эксперименты над формой и содержанием в литературе, музыке и изобразительном искусстве), с другой стороны – с математикой (понятие системности соответствует алгебраической концепции языка). Подобная концепция изменила понятие языковой интерпретации как таковой: Явления стали трактоваться не относительно причин их возникновения, а относительно настоящего и будущего. Толкование перестало быть независимым от намерений человека (несмотря на то, что намерения могут быть безличными, «бессознательными» во фрейдистском понимании этого слова).

Функционирование же языкового механизма проявляется через речевую деятельность носителей языка. Результатом речи являются так называемые «правильные тексты» – последовательности речевых единиц, подчиняющиеся определённым закономерностям, многие из которых допускают математическое описание. Изучением способов математического описания правильных текстов (в первую очередь, предложений) занимается теория способов описания синтаксической структуры. В подобной структуре языковые аналогии определены не с помощью изначально присущих им качеств, а с помощью системных («структурных») отношений.

На Западе соссюровские идеи развивают младшие современники великого швейцарского лингвиста: в Дании – уже упомянутый мною Л. Ельмслев, давший начало алгебраической теории языка в своём труде «Основы лингвистической теории», в США – Э. Сепир, Л. Блумфилд, Ц. Харрис, в Чехии – русский учёный-эмигрант Н. Трубецкой.

Статистическими же закономерностями в изучении языка стал заниматься не кто иной, как основоположник генетики Георг Мендель. Только в 1968 году филологи обнаружили, что, оказывается, в последние годы жизни он был увлечен изучением лингвистических явлений с помощью методов математики. Этот метод Мендель привнёс в лингвистику из биологии; в девяностые годы девятнадцатого века лишь самые смелые лингвисты и биологи заявляли о целесообразности подобного анализа. В архиве монастыря св. Томаша в г. Брно, аббатом которого был Мендель, были найдены листки со столбцами фамилий , оканчивающимися на «mann», «bauer», «mayer», и с какими-то дробями и вычислениями. Стремясь обнаружить формальные законы происхождения фамильных имен, Мендель производит сложные подсчеты, в которых учитывает количество гласных и согласных в немецком языке, общее число рассматриваемых им слов, количество фамилий и т.д.

В нашей стране структурная лингвистика начала развиваться примерно в то же время, что и на Западе – на рубеже XIX-XX веков. Одновременно с Ф. де Соссюром понятие языка как системы разрабатывали в своих трудах профессора Казанского университета Ф.Ф. Фортунатов и И.А. Бодуэн де Куртенэ. Последний на протяжении долгого времени переписывался с де Соссюром, соответственно, женевская и казанская школы языкознания сотрудничали друг с другом. Если Соссюра можно назвать идеологом «точных» методов в лингвистике, то Бодуэн де Куртенэ заложил практические основы их применения. Он первым отделил лингвистику (как *точную* науку, использующую статистические методы и функциональную зависимость) от филологии (общности гуманитарных дисциплин, изучающих духовную культуру через язык и речь). Сам учёный считал, что «языкознание может принести пользу в ближайшем будущем, лишь освободившись от обязательного союза с филологией и историей литературы» [13, c.102]. «Испытательным полигоном» для внедрения математических методов в лингвистику стала фонология – звуки как «атомы» языковой системы, обладающие ограниченным количеством легко измеримых свойств, были самым удобным материалом для формальных, строгих методов описания. Фонология отрицает наличие смысла у звука, так что в исследованиях устранялся «человеческий» фактор. В этом смысле фонемы подобны физическим или биологическим объектам.

Фонемы, как самые мелкие языковые элементы, приемлемые для восприятия, представляют собой отдельную сферу, отдельную «феноменологическую реальность». Например, в английском языке звук «т» может произноситься по-разному, но во всех случаях человек, владеющий английским, будет воспринимать его как «т». Главное, что фонема будет выполнять свою главную – смыслоразличительную – функцию. Более того – различия между языками таковы, что разновидности одного звука в одном языке могут соответствовать разным фонемам в другом; например «л» и «р» в английском различны, в то время как в других языках это разновидности одной фонемы (подобно английскому «т», произнесённому с придыханием или без). Обширный словарный запас любого естественного языка представляет собой набор сочетаний гораздо меньшего количества фонем. В английском, например, для произнесения и написания около миллиона слов используется всего 40 фонем.

Звуки языка представляют собой системно организованный набор черт. В 1920е –1930е, вслед за Соссюром, Якобсон и Н.С.Трубецкой выделили «отличительные черты» фонем. Эти черты основаны на строении органов речи – языка, зубов, голосовых связок. Скажем, в английском разница между «т» и «д» заключается в наличии или отсутствии «голоса» (напряжении голосовых связок) и в уровне голоса, отличающем одну фонему от другой. Таким образом, фонологию можно считать примером общего языкового правила, описанного Соссюром: «В языке есть только различия» [24, c. 145]. Более важно даже не это: различие обычно подразумевает точные условия, между которыми оно и находится; но в языке существуют только различия без точных условий. Рассматриваем ли мы «обозначение» или «обозначаемое» – в языке не существует ни понятий, ни звуков, которые существовали бы до того, как развилась языковая система.

Таким образом, в соссюровском языкознании изучаемый феномен понимается как свод сопоставлений и противопоставлений языка. Язык – это и выражение значения слов, и средство общения, причём эти две функции никогда не совпадают. Мы можем заметить чередование формы и содержания: языковые контрасты определяют его структурные единицы, и эти единицы взаимодействуют, чтобы создать определённое значимое содержание. Так как элементы языка случайны, ни контраст, ни сочетание не могут быть основой. Значит, в языке отличительные признаки формируют фонетический контраст на другом уровне понимания, фонемы соединяются в морфемы, морфемы – в слова, слова – в предложения и т.д. В любом случае, целая фонема, слово, предложение и т.д. представляет собой нечто большее, чем просто сумма составляющих.

Соссюр предложил идею новой науки двадцатого века, отдельно от лингвистики изучающей роль знаков в обществе. Соссюр назвал эту науку семиологией (от греческого «semeîon» - знак). «Наука» семиотики, развивавшаяся в Восточной Европе в 1920е –1930е и в Париже в 1950е – 1960е, расширила изучение языка и лингвистических структур до литературных находок, составленных (или сформулированных) с помощью этих структур. Кроме того, на закате своей карьеры, параллельно совему курсу общей лингвистики, Соссюр занялся «семиотическим» анализом поздней римской поэзии, пытаясь открыть умышленно составленные анаграммы имён собственных. Этот метод был во многом противоположен рационализму в его лингвистическом анализе: он был попыткой, изучить в системе проблему «вероятности» в языке. Такое исследование помогает сосредоточиться на «вещественной стороне» вероятности; «ключевое слово», анаграмму которого ищет Соссюр, как утверждает Жан Старобинский, «инструмент для поэта, а не источник жизни стихотворения». Стихотворение служит для того, чтобы поменять местами звуки ключевого слова. По словам Старобинского, в этом анализе «Соссюр не углубляется в поиски скрытых смыслов». Напротив, в его работах заметно желание избежать вопросов, связанных с сознанием: «так как поэзия выражается не только в словах, но и в том, что порождают эти слова, она выходит из-под контроля сознания и зависит только от законов языка» (cм. Приложение 1).

Попытка Соссюра изучить имена собственные в поздней римской поэзии подчёркивает одну из составляющих его лингвистического анализа – произвольную природу знаков, а также формальную сущность соссюровской лингвистики, что исключает возможность анализа смысла. Тодоров делает вывод, что в наши дни труды Соссюра выглядят на редкость последовательными в нежелании изучать символы явления, имеющие чётко определённое значение [Приложение 1]. Исследуя анаграммы, Соссюр обращает внимание только на повторение, но не на предшествующие варианты. . . . Изучая «Песнь о Нибелунгах», он определяет символы только для того, чтобы присвоить их ошибочным чтениям: если они неумышленны, символов не существует. В конце концов, в своих трудах по общей лингвистике он делает предположение о существовании семиологии, описывающей не только лингвистические знаки; но это предположение ограничивается тем, что семиoлогия может описывать только случайные, произвольные знаки.

Раз это действительно так, то только потому, что не мог представить «намерение» без предмета; он не мог до конца преодолеть пропасть между формой и содержанием – в его трудах это превращалось в вопрос. Вместо этого он обращался к «языковой законности». Находясь между, с одной стороны, концепциями девятнадцатого века, основанными на истории и субъективных догадках, и методах случайной интерпретации, основанных на этих концепциях, и, с другой стороны, структуралистскими концепциями, стирающими противостояние между формой и содержанием (субъектом и объектом), значением и происхождением в структурализме, психоанализе и даже квантовой механике – труды Фердинанда де Соссюра по лингвистике и семиотике обозначают поворотный момент в изучении значений в языке и культуре.

Русские учёные были представлены и на Первом международном конгрессе лингвистов в Гааге в 1928 году. С. Карцевский, Р. Якобсон и Н. Трубецкой выступили с докладом, в котором рассматривалась иерархическая структура языка – в духе самых современных для начала прошлого века представлений. Якобсон в своих трудах развивал идеи Соссюра о том, что базовые элементы языка должны изучаться, в первую очередь, в связи со своими функциями, а не с причинами их возникновения.

К сожалению, после прихода в 1924 году к власти Сталина отечественное языкознание, как и многие другие науки, отбрасывает назад. Многие талантливые учёные вынуждены были эмигрировать, были высланы из страны или погибли в лагерях. Только с середины 1950-х годов стал возможен некоторый плюрализм теорий – об этом в разделе 1.2.

**1.2 Применение математических методов в лингвистике во второй половине ХХ века**

К середине ХХ века сформировалось четыре мировых лингвистических школы, каждая из которых оказалась родоначальником определённого «точного» метода. *Ленинградская фонологическая школа* (её родоначальником был ученик Бодуэна де Куртенэ Л.В. Щерба) использовала в качестве основного критерия обобщения звука в виде фонемы психолингвистический эксперимент, основанный на анализе речи носителей языка.

Учёные *Пражского лингвистического кружка*, в частности – его основатель Н.С. Трубецкой, эмигрировавший из России, разработали теорию оппозиций – семантическая структура языка была описана ими как набор оппозитивно постороенных семантических единиц – сем. Эта теория применялась в изучении не только языка, но и художественной культуры.

Идеологами *американского дескриптивизма* были языковеды Л. Блумфилд и Э. Сепир. Язык представлялся дескриптивистам в виде совокупности речевых высказываний, которые и были главным объектом их исследования. В центре их внимания оказались правила научного описания (отсюда название) текстов: изучение организации, аранжировка и классификация их элементов. Формализация аналитических процедур в области фонологии и морфологии (разработка принципов исследования языка на разных уровнях, дистрибутивного анализа, метода непосредственно составляющих и т.д.) привела к постановке общих вопросов лингвистического моделирования. Невнимание к плану содержания языка, а также парадигматической стороне языка не позволило дескриптивистам достаточно полно интерпретировать язык как систему.

В 1960-х годах развивается теория формальных грамматик, возникшая, главным образом, благодаря работам американского философа и лингвиста Н. Хомского. Он по праву считается одним из наиболее известных современных учёных и общественных деятелей, ему посвящено множество статей, монографий и даже полнометражный документальный фильм. По имени принципиально нового способа описания синтаксической структуры, изобретённого Хомским – генеративной (порождающей) грамматики – соответствующее течение в лингвистике получило название *генеративизма*.

Хомский, потомок выходцев из России, с 1945 года изучал в Пенсильванском университете лингвистику, математику и философию, находясь под сильным влиянием своего учителя Зелига Хэрриса – как и Хэррис, Хомский считал и считает свои политические взгляды близкими к анархизму (до сих пор он известен как критик существующего политического строя США и как один из духовных лидеров антиглобализма).

Первая крупная научная работа Хомского, магистерская диссертация «Морфология современного иврита*»* (1951), так и осталась неопубликованной. Докторскую степень Хомский получил в Пенсильванском университете в 1955, однако большая часть исследований, положенных в основу диссертации (полностью опубликованной только в 1975 под названием «Логическая структура лингвистической теории») и его первой монографии «Синтаксические структуры» (Syntactic Structures, 1957, рус. пер. 1962), была выполнена в Гарвардском университете в 1951–1955. В том же 1955 ученый перешел в Массачусетский технологический институт, профессором которого он стал в 1962.

В своём развитии теория Хомского прошла несколько этапов.

В первой монографии «Синтактические структуры» учёный представил язык как механизмепорождения бесконечного множества предложений с помощью конечного набора грамматических средств. Для описания языковых свойств он предложил понятия глубинной (скрытой от непосредственного восприятия и порождаемой системой рекурсивных, т.е. могущих применяться многократно, правил) и поверхностной (непосредственно воспринимаемой) грамматических структур, а также трансформаций, описывающих переход от глубинных структур к поверхностным. Одной глубинной структуре могут соответствовать несколько поверхностных (например, пассивная конструкция *Указ подписывается президентом* выводится из той же глубинной структуры, что и активная конструкция *Президент подписывает указ*) и наоборот (так, неоднозначность *Мать любит дочь* описывается как результат совпадения поверхностных структур, восходящих к двум различным глубинным, в одной из которых мать – та, кто любит дочь, а в другой – та, кого любит дочь).

Стандартной теорией Хомского считается модель «Аспектов», изложенная в книге Хомского «Аспекты теории синтаксиса». В этой модели в формальную теорию впервые вводились правила семантической интерпретации, приписывающих значение глубинным структурам. В «Аспектах» языковая компетенция противопоставлена употреблению языка (performance), принята так называемая гипотеза Катца – Постала о сохранении смысла при трансформации, в связи с чем исключено понятие факультативной трансформации, а также введен аппарат синтаксических признаков, описывающих лексическую сочетаемость.

В 1970-е Хомский работает над теорией управления и связывания (GB-теория – от слов *government* и *binding*) – более общей, нежели предыдущая. В ней учёный отказался от специфических правил, описывающих синтаксические структуры конкретных языков. Все трансформации были заменены одной универсальной трансформацией перемещения. В рамках GB-теории существуют и частные модули, каждый из которых отвечает за свою часть грамматики.

Уже недавно, в 1995 году, Хомский выдвинул минималистскую программу, где человеческий язык описывается подобно машинному. Это лишь программа – не модель и не теория. В ней Хомский выделяет две главных подсистемы языкового аппарата человека: лексикон и вычислительную систему, а также два интерфейса – фонетический и логический.

Формальные грамматики Хомского стали классическими для описания не только естественных, но и искусственных языков – в частности, языков программирования. Развитие структурной лингвистики во второй половине ХХ века можно по праву считать «хомскианской революцией».

*Московская фонологическая школа*, представителями которой были А.А. Реформатский, В.Н. Сидоров, П.С. Кузнецов, А.М. Сухотин, Р.И. Аванесов, использовала подобную же теорию для изучения фонетики. Постепенно «точные» методы начинают применяться касаемо не только фонетики, но и синтаксиса. Структурностью языка начинают заниматься и лингвисты, и математики – как у нас, так и за рубежом. В 1950-60е в СССР начинается новый этап во взаимодействии математики и лингвистики, связанный с разработкой систем машинного перевода.

Толчком к началу этих работ в нашей стране послужили первые разработки в области машинного перевода в США (хотя первое механизированное переводное устройство П.П. Смирнова-Троянского было изобретено в CCCР ещё в 1933 году, оно, будучи примитивным, не получило распространения). В 1947 году А.Бутт и Д. Бриттен придумали код для пословного перевода с помощью ЭВМ, годом позже Р.Риченс предложил правило разбиения слов на основу и окончание при машинном переводе. В те годы довольно сильно отличались от современных. Это были очень большие и дорогие машины, которые занимали целые комнаты и требовали для своего обслуживания большой штат инженеров, операторов и программистов. В основном эти компьютеры использовались для осуществления математических расчетов для нужд военных учреждений – новое в математике, физике и технике служило, в первую очередь, военному делу. На ранних этапах разработка МП активно поддерживалась военными, при этом (в условиях «холодной войны») в США развивалось русско-английское направление, а в СССР - англо-русское.

В январе 1954 года в Массачусетском техническом университете состоялся «Джорджтаунский эксперимент»– первая публичная демонстрация перевода с русского языка на английский на машине ИБМ-701. Реферат сообщения об удачном прохождении эксперимента, сделанный Д.Ю. Пановым, появился в РЖ «Математика», 1954, №10: «Перевод с одного языка на другой при помощи машины: отчёт о первом успешном испытании».

К работам по машинному переводу Д. Ю. Панов (в то время директор Института научной информации – ИНИ, позднее ВИНИТИ) привлёк И. К. Бельскую, которая позднее возглавит группу машинного перевода в Институте точной математики и вычислительной техники АН СССР. К концу 1955 года относится первый опыт перевода с английского языка на русский при помощи машины БЭСМ. Программы для БЭСМ составляли Н.П. Трифонов и Л.Н. Королёв, кандидатская диссертация которого была посвящена методам построения словарей для машинного перевода.

Параллельно работы по машинному переводу велись в Отделении прикладной математики Математического института АН СССР (сейчас Институт прикладной математики имени М.В. Келдыша РАН). По инициативе математика А.А. Ляпунова. К работам по переводу текстов на машине «Стрела» с французского языка на русский он привлёк аспирантку МИАН О.С. Кулагину и своих учениц Т.Д. Вентцель и Н.Н. Рикко. Представления Ляпунова и Кулагиной о возможности использования техники для перевода с одного языка на другой были опубликованы в журнале «Природа», 1955, №8. С конца 1955 года к ним присоединилась Т.Н. Молошная, затем приступившая к самостоятельной работе над алгоритмом англо-русского перевода.

Р.Фрумкина [37, c.12], занимавшаяся в то время алогритмом перевода с испанского, вспоминает, что на этом этапе работ сложно было делать какие-то последовательные шаги. Гораздо чаще приходилось следовать эвристическому опыту – своему или коллег.

Однако первое поколение систем машинного перевода было весьма несовершенным. Все они базировались на алгоритмах последовательного перевода «слово за словом», «фраза за фразой» – смысловые связи между словами и предложениями никак не учитывались. Для примера можно привести предложения: «John was looking for his toy box. Finally he found it. The box was in the pen. John was very happy. (Джон искал свою игрушечную коробку. Наконец он ее нашел. Коробка была в манеже. Джон был очень счастлив.)». «Pen» в данном контексте – не «ручка» (инструмент для письма), а «детский манеж» (play-pen). Знание синонимов, антониов и переносных значений сложно вводить в компьютер. Перспективным направлением становилась разработка машинных систем, ориентированных на использование человеком-переводчиком.

Со временем на смену системам прямого перевода пришли Т-системы (от английского слова «transfer» – преобразование), в которых перевод осуществлялся на уровне синтаксических структур. В алгоритмах Т-систем использовался механизм, позволяющий построить синтаксическую структуру по правилам грамматики языка входного предложения (подобно тому, как учат иностранному языку в средней школе), а затем синтезировать выходное предложение, преобразуя синтаксическую структуру и подставляя из словаря нужные слова.

Ляпунов говорил о переводе путём извлечения смысла переводимого текста и его представления на другом языке. Подход к построению систем машинного перевода, основанный на получении смыслового представления входного предложения путём его семантического анализа и синтеза входного предложения по полученному смысловому представлению, до сих пор считается наиболее совершенным. Такие системы называют И-системами (от слова «интерлингва»). Однако задача по их созданию, поставленная ещё в конце 50-х – начале 60-х, не решена полностью до сих пор, несмотря на усилия Международной федерации IFIP – мирового сообщества учёных в области обработки информации.

Учёные задумались над тем, как нужно формализовать и строить алгоритмы для работы с текстами, какие словари надо вводить в машину, какие лингвистические закономерности следует использовать при машинном переводе. Такими представлениями традиционная лингвистика не располагала – не только в части семантики, но и в части синтаксиса. Ни для одного языка в то время не существовало перечней синтаксических конструкций, не были изучены условия их сочетаемости и взаимозмаеняемости, не были разработаны правила построения крупных единиц синтаксической структуры из более мелких составляющих элементов.

Потребность в создании теоретических основ машинного перевода и привела к формированию и развитию математической лингвистики. Ведущую роль в этом деле в СССР сыграли математики А.А. Ляпунов, О.С. Кулагина, В.А. Успенский, лингвисты В.Ю. Розенцвейг, П.С. Кузнецов, Р.М. Фрумкина, А.А. Реформатский, И.А. Мельчук, В.В. Иванов. Диссертация Кулагиной была посвящена исследованию формальной теории грамматик (одновременно с Н.Хомским в США), Кузнецов выдвинул задачу аксиоматизации лингвистики, восходящую к работам Ф.Ф. Фортунатова.

6 мая 1960 года было принято Постановление Президиума АН СССР «О развитии структурных и математических методов исследования языка», в Институте языкознания и Институте русского языка были созданы соответствующие подразделения. С 1960 года в ведущих гуманитарных вузах страны – филологическом факультете МГУ, Ленинрадском, Новосибирском университетах, МГПИИЯ – началась подготовка кадров в области автоматической обработки текста.

Однако работы по машинному переводу этого периода, называемого «классическим», представляют собой скорее теоретический, нежели практический интерес. Экономически эффективные системы машинного перевода стали создаваться только в восьмидесятые годы прошлого века. Об этом я расскажу позже, в разделе 2.1, «Машинный перевод».

К 1960-м – 70-м годам относятся глубокие теоретические разработки, использующие методы теории множеств и математической логики, такие, как теория поля и теория нечётких множеств.

Автором теории поля в лингвистике был советский поэт, переводчик и лингвист В.Г. Адмони. Свою теорию он изначально разрабатывал на основе немецкого языка. У Адмони понятие «поле» обозначает произвольное непустое множество языковых элементов (например, «лексическое поле», «семантическое поле»).

Структура поля неоднородна: оно состоит из ядра, элементы которого обладают полным набором признаков, определяющих множество, и периферии, элементы которой могут обладать как признаками данного множества (не всеми), так и соседних. Приведу пример, иллюстрирующий данное высказывание: скажем, в английском языке поле сложных слов («day-dream» – «мечтать» трудноотделимо от поля словосочетаний («tear gas» – «слезоточивый газ»).

С теорией поля тесно связана уже упомянутая выше теория нечётких множеств. В СССР её обоснованием занимались лингвисты В.Г. Адмони, И.П. Иванова, Г.Г. Поченцов, однако её родоначальником был американский математик Л.Заде, в 1965 году выпустивший статью «Fuzzy Logic». Давая математическое обоснование теории нечётких множеств, Заде рассматривал их на лингвистическом материале.

В этой теории речь идёт уже не столько о принадлежности элементов к данному множеству (А∈а), сколько о степени этой принадлежности (μА∈а), так как периферийные элементы могут в той или иной мере принадлежать нескольким полям. Заде (Лофти-заде) был выходцем из Азербайджана, до 12 лет имел практику общения на четырех языках - азербайджанском, русском, английском и персидском - и пользовался тремя различными алфавитами: кириллицей, латинским, арабским. Когда ученого спрашивают, что общего между теорией нечетких множеств и лингвистикой, он не отрицает этой связи, но уточняет: «Я не уверен, что изучение этих языков оказало большое влияние на мое мышление. Если это и имело место, то разве что подсознательно». В юности Заде учился в Тегеране в пресвитерианской школе, а после Второй мировой войны эмигрировал в США. «Вопрос не в том, являюсь ли я американцем, русским, азербайджанцем или кем-то еще, - сказал он в одной из бесед, - я сформирован всеми этими культурами и народами и чувствую себя достаточно комфортабельно среди каждого из них» [34, c.13]. В этих словах есть нечто родственное тому, что характеризует теорию нечетких множеств – отход от однозначных определений и резких категорий.

В нашей стране в 70е переводятся и изучаются труды западных лингвистов ХХ века. И.А. Мельчук перевёл на русский язык сочинения Н. Хомского. Н.А. Слюсарева в своей книге «Теория Ф. де Соссюра в свете современной лингвистики» связывает постулаты соссюровского учения с актуальными проблемами лингвистики 70-х. Намечается тенденция к дальнейшей математизации лингвистики. В ведущих отечественных вузахидёт подготовка кадров по специальности «Математическая (теоретическая, прикладная) лингвистика». В это же время на Западе происходит резкий скачок в развитии вычислительной техники, для чего требуются всё более новые лингвистические основы.

В 1980-е годы профессор Института востоковедения АН Ю.К. Лекомцев, занимаясь анализом языка лингвистики через анализ схем, таблиц и других видов записи, используемых в лингвистических описаниях, рассматривает математические системы, пригодные для этих целей (в основном – системы матричной алгебры).

Таким образом, на протяжении всего ХХ века шло сближение точных и гуманитарных наук. Взаимодействие математики с лингвистикой всё чаще находило практическое применение. Об этом – в следующей главе.

**Глава 2. Отдельные примеры использования математики в лингвистике**

**2.1 Машинный перевод**

Идея перевода с одного языка на другой при помощи универсального механизма возникла несколькими веками раньше, чем начались первые разработки в этой области – ещё в 1649 году Рене Декарт предложил идею языка, в котором эквивалентные идеи разных языков выражались бы одним символом. Первые попытки осуществить эту идею в 1930-40е, начало теоретических разработок в середине века, усовершенствование систем перевода при помощи техники в 1970-80е, бурное развитие переводческой техники в последнее десятилетие – таковы этапы развития машинного перевода как отрасли. Именно из работ по машинному переводу выросла компьютерная лингвистика как наука.

С развитием вычислительной техники в конце 70х – начале 80х исследователи задались более реалистичными и экономически выгодными целями – машина становилась не конкурентом ( как предполагалось раньше), а помощником человека-переводчика. Машинный перевод перестаёт служить исключительно военным задачам (все советские и американские изобретения и исследования, ориентированные, в первую очередь, на русский и английский языки, в той или иной мере способствовали «холодной войне»). В 1978 году слова естественного языка были переданы в объединённой сети Arpa, шестью годами позже в США появились первые программы перевода для микрокомпьютеров.

В 70е Комиссия Европейских Общин покупает англо-французскую версию компьютерного переводчика Systran, заказывая также франко-аглийскую и итало-английскую версии, и систему перевода с русского на английский, использовавшуюся американскими Вооружёнными Силами. Так были заложены основы проекта EUROTRA.

О возрождении машинного перевода в 70-80-е гг. свидетельствуют следующие факты: Комиссия Европейских общин (CEC) покупает англо-французскую версию Systran, а также систему перевода с русского на английский (последняя развивалась после доклада ALPAC и продолжала использоваться ВВС США и НАСА); кроме того, CEC заказывает разработку франко-английской и итало-английской версий. Одновременно происходит быстрое расширение деятельности по созданию систем машинного перевода в Японии; в США Панамериканская организация здравоохранения (PAHO) заказывает разработку испано-английского направления (система SPANAM); ВВС США финансируют разработку системы машинного перевода в Лингвистическом исследовательском центре при Техасском университете в Остине; группа TAUM в Канаде достигает заметных успехов в разработке своей системы METEO (для перевода метеорологических сводок). Целый ряд проектов, начатых в 70-80-е гг. впоследствии развились в полноценные коммерческие системы.

За период 1978-93 в США на исследования в области машинного перевода истрачено 20 миллионов долларов, в Европе - 70 миллионов, в Японии - 200 миллионов.

Одной из новых разработок стала технология TM (translation memory), работающая по принципу накопления: в процессе перевода сохраняется исходный сегмент (предложение) и его перевод, в результате чего образуется лингвистическая база данных; если идентичный или подобный исходному сегмент обнаруживается во вновь переводимом тексте, он отображается вместе с переводом и указанием совпадения в процентах. Затем переводчик принимает решение (редактировать, отклонить или принять перевод), результат которого сохраняется системой, поэтому не нужно дважды переводить одно и то же предложение. В настоящее время разработчиком известной коммерческой системы, основанной на технологии TM, является система TRADOS (основана в 1984 г.).

В настоящее несколько десятков компаний занимаются разработкой коммерческих систем машинного перевода, в их числе: Systran, IBM, L&H (Lernout & Hauspie), Transparent Language, Cross Language, Trident Software, Atril, Trados, Caterpillar Co., LingoWare; Ata Software; Lingvistica b.v. и др. Появилась возможность воспользоваться услугами автоматических переводчиков непосредственно в Сети: alphaWorks; PROMT's Online Translator; LogoMedia.net; AltaVista's Babel Fish Translation Service; InfiniT.com; Translating the Internet.

Коммерчески эффективные переводческие системы появились во второй половине 80х и в нашей стране. Расширилось само понятие машинного перевода (к нему стали относить «создание целого ряда автоматических и автоматизированных систем и устройств, выполняющих автоматически или полуавтоматически весь цикл перевода либо отдельные задачи в диалоге с человеком» [29, c.13]), увеличились государственные ассигнования на развитие этой отрасли.

Основными языками отечественных переводческих систем стали русский, английский, немецкий, французский и японский. Во Всесоюзном центре переводов (ВЦП) была разработана система перевода с английского и немецкого языков на русский на машине ЭВМ ЕС-1035 –АНРАП. Она состояла из трёх словарей – входных английского и немецкого и выходного русского – под единым программным обеспечением. Существовало несколько сменных специализированных словарей – по вычислительной технике, программированию, радиоэлектронике, машиностроению, сельскому хозяйству, металлургии. Система могла работать в двух режимах – автоматическом и интерактивном, когда на экране отображался пофразно исходный текст и перевод, который человек мог отредактировать. Скорость перевода текста на АНРАП (от начала набора до окончания печати) составляла примерно 100 страниц в час.

В 1989 году было создано семейство коммерческих переводчиков типа СПРИНТ, работавших с русским, английским, немецким и японским языками. Их главным преимуществом стала их совместимость с IBM PC – таким образом отечественные системы машинного перевода достигали международного уровня качества. В это же время разрабатывается система машинного перевода с французского языка на русский ФРАП, включающая в себя 4 этапа анализа текста: графематический, морфологический, синтаксический и семантический. В ЛГПИ им. Герцена шла работа над четырёхязычной (английский, французский. Испанский, русский) системой СИЛОД-МП (в промышленном режиме эксплуатировались англо-русский и франко-русский словари.

Для специализированного перевода текстов по электротехнике существовала система ЭТАП-2. Анализ входного текста в ней осуществлялся на двух уровнях – морфологическом и синтаксическом. Словарь ЭТАП-2 содержал около 4 тысяч статей; этап преобразования текста – около 1000 правил (96 общих, 342 частных, остальные – словарные). Всё это обеспечивало удовлетворительное качество перевода (скажем, заголовок патента «Optical phase grid arrangement and coupling device having such an arrangement» переводился как «Устройство оптической фазовой сетки и соединяющее устройство с таким устройством» [29, c.20] – несмотря на тавтологию, смысл сохранён).

В Минском педагогическом институте иностранных языков на базе англо-русского словаря словоформ и оборотов была изобретена система машинного перевода заголовков), в Институте востоковедения АН – система перевода с японского на русский. Созданная в Московском НИИ систем автоматизации первая автоматическая словарно-терминологическая служба (СЛОТЕРМ) по вычислительной технике и программированию содержала примерно 20000 терминов в толковом словаре и специальных словарях для лингвистических исследований.

Системы машинного перевода постепенно стали использоваться не только по прямому назначению, но и как важный компонент автоматических обучающих систем (для обучения переводу, контроля орфографических и грамматических знаний).

90-е годы принесли с собой бурное развитие рынка ПК (от настольных до карманных) и информационных технологий, широкое использование сети Интернет (которая становится все более интернациональной и многоязыкой). Все это сделало востребованным дальнейшее развитие автоматизированных переводческих систем. С начала 1990-х гг. на рынок систем ПК выходят и отечественные разработчики.

В июле 1990 года на выставке PC Forum в Москве была представлена первая в России коммерческая система машинного перевода под названием PROMT (PROgrammer's Machine Translation). В 1991 г. было создано ЗАО "ПРОект МТ", и уже в 1992 г. компания "ПРОМТ" выиграла конкурс NASA на поставку систем МП (ПРОМТ была единственной неамериканской фирмой на этом конкурсе). В 1992 г. "ПРОМТ" выпускает целое семейство систем под новым названием STYLUS для перевода с английского, немецкого, французского, итальянского и испанского языков на русский и с русского на английский, а в 1993 г. на базе STYLUS создается первая в мире система машинного перевода для Windows. В 1994 г. вышла версия STYLUS 2.0 для Windows 3.Х/95/NT, а в 1995-1996 гг. представлено третье поколение систем машинного перевода, полностью 32-разрядных STYLUS 3.0 для Windows 95/NT, одновременно с этим успешно завершена разработка совершенно новых, первых в мире русско-немецкой и русско-французской систем машинного перевода.

В 1997 г. подписано соглашение с французской фирмой Softissimo о создании систем перевода с французского языка на немецкий и английский и обратно, а в декабре этого года была выпущена первая в мире система немецко-французского перевода. В этом же году компания "ПРОМТ" компания выпустила систему, реализованную по технологии Гигант , поддерживающей несколько языковых направлений в одной оболочке, а также специальный переводчик для работы в Интернете WebTranSite.

В 1998 г. выпускается целое созвездие программ под новым названием PROMT 98. Через год компания ПРОМТ выпустила два новых продукта: уникальный пакет программ для работы в Интернете - PROMT Internet, и переводчик для корпоративных почтовых систем - PROMT Mail Translator. В ноябре 1999 года PROMT была признана лучшей системой машинного перевода среди тестируемых французским журналом PC Expert, обойдя конкурентов по сумме показателей на 30 процентов. Для корпоративных клиентов разработаны также специальные серверные решения - корпоративный сервер переводов PROMT Translation Server (PTS) и Интернет-решение PROMT Internet Translation Server (PITS). В 2000 г. "ПРОМТ" обновила всю линию своих программных продуктов, выпустив МП системы нового поколения: PROMT Translation Office 2000, PROMT Internet 2000 и Magic Gooddy 2000.

Перевод в режиме он-лайн при поддержке системы "ПРОМТ" используется на ряде отечественных и зарубежных сайтов: PROMT's Online Translator, InfiniT.com, Translate.Ru, Lycos и др., а также в учреждениях различного профиля для перевода деловой документации, статей и писем (существуют системы перевода, встраиваемые непосредственно в Outlook Express и другие почтовые клиенты).

В наше время появляются новые технологии машинного перевода, основанные на использовании систем искусственного интеллекта, статистических методах. О последних – в следующем разделе.

**2.2 Статистические методы в изучении языка**

Немалое внимание в современной лингвистике отводится изучению языковых явлений методами количественной математики. Количественные данные часто помогают более глубоко осмыслить изучаемые явления, их место и роль в системе смежных явлений. Ответ на вопрос «сколько» помогает ответить и на вопросы «что», «как», «почему» – таков эвристический потенциал количественной характеристики.

Немалую роль статистические методы играют в разработке систем машинного перевода (см. раздел 2.1). При статистическом подходе проблема перевода рассматривается в терминах канала с помехами. Представим себе, что нам нужно перевести предложение с английского на русский. Принцип канала с помехами предлагает нам следующее объяснение отношений между английской и русской фразой: английское предложение представляет собой не что иное, как русское предложение, искаженное неким шумом. Для того чтобы восстановить исходное русское предложение, нам нужно знать, что именно люди обычно говорят по-русски и как русские фразы искажаются до состояния английского. Перевод осуществляется путем поиска такого русского предложения, которое максимизирует произведения безусловной вероятности русского предложения и вероятности английского предложения (оригинала) при условии данного русского предложения. Согласно теореме Байеса, это русское предложение является наиболее вероятным переводом английского:



где e – предложение перевода, а f – предложение оригинала

Таким образом, нам требуется модель источника и модель канала, или модель языка и модель перевода. Модель языка должна присваивать оценку вероятности любому предложению конечного языка (в нашем случае, русского), а модель перевода –предложению оригинала. (cм. табл.1)

Табл.1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *amount* | *bonus* | *compensation* | *payment* | *rate* |
| *выплата* | 15% | 8% | 6% | 71% | 0% |
| *оплата* | 0% | 0% | 0% | 97% | 3% |

В общем случае система машинного перевода работает в двух режимах:

1. Обучение системы: берется тренировочный корпус параллельных текстов, и с помощью линейного программирования ищутся такие значения таблиц переводных соответствий, которые максимизируют вероятность (например) русской части корпуса при имеющейся английской согласно выбранной модели перевода. На русской части того же корпуса строится модель русского языка.

2. Эксплуатация: на основе полученных данных для незнакомого английского предложения ищется русское, максимизирующее произведение вероятностей, присваиваемых моделью языка и моделью перевода. Программа, используемая для такого поиска, называется дешифратором.

Самой простой статистической моделью перевода является модель дословного перевода. В этой модели предполагается, что для перевода предложения с одного языка на другой достаточно перевести все слова (создать «мешок слов»), а расстановку их в правильном порядке обеспечит модель Для приведения P(a, f | e) к P(a | e, f), т.е. вероятности данного выравнивания при данной паре предложений, каждая вероятность P(a, f | e) нормализуется по сумме вероятностей всех выравниваний данной пары предложений:

Реализация алгоритма Витерби, используемая для обучения Модели №1, состоит в следующем:

1.Вся таблица вероятностей переводных соответствий заполняется одинаковыми значениями.

2. Для всех возможных вариантов попарных связей слов вычисляется вероятность P(a, f | e):

3. Значения P(a, f | e) нормализуются для получения значений P(a | e, f).

4. Подсчитывается частота каждой переводной пары, взвешенная по вероятности каждого варианта выравнивания.

5. Полученные взвешенные частоты нормализуются и формируют новую таблицу вероятностей переводных соответствий

6. Алгоритм повторяется с шага 2.

Рассмотрим в качестве примера тренировку подобной модели на корпусе из двух пар предложений (рис.2):

- Белый Дом/White House

* Дом/House

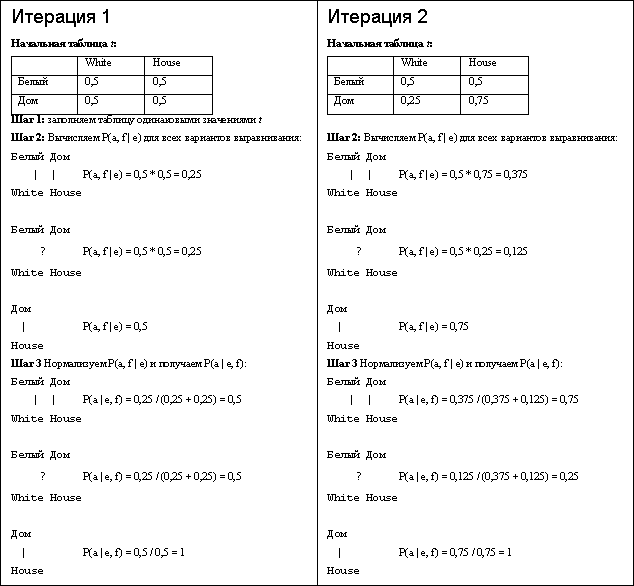


Рис.1

После большого числа итераций мы получим таблицу (табл.2.), из которой видно, что перевод осуществляется с высокой точностью.

Табл.2

|  |  |  |
| --- | --- | --- |
|  | *White* | *House* |
| *белый* | 0,9999 | 0, 0001 |
| *дом* | 0, 0001 | 0,9999 |

Также статистические методы широко используются в изучении лексики, морфологии, синтаксиса, стилистики. Учёные Пермского государственного университета провели исследование, в основе которого лежало утверждение о том, что стереотипные словосочестания являются важным «строительным материалом» текста [13, c.24]. Эти словосочетания состоят из «ядерных» повторяющихся слов и зависимых слов-конкретизавторов и имеют ярко выраженную стилистическую окраску.

В научном стиле «ядерными» словами можно назвать: *исследование, изучение, задача, проблема, вопрос, явление, факт, наблюдение, анализ* и др. В публицистике «ядерными» будут уже другие слова, обладающие повышенной ценностью именно для текста газеты: *время, лицо, власть, дело, действие, закон, жизнь, история, место* и т.д. (всего 29)

Особый интерес для лингвистов представляет также профессиональная диффереренциация общенародного языка, своеобразие использования лексики и грамматики в зависимости от рода занятий. Известно, что шофёры в профессиональной речи употребляю форму ш***о***фер, медики говорят к***о***клюш вместо кокл***ю***ш – подобных примеров можно привести. Задача статистики – проследить за вариативностью произношения и изменением языковой нормы.

Профессиональные различия ведут за собой различия не только грамматические, но и лексические. В Якутском государственном университете им. М.К. Аммосова было проанализировано по 50 анкет с наиболее часто встречающимися реакциями на некоторые слова среди медиков и строителей (табл.3) [13, c.78].

Табл.3

|  |  |  |
| --- | --- | --- |
| Стимул | Медики | Строители |
| *человек* | *пациент (10), личность (5)* | *мужчина (5)* |
| *добро* | *помощь (8), помогать (7)* | *зло (16)* |
| *жизнь* | *смерть (10)* | *прекрасная (5)* |
| *смерть* | *труп (8)* | *жизнь (6)* |
| *огонь* | *жар (8), ожог (6)* | *пожар (7)* |
| *палец* | *рука (14), панариций (5)* | *большой (7), указательный (6)* |
| *глаза* | *зрение (6), зрачок, окулист (по 5)* | *карие (10), большие (6)* |
| *голова* | *ум (14), мозги (5)* | *большая (9), умная (8), ум (6)* |
| *терять* | *сознание, жизнь (по 4)* | *деньги (5), находить (4)* |

Можно заметить, что медики чаще, чем строители, дают ассоциации, связанные с их профессиональной деятельностью, так как приведённые в анкете слова-стимулы имеют к их профессии больше отношения, чем к профессии строителя.

Статистические закономерности в языке используются для создания частотных словарей – словарей, в которых приводятся числовые характеристики употребительности слов (словоформ, словосочетаний) какого-либо языка – языка писателя, какого-либо произведения и т. п. Обычно в качестве характеристики употребительности используется частота встречаемости слова в тексте определенного объема

Модель восприятия речи невозможна без словаря как своего существеннейшего компонента. При восприятии речи основной оперативной единицей выступает слово. Из этого следует, в частности, что каждое слово воспринимаемого текста должно быть отождествлено с соответствующей единицей внутреннего словаря слушающего (или читающего). Естественно считать, что уже с самого начала поиск ограничен некоторыми подобластями словаря. Согласно большинству современных теорий восприятия речи, собственно фонетический анализ звучащего текста в типичном случае дает лишь некоторую частичную информацию о возможном фонологическом облике слова, и такого рода информации отвечает не одно, а определенное МНОЖЕСТВО слов словаря; следовательно, возникает две задачи:

(а) выделить соответствующее множество по тем или иным параметрам;

(б) в пределах очерченного множества (если оно выделено адекватно) произвести «отсев» всех слов, кроме того единственного, которое и соответствует наилучшим образом данному слову распознаваемого текста. Одна из стратегий «отсева» – исключение низкочастотных слов. Отсюда следует, что словарь для восприятия речи – это частотный словарь. Именно создание компьютерной версии частотного словаря русского языка и является первоначальной задачей представляемого проекта.

На материале русского языка существует 5 частотных словарей (не считая отраслевых). Отметим лишь некоторые общие недостатки имеющихся словарей.

Все известные частотные словари русского языка построены на обработке массивов письменных (печатных) текстов. Отчасти по этой причине, когда тождество слова во многом опирается на совпадение формальное, графическое, недостаточно учитывается семантика. В результате оказываются смещенными, искаженными и частотные характеристики; например, если слова из сочетания «друг друга» составитель частотного словаря включает в общую статистику употребления слова «друг», то едва ли это оправданно: учитывая семантику, мы должны признать, что это уже другие слова, а точнее, что самостоятельной словарной единицей выступает лишь само по себе сочетание в целом.

Также во всех существующих словарях слова помещены лишь в своих основных формах: существительные в форме единственного числа, именительного падежа, глаголы в форме инфинитива и т.д. Некоторые из словарей дают информацию о частотности словоформ, но обычно делают это недостаточно последовательно, не исчерпывающим образом. Частотности разных словоформ одного и того же слова заведомо не совпадают. Разработчик же модели восприятия речи должен учитывать, что в реальном перцептивном процессе распознаванию подлежит именно конкретная словоформа, «погруженная» в текст: на базе анализа начального участка экспонента словоформы формируется множество слов с идентичным началом, причем начальный участок словоформы не обязательно тождествен начальному участку словарной формы. Именно словоформе принадлежит конкретная ритмическая структура – также чрезвычайно важный параметр для перцептивного отбора слов. Наконец, в итоговом представлении распознанного высказывания опять-таки слова представлены соответствующими словоформами.

Существует множество работ, в которых демонстрируется важность частотности в процессе восприятии речи. Но нам не известны работы, где использовалась бы частотность словоформ – напротив, все авторы практически игнорируют частотность отдельных словоформ, обращаясь исключительно к лексемам. Если полученные ими результаты не считать артефактами, приходится допустить, что носителю языка каким-то образом доступна информация о соотношении частотностей словоформ и словарной формы, т.е., фактически, лексемы. Причем такого рода переход от словоформы к лексеме, конечно, невозможно объяснить естественным знанием соответствующей парадигмы, поскольку информация о частотности должна использоваться до окончательной идентификации слова, иначе она просто теряет смысл.

По первичным статистическим характеристикам можно определить с заданной относительной погрешностью ту часть словника, в которую входят слова с высокой частотой появления независимо от типа текста. Возможно также, введя ступенчатое упорядочение в словарь, получить серию словников, охватывающих первые 100, 1000, 5000 и т. д. частых слов. Статистические характеристики словаря вызывают интерес в связи со смысловым анализом лексики. Изучение предметно-идеологическнх групп и семантических полей показывает, что лексические объединения поддерживаются семантическими связями, которые концентрируются вокруг лексем с наиболее общим значением. Описание значений в пределах лексико-семантического поля может проводиться посредством идентификации слов с наиболее абстрактными по смыслу лексемами. По-видимому, «пустые» (с точки зрения номинативных потенций) единицы словаря составляют статистически однородный пласт.

Не меньшую ценность имеют и словники по отдельным жанрам. Изучение меры их сходства и характера статистических распределений даст интересные сведения о качественном расслоении лексики в зависимости от сферы речеупотребления.

Составление больших частотных словарей требует обращения к вычислительной технике. Введение частичной механизации и автоматизации в процесс работы над словарем представляет интерес как эксперимент машинной обработки словников к разным текстам. Такой словарь требует более строгой системы обработки и накопления словарного материала. В миниатюре это информационно-поисковая система, которая способна выдавать сведения о различных сторонах текста и словаря. Некоторые основные запросы к этой системе планируются с самого начала: общее количество инвентаризованных слов, статистические характеристики отдельного слова и целых словников, упорядочение частых и редких зон словника и т. п. Машинная картотека позволяет автоматически строить обратные словари по отдельным жанрам и источникам. Множество других полезных статистических сведений о языке будет извлечено из накопленного массива информации. Компьютерный частотный словарь создает экспериментальную базу для перехода к более обширной автоматизации словарных работ.

Статистические данные частотных словарей могут быть широко использованы и при решении других лингвистических задач – например, при анализе и определении активных средств словообразования современного русского языка, решении вопросов усовершенствования графики и орфографии, которые связаны с учетом статистических сведений о словарном составе (при этом важно учитывать вероятностные характеристики комбинаций графем, реализованные в словах типы буквосочетаний), практической транскрипции и транслитерации. Статистические параметры словаря будут полезны и при решении вопросов автоматизации печатного дела, распознавания и автоматического чтения буквенного текста.

Современные толковые словари и грамматики русского языка в основном построены на базе литературно-художественных текстов. Существуют частотные словари языка А.С. Пушкина, А.С. Грибоедова, Ф.М. Достоевского, В.В. Высоцкого и многих других авторов. На кафедре истории и теории литературы Смоленского гос. педагогического университета ряд лет ведётся работа по составлению частотных словарей стихотворных и прозаических текстов. Для настоящего исследования отобраны частотные словари всей лирики Пушкина и ещё двух поэтов золотого века – «Горя от ума» Грибоедова и всей поэзии Лермонтова; Пастернака и ещё пяти поэтов серебряного века - Бальмонта 1894-1903 гг., «Стихов о Прекрасной Даме» Блока, «Камня» Мандельштама, «Огненного столпа» Гумилёва, «Anno Domini MCMXXI» Ахматовой и «Сестры моей жизни» Пастернака и ещё четырёх поэтов века железного – «Стихотворений Юрия Живаго», “Когда разгуляется”, всего корпуса лирики М. Петровых, «Дорога далека», «Ветрового стекла», «Прощания со снегом» и «Подковы» Межирова, «Антимиров» Вознесенского и «Снежницы» Рыленкова.

Следует отметить, что эти словари по природе своей различны: одни представляют лексику одного драматического произведения, другие – книги лирики, или нескольких книг, или всего корпуса стихов поэта. Результаты анализа, представленные в настоящей работе, следует воспринимать с осторожностью, их нельзя абсолютизировать. Однако с помощью специальных мер разницу онтологической природы текстов можно до известной степени уменьшить.

В последние годы все более отчетливо осознается противопоставление разговорной и книжной речи. Особенно остро обсуждается этот вопрос среди методистов, которые требуют поворота обучения в сторону разговорного языка. Однако специфика разговорной речи до сих пор остается необъясненной.

Обработка словарей выполнялась путем создания пользовательского приложения в среде офисной программы EXCEL97. Приложение включает четыре рабочих листа книги EXCEL – «Титульный лист», лист «Словари» с исходными данными, «Близости» и «Расстояния» с результатами , а также набор макросов.

Исходная информация вводится на лист «Словари». В ячейки EXCEL, записываются словари исследуемых текстов, последний столбец S формируется из полученных результатов и равен количеству слов, встречающихся в других словарях. Таблицы «Близости» и «Расстояния» содержат рассчитанные меры близости M, корреляции R и расстояния D.

Макросы приложения представляют собой событийные программные процедуры на языке Visual Basic for Application (VBA). Основу процедур составляют библиотечные объекты VBA и методы их обработки. Так, для операций с рабочими листами приложения используется ключевой объект Worksheet (рабочий лист) и соответствующий ему метод активизации листа Activate (активизировать). Задание диапазона анализируемых исходных данных на листе «Словари» выполняется методом Select (выбрать) объекта Range (диапазон), а передача слов в качестве значений переменным выполняется как свойство Value (значение) этого же объекта Range.

Несмотря на то, что ранговый корреляционный анализ заставляет с осторожностью говорить о зависимости тематики между разными текстами, большая часть самых частотных слов каждого текста имеет соответствия в одном или нескольких других текстах. В колонке S показано количество таких слов среди 15 наиболее частотных у каждого автора. Полужирным шрифтом выделены слова, встречающиеся в нашей таблице только у одного поэта. Нет выделенных слов вовсе у Блока, Ахматовой и Петровых, у них S = 15. У этих трёх поэтов все 15 самых частотных слов одни и те же, различаются они только местом в списке. Но даже у Пушкина, лексика которого наиболее оригинальна, S = 8, а выделенных слов 7.

Результаты показывают, что существует определённый слой лексики, концентрирующий основные темы поэзии. Как правило, эти слова коротки: из общего числа (225) словоупотреблений односложных 88, двусложных 127, трёхсложных 10. Зачастую эти слова представляют основные мифологемы и могут распадаться на пары: *ночь - день, земля - небо (солнце), Бог - человек (люди), жизнь - смерть, тело - душа, Рим - мир* (у Мандельштама); могут объединяться в мифологемы более высокого уровня: *небо, звезда, солнце, земля; в человеке как правило выделяются тело, сердце, кровь, рука, нога, щека, глаза*. Из человеческих состояний предпочтение отдаётся сну и любви. К миру человека принадлежат дом и города – Москва, Рим, Париж. Творчество представлено лексемами *слово* и *песня*.

У Грибоедова и Лермонтова среди наиболее частотных слов почти нет слов, обозначающих природу. У них втрое больше слов, обозначающих человека, части его тела, элементы его духовного мира. У Пушкина и поэтов ХХ в. обозначений человека и природы приблизительно поровну. В этом важном аспекте тематики, можем сказать, ХХ в. пошёл за Пушкиным.

Минимальная тема *дело* среди самых частотных слов встречается только у Грибоедова и Пушкина. У Лермонтова и поэтов ХХ в. она уступает место минимальной теме *слово*. Слово не исключает дела (библейская трактовка темы: в Новом Завете всё учение Иисуса Христа рассматривается как слово Божье или слово Иисуса, а апостолы иногда называют себя служителями Слова). Сакральный смысл лексемы слово убедительно проявляется, например, в стихе Пастернака «И образ мира, в Слове явленный». Сакральный смысл лексемы *слово* в со- и противопоставлении с человеческими делами убедительно проявляется в одноимённом стихотворении Гумилёва.

Лексемы, которые встречаются только в одном тексте, характеризуют своеобразие данной книги или совокупности книг. Например, слово «ум» - самое частотное в комедии Грибоедова «Горе от ума» – но оно не встречается среди частотных слов других текстов. Тема ума безусловно является наиболее значимой в комедии. Эта лексема сопровождает образ Чацкого, причём имя Чацкого является наиболее частотным в комедии. Таким образом, в произведении органически сочетаются самое частотное имя нарицательное с самым частотным именем собственным.

Самый высокий коэффициент корреляции связывает тематику трагических книг Гумилёва «Огненный столп» и Ахматовой «Anno Domini MCMXXI». Среди 15 наиболее частотных имён существительных здесь 10 общих, в том числе кровь, сердце, душа, любовь, слово, небо. Напомним, что в книгу Ахматовой вошла миниатюра «Не бывать тебе в живых...», написанная между арестом Гумилёва и его расстрелом.

Темы свечи и толпы в исследованном материале встречаются только в «Стихотворениях Юрия Живаго». Тема свечи в стихах из романа имеет множество контекстуальных значений: она связана с образом Иисуса Христа, с темами веры, бессмертия, творчества, любовного свидания. Свеча – важнейший источник света в центральных сценах романа. Тема толпы развивается в связи с основной идеей романа, в котором частная жизнь человека с её незыблемыми ценностями противопоставлена безнравственности нового государства, построенного на началах угождения толпе.

Работа предполагает и третий этап, тоже отраженный в программе, - это вычисление разности порядковых номеров слов, общих для двух словарей и среднего показателя расстояния между одинаковыми словами двух словарей. Этот этап позволяет от общих тенденций взаимодействия словарей, выявленных с помощью статистики, перейти на уровень, приближающийся к тексту. Например, статистически значимо коррелируют книги Гумилева и Ахматовой. Мы смотрим, какие слова оказались общими для их словарей, и выбираем прежде всего те, разница между порядковыми номерами которых минимальна или равна нулю. Именно эти слова имеют один и тот же ранговый номер и, следовательно, именно эти минимальные темы в сознании двух поэтов одинаково важны. Далее следует переходить на уровень текстов и контекстов.

Количественные методы также помогают изучить особенности народов – носителей языка. Скажем, в русском языке 6 падежей, в английском падежей нет, а в отдельных языках народов Дагестана количество падежей доходит до 40. Л.Перловский в своей статье «Сознание, язык и культура» соотносит эти характеристики со склонностью народов к индивидуализму или коллективизму [36, c.28], с восприятием вещей и явлений отдельно или в связи с другими. Ведь именно в англоязычном мире (падежей нет – вещь воспринимается «сама по себе») появились такие понятия, как свобода личности, либерализм и демократия (замечу, что я употребляю эти понятия только в связи с языком, без какой-либо оценочной характеристики). Несмотря на то, что подобные догадки пока остаются лишь на уровне смелых научных гипотез, они помогают посмотреть по-новому на уже знакомые явления.

Как мы видим, количественные характеристики могут применяться в совершенно различных областях языкознания, что всё больше стирает границы между «точными» и «гуманитарными» методами. Лингвистика всё чаще прибегает к помощи не только математики, но и вычислительной техники для решения своих задач.

**2.3 Изучение языка методами формальной логики**

С неколичественными методами математики, в частности, с логикой, современная теоретическая лингвистика взаимодействует не менее плодотворно, чем с количественными. Быстрое развитие компьютерных технологий и возрастание их роли в современном мире потребовало пересмотра подхода к взаимодействию языка и логики в целом.

Методы логики широко используются в разработке формализованных языков, в частности, языков программирования, элементами которых являются некоторые символы (сродни математическим), выбираемые (или конструируемые из выбранных ранее символов) и интерпретируемые определённым образом, связанным ни с каким «традиционным» употреблением, пониманием и функциями таких же символов в других контекстах. Программист постоянно имеет дело с логикой в своей работе. Смысл программирования состоит как раз в том, чтобы научить компьютер рассуждать (в широком смысле слова). При этом методы "рассуждения" оказываются самыми разными. Каждый программист тратит определенное время на поиск ошибок в своих и чужих программах. То есть, на поиск ошибок в рассуждениях, в логике. И это тоже накладывает свой отпечаток. Гораздо легче обнаруживаешь логические ошибки и в обычной речи. Относительная простота языков, изучаемых логиками, позволяет им выяснять структуры этих языков более четко, чем это достижимо для лингвистов, анализирующих исключительно сложные естественные языки. Ввиду того, что языки, изучаемые логиками, используют отношения, скопированные с естественных языков, логики способны внести существенный вклад в общую теорию языка. Ситуация здесь подобна той, которая имеет место в физике: физик также формулирует теоремы для идеально упрощенных случаев, которые не происходят в природе вообще - он формулирует законы для идеальных газов, идеальных жидкостей, говорит о движении при отсутствии трения и т.д. Для этих идеализированных случаев можно установить простые законы, которые значительно способствовали бы пониманию того, что происходит в действительности и что, вероятно, осталось бы неизвестным физике, если бы она пробовала рассматривать действительность непосредственно, во всей сложности.

В изучении естественных языков логические методы используются для того, чтобы изучающие язык могли не тупо «зазубрить» как можно больше слов, а лучше понять его структуру. Ещё Л. Щерба использовал на своих лекциях пример предложения, построенного по законам русского языка: «Глокая куздра штеко будланула бокра и курдячит бокренка», – а потом спрашивал у студентов, что это значит. Несмотря на то, что смысл слов в предложении оставался непонятен (их просто не существует в русском языке), можно было чётко ответить: «куздра» – подлежащее, существительное женского рода, в единственном числе, именительном падеже, «бокр» – одушевлённое, и т.д. Перевод фразы получается примерно таков: «Нечто женского рода в один прием совершило что-то над каким-то существом мужского рода, а потом начало что-то такое вытворять длительное, постепенное с его детенышем» [46]. Подобным же примером текста (художественного) из несуществующих слов, построенного полностью по законам языка, является «Бармаглот» Льюиса Кэрролла (в «Алисе в стране чудес» Кэрролл устами своего персонажа Шалтая-Болтая объясняет и значение придуманных им слов: «варкалось» – восемь часов вечера, когда уже пора варить ужин, «хливкий» – хлипкий и ловкий, «шорёк» – помесь хорька, барсука и штопора, «пыряться» – прыгать, нырять, вертеться, «нава» – трава под солнечными часами (простирается немного направо, немного налево и немного назад), «хрюкотать» – хрюкать и хохотать, «зелюк» – зелёный индюк, «мюмзик» – птица; перья у неё растрёпаны и торчат во все стороны, как веник, «мова» – далеко от дома) [46].

Одно из основных понятий современной логики и теоретической лингвистики, используемое при исследовании языков различных логико-математических исчислений, естественных языков, для описания отношений между языками различных «уровней» и для характеристики отношений между рассматриваемыми языками и описываемыми с их помощью предметными областями – понятие метаязыка. Метаязык – это язык, используемый для выражения суждений о другом языке, языке-объекте. С помощью метаязыка изучают структуру знакосочетаний (выражений) языка-объекта, доказывают теоремы о его выразительных свойствах, об отношении его к другим языкам и т. п. Изучаемый язык называется также предметным языком по отношению к данному метаязыку. Как предметный язык, так и метаязык могут быть обычными (естественными) языками. Метаязык может отличаться от языка-объекта (например, в учебнике английского языка для русских русский язык является метаязыком, а английский – языком-объектом), но может и совпадать с ним или отличаться лишь частично, например специальной терминологией (русская лингвистическая терминология – элемент метаязыка для описания русского языка; т. н. семантические множители – часть метаязыка описания семантики естественных языков).

Понятие «метаязык» стало весьма плодотворным в связи с изучением формализованных языков, строящихся в рамках математической логики. В отличие от формализованных предметных языков, в этом случае метаязык, средствами которого формулируется метатеория (изучающая свойства предметной теории, формулируемой на предметном языке), является, как правило, обычным естественным языком, некоторым специальным образом ограниченным фрагментом естественного языка, не содержащим всякого рода двусмысленностей, метафор, «метафизических» понятий и т. п. элементов обычного языка, препятствующих использованию его в качестве орудия точного научного исследования. При этом метаязык сам может быть формализован и (независимо от этого) оказаться предметом исследования, проводимого средствами метаметаязыка, причём такой ряд можно «мыслить» растущим бесконечно.

Логика учит нас плодотворному разграничению языка-объекта и метаязыка. Язык-объект - это сам предмет логического исследования, а метаязык - тот неизбежно искусственный язык, на котором такое исследование ведется. Логическое мышление как раз и состоит в том, чтобы сформулировать на языке символов (метаязыке) отношения и структуру реального языка (языка-объекта).

Метаязык должен быть во всяком случае «не беднее» своего предметного языка (т. е. для каждого выражения последнего в метаязыке должно иметься его имя- «перевод») – иначе, при невыполнении этих требований (что заведомо имеет место в естественных языках, если специальными соглашениями не предусмотрено обратное) возникают семантические парадоксы (антиномии).

По мере создания все новых и новых языков программирования в связи с проблемой программирования трансляторов появилась острая необходимость в создании метаязыков. В настоящее время наиболее употребительным для описания синтаксиса языков программирования является метаязык форм Бэкуса-Наура (сокращенно БНФ). Он представляет собой компактную форму в виде некоторых формул, похожих на математические. Для каждого понятия языка существует единственная метаформула (нормальная формула). Она состоит из левой и правой частей. В левой части указывается определяемое понятие, а в правой - задается множество допустимых конструкций языка, которые объединяются в это понятие. В формуле используют специальные метасимволы в виде угловых скобок, в которых заключено определяемое понятие (в левой части формулы) или ранее определенное понятие (в ее правой части), а разделение левой и правой частей указывается метасимволом "::=", смысл которого эквивалентен словам "по определению есть". Металингвистические формулы в некотором виде заложены в трансляторы; с их помощью ведется проверка конструкций, используемых программистом, на формальное соответствие какой-нибудь из конструкций, синтаксически допустимых в этом языке. Существуют и отдельные метаязыки различных наук – таким образом, знания существуют в виде различных метаязыков.

Логические методы также послужили основой для создания систем искусственного интеллекта, основанных на концепции коннекционизма. Коннекционизм – это особое течение в философской науке, предметом которого являются вопросы познания. В рамках этого течения предпринимаются попытки объяснить интеллектуальные способности человека, используя искусственные нейронные сети. Составленные из большого числа структурных единиц, аналогичных нейронам, с заданным для каждого элемента весом, определяющим силу связи с другими элементами, нейронные сети представляют собой упрощённые модели человеческого мозга. Эксперименты с нейронными сетями подобного рода продемонстрировали их способность к обучению выполнения таких задач, как распознавание образов, чтение и определение простых грамматических структур.

Философы начали проявлять интерес к коннекционизму, так как коннекционистский подход обещал обеспечить альтернативу классической теории разума и широко распространённой в рамках этой теории идеи, согласно которой механизмы работы разума имеют сходство с обработкой символического языка цифровым компьютером. Эта концепция весьма спорна, однако в последние годы она находит всё больше сторонников.

Логическое изучение языка продолжает соссюровскую концепцию о языке как системе. То, что оно постоянно продолжается, ещё раз подтверждает смелость научных догадок начала прошлого века. Последний раздел своей работы я посвящу перспективам развития математических методов в лингвистике в наши дни.

**2.4 Перспективы применения математических методов в лингвистике**

В эпоху компьютерных технологий методы математической лингвистики получили новую перспективу развития. Поиск решения проблем лингвистического анализа все активнее реализуется теперь на уровне информационных систем. Вместе с тем автоматизация процесса обработки языкового материала, предоставляя исследователю значительные возможности и преимущества, неизбежно выдвигает перед ним новые требования и задачи.

Соединение «точного» и «гуманитарного» знания стало плодородной почвой для новых открытий в области лингвистики, информатики и философии.

Машинный перевод с одного языка на другой остаётся быстро развивающейся отраслью информационных технологий. Несмотря на то, что перевод при помощи компьютера никогда не сравнится по качеству с переводом, сделанным человеком (особенно это касается художественных текстов), машина стала неотъемлемым помощником человека в переводе больших объёмов текста. Считается, что в ближайшем будущем будут созданы более совершенные переводческие системы, основанные, в первую очередь, на семантическом анализе текста.

Не менее перспективным направлением остаётся взаимодействие лингвистики и логики, служащее философским фундаментом для осмысления информационных технологий и так называемой «виртуальной реальности». В ближайшем будущем продолжится работа над созданием систем искусственного интеллекта – хотя, опять же, он никогда не будет равен человеческому по его возможностям. Подобная конкуренция бессмысленна: в наше время машина должна стать (и становится) не соперником, а помощником человека, не чем–то из области фантастики, а частью реального мира.

Продолжается изучение языка методами статистики, что позволяет более точно определить его качественные свойства. Важно, чтобы наиболее смелые гипотезы о языке находили своё математическое, а, следовательно, и логическое, доказательство.

Наиболее значимо то, что различные отрасли применения математики в лингвистике, до этого достаточно разрозненные, в последние годы соотносятся между собой, соединяясь в стройную систему, по аналогии с системой языка, открытой столетие назад Фердинандом де Соссюром и Иваном Бодуэном де Куртенэ. В этом – преемственность научного знания.

Лингвистика в современном мире стала фундаментом для развития информационных технологий. Пока информатика остаётся бурно развивающейся отраслью человеческой деятельности, союз математики и лингвистики продолжит играть свою роль в развитии науки.

Заключение

За ХХ век компьютерные технологии проделали большой путь – от военного применения к мирному, от узкого круга целей до проникновения во все отрасли человеческой жизни. Математика как наука находила всё новое практическое значение с развитием вычислительной техники. Этот процесс продолжается и сегодня.

Немыслимый раньше «тандем» «физиков» и «лириков» стал реальностью. Для полноценного взаимодействия математики и информатики с гуманитарными науками потребовались квалифицированные специалисты как с той, так и с другой стороны. В то время как специалистам-компьютерщикам всё более нужны систематические гуманитарные знания (лингвистические, культурологические, философские) , чтобы осмыслять изменения в окружающей их реальности, во взаимодействии человека и техники, разрабатывать всё новые и новые языковые и мыслительные концепции, писать программы, то любой «гуманитарий» в наше время для своего профессионального роста должен овладеть хотя бы азами работы с компьютером.

Математика, будучи тесно взаимосвязанной с информатикой, продолжает развиваться и взаимодействовать с естественнонаучным и гуманитарным знанием. В новом веке не ослабевает, а , наоборот, усиливается тенденция к математизации науки. На количественных данных осмысливаются закономерности развития языка, его исторические и философские характеристики.

Математический формализм более всего подходит для описания закономерностей в лингвистике (как , впрочем, и в других науках – и гуманитарных, и естественных). Ситуация порой складывается в науке так, что без применения соответствующего математического языка понять характер физического, химического и т.п. процесса невозможно. Создавая планетарную модель атома, известный английский физик XX в. Э. Резерфорд испытал математические трудности. Вначале его теорию не приняли: она не звучала доказательно, и виной тому явилось незнание Резерфордом теории вероятности, на основе механизма которой только и возможно было понять модельное представление атомных взаимодействий. Осознав это, выдающийся уже к тому времени ученый, обладатель Нобелевской премии , записался в семинар математика профессора Лэмба и в течение двух лет вместе со студентами прослушал курс и отработал практикум по теории вероятности. На ее основе Резерфорд смог описать поведение электрона, придав своей структурной модели убедительную точность и получив признание. То же – и с языкознанием.

Напрашивается вопрос, что же содержится в объективных явлениях такое математическое, благодаря чему они и поддаются описанию на языке математики, на языке количественных характеристик? Это однородные единицы вещества, распределяемые в пространстве и времени. Те науки, которые дальше других прошли путь к выделению однородности, и оказываются лучше приспособленными для использования в них математики.

Стремительно развившаяся в 90е годы сеть Интернет объединила под собой представителей различных стран, народов и культур. Несмотря на то, что основным языком международного общения продолжает служить английский, Интернет в наше время стал многоязычным. Это обусловило развитие коммерчески успешных систем машинного перевода, широко использующихся в различных областях человеческой деятельности.

Компьютерные сети стали объектом философского осмысления – создавались всё новые лингвистические, логические, мировоззренческие концепции, помогающие понять «виртуальную реальность». Во многих художественных произведениях создавались сценарии – чаще пессимистические – о господстве машин над человеком, а виртуальной реаьности – над окружающим миром. Далеко не всегда подобные прогнозы оказывались бессмысленными. Информационные технологии – не только перспективная отрасль вложения человеческих знаний, это ещё и способ контроля над информацией, а, следовательно, и над человеческой мыслью.

У этого явления есть как отрицательная, так и положительная сторона. Отрицательная – потому что контроль над информацией противоречит неотъемлемому человеческому праву на свободный доступ к ней. Положительная – потому что отсутствие этого контроля может привести к катастрофическим последствиям для человечества. Достаточно вспомнить один из наиболее мудрых фильмов последнего десятилетия – «Когда наступит конец света» Вима Вендерса, герои которого полностью погрузились в «виртуальную реальность» собственных снов, записываемых на компьютер. Однако ни один учёный и ни один художник не может дать однозначного ответа на вопрос: что же ждёт науку и технику в будущем.

Ориентировка на «будущее», порой кажущееся фантастическим, была отличительной особенностью науки середины ХХ века, когда изобретатели стремились создать совершенные образцы техники, которые могут работать без вмешательства человека. Время показало утопичность подобных изысканий. Однако было бы излишним осуждать учёных за это – без их энтузиазма в 1950е – 60е информационные технологии не сделали бы столь мощного скачка в 90е, и мы бы не имели того, что имеем сейчас.

Последние десятилетия ХХ века изменили приоритеты науки – исследовательский, изобретательский пафос уступил место коммерческому интересу. Опять же – это не хорошо и не плохо. Это – реальность, в которой наука оказывается всё более интегрированной в повседневную жизнь.

Наступивший XXI век продолжил эту тенденцию, и в наше время за изобретениями стоят не только слава и признание, но, в первую очередь, деньги. Ещё и поэтому важно заботиться о том, чтобы новейшие достижения науки и техники не попали в руки террористических группировок или диктаторских режимов. Задача сложная до невозможности; максимально осуществить её – задача всего мирового сообщества.

Информация – оружие, причём оружие не менее опасное, чем ядерное или химическое – только действует оно не физически, а, скорее, психологически. Человечеству надо задуматься о том, что для него в этом случае важнее – свобода или контроль.

Новейшие философские концепции, связанные с развитием информационных технологий и попыткой их осмыслить, показали ограниченность как естественнонаучного материализма, господствовавшего на протяжении ХIХ – начала ХХ веков, так и крайнего идеализма, отрицающего значимость материального мира. Современной мысли, особенно мысли Запада, важно преодолеть этот дуализм в мышлении, когда окружающий мир чётко делится на материальное и идеальное. Путь к этому – диалог культур, сопоставление разных точек зрения на окружающие явления.

Как ни парадоксально, информационные технологии могут сыграть не последнюю роль в этом процессе. Компьютерные сети, и особенно Интернет – это не только ресурс для развлечения и бурной коммерческой деятельности, это ещё и средство осмысленного, спорного общения между представителями различных цивилизаций в современном мире, а также для диалога прошлого с настоящим. Можно сказать, что Интернет раздвигает пространственные и временные рамки.

А в диалоге культур посредством информационных технологий по-прежнему важна роль языка как древнейшего универсального средства общения. Именно поэтому лингвистика во взаимодействии с математикой, философией и информатикой пережила своё второе рождение и продолжает развиваться поныне. Тенденция настоящего продолжится и в будущем – «until the end of the world», как 15 лет назад предсказывал всё тот же В. Вендерс. Правда, неизвестно, когда произойдёт этот конец – но важно ли это сейчас, ведь будущее рано или поздно всё равно станет настоящим.

Приложение 1

Ferdinand de Saussure

The Swiss linguist Ferdinand de Saussure (1857-1913) is widely considered to be the founder of modern linguistics in its attempts to describe the structure of language rather than the history of particular languages and language forms. In fact, the method of Structuralism in linguistics and literary studies and a significant branch of Semiotics find their major starting point in his work at the turn of the twentieth century. It has even been argued that the complex of strategies and conceptions that has come to be called "poststructuralism" – the work of Jacques Derrida, Michel Foucault, Jacques Lacan, Julia Kristeva, Roland Barthes, and others – is suggested by Saussure's work in linguistics and anagrammatic readings of late Latin poetry. If this is so, it can be seen most clearly in the way that Saussure's work in linguistics and interpretation participates in transformations in modes of understanding across a wide range of intellectual disciplines from physics to literary modernism to psychoanalysis and philosophy in the early twentieth century. As Algirdas Julien Greimas and Joseph Courtés argue in Semiotics and Language: An Analytic Dictionary, under the heading "Interpretation," a new mode of interpretation arose in the early twentieth century which they identify with Saussurean linguistics, Husserlian Phenomenology, and Freudian psychoanalysis. In this mode, "interpretation is no longer a matter of attributing a given content to a form which would otherwise lack one; rather, it is a paraphrase which formulates in another fashion the equivalent content of a signifying element within a given semiotic system" (159). In this understanding of "interpretation," form and content are not distinct; rather, every "form" is, alternatively, a semantic "content" as well, a "signifying form," so that interpretation offers an analogical paraphrase of something that already signifies within some other system of signification.

Such a reinterpretation of form and understanding – which Claude Lévi-Strauss describes in one of his most programmatic articulations of the concept of structuralism, in "Structure and Form: Reflections on a Work by Vladimir Propp" – is implicit in Saussure's posthumous Course in General Linguistics (1916, trans., 1959, 1983). In his lifetime, Saussure published relatively little, and his major work, the Course, was the transcription by his students of several courses in general linguistics he offered in 1907-11. In the Course Saussure called for the "scientific" study of language as opposed to the work in historical linguistics that had been done in the nineteenth century. That work is one of the great achievements of Western intellect: taking particular words as the building blocks of language, historical (or "diachronic") linguistics traced the origin and development of Western languages from a putative common language source, first an "Indo-European" language and then an earlier "proto-Indo-European" language.

It is precisely this study of the unique occurrences of words, with the concomitant assumption that the basic "unit" of language is, in fact, the positive existence of these "word-elements," that Saussure questioned. His work was an attempt to reduce the mass of facts about language, studied so minutely by historical linguistics, to a manageable number of propositions. The "comparative school" of nineteenth-century Philology, Saussure says in the Course, "did not succeed in setting up the true science of linguistics" because "it failed to seek out the nature of its object of study" ([1959] 3). That "nature," he argues, is to be found not simply in the "elemental" words that a language comprises – the seeming "positive" facts (or "substances") of language – but in the formal relationships that give rise to those "substances."

Saussure's systematic reexamination of language is based upon three assumptions. The first is that the scientific study of language needs to develop and study the system rather than the history of linguistic phenomena. For this reason, he distinguishes between the particular occurrences of language – its particular "speech-events," which he designates as parole – and the proper object of linguistics, the system (or "code") governing those events, which he designates as langue. Such a systematic study, moreover, calls for a "synchronic" conception of the relationship among the elements of language at a particular instant rather than the "diachronic" study of the development of language through history.

This assumption gave rise to what Roman Jakobson in 1929 came to designate as "structuralism," in which "any set of phenomena examined by contemporary science is treated not as a mechanical agglomeration but as a structural whole [in which] the mechanical conception of processes yields to the question of their function" ("Romantic" 711). In this passage Jakobson is articulating Saussure's intention to define linguistics as a scientific system as opposed to a simple, "mechanical" accounting of historical accidents. Along with this, moreover, Jakobson is also describing the second foundational assumption in Saussurean – we can now call it "structural" – linguistics: that the basic elements of language can only be studied in relation to their functions rather than in relation to their causes. Instead of studying particular and unique events and entities (i.e., the history of particular Indo-European "words"), those events and entities have to be situated within a systemic framework in which they are related to other so-called events and entities. This is a radical reorientation in conceiving of experience and phenomena, one whose importance the philosopher Ernst Cassirer has compared to "the new science of Galileo which in the seventeenth century changed our whole concept of the physical world" (cited in Culler, Pursuit 24). This change, as Greimas and Courtés note, reconceives "interpretation" and thus reconceives explanation and understanding themselves. Instead of explanation's being in terms of a phenomenon's causes, so that, as an "effect," it is in some ways subordinate to its causes, explanation here consists in subordinating a phenomenon to its future-oriented "function" or "purpose." Explanation is no longer independent of human intentions or purposes (even though those intentions can be impersonal, communal, or, in Freudian terms, "unconscious").

In his linguistics Saussure accomplishes this transformation specifically in the redefinition of the linguistic "word," which he describes as the linguistic "sign" and defines in functionalist terms. The sign, he argues, is the union of "a concept and a sound image," which he called "signified [signifié] and signifier [signifiant]" (66-67; Roy Harris's 1983 translation offers the terms "signification" and "signal" [67]). The nature of their "combination" is "functional" in that neither the signified nor the signifier is the "cause" of the other; rather, "each [derives] its values from the other" (8). In this way, Saussure defines the basic element of language, the sign, relationally and makes the basic assumption of historical linguistics, namely, the identity of the elemental units of language and signification (i.e., "words"), subject to rigorous analysis. The reason we can recognize different occurrences of the word "tree" as the "same" word is not because the word is defined by inherent qualities – it is not a "mechanical agglomeration" of such qualities – but because it is defined as an element in a system, the "structural whole," of language.

Such a relational (or "diacritical") definition of an entity governs the conception of all the elements of language in structural linguistics. This is clearest in the most impressive achievement of Saussurean linguistics, the development of the concepts of the "phonemes" and "distinctive features" of language. Phonemes are the smallest articulated and signifying units of a language. They are not the sounds that occur in language but the "sound images" Saussure mentions, which are apprehended by speakers – phenomenally apprehended – as conveying meaning. (Thus, Elmar Holenstein describes Jakobson's linguistics, which follows Saussure in important ways, as "phenomenological structuralism.") It is for this reason that the leading spokesperson for Prague School Structuralism, Jan Mukarovsky, noted in 1937 that "structure . . . is a phenomenological and not an empirical reality; it is not the work itself, but a set of functional relationships which are located in the consciousness of a collective (generation, milieu, etc.)" (cited in Galan 35). Similarly, Lévi-Strauss, the leading spokesperson for French structuralism, noted in 1960 that "structure has no distinct content; it is content itself, and the logical organization in which it is arrested [or apprehended] is conceived as a property of the real" (167; see also Jakobson, Fundamentals 27-28).

Phonemes, then, the smallest perceptible elements of language, are not positive objects but a "phenomenological reality." In English, for instance, the phoneme /t/ can be pronounced in many different ways, but in all cases an English speaker will recognize it as functioning as a /t/. An aspirated t (i.e., a t pronounced with an h-like breath after it), a high-pitched or low-pitched t sound, an extended t sound, and so on, will all function in the same manner in distinguishing the meaning of "to" and "do" in English. Moreover, the differences between languages are such that phonological variations in one language can constitute distinct phonemes in another; thus, English distinguishes between /l/ and /r/, whereas other languages are so structured that these articulations are considered variations of the same phoneme (like the aspirated and unaspirated t in English). In every natural language, the vast number of possible words is a combination of a small number of phonemes. English, for instance, possesses less than 40 phonemes that combine to form over a million different words.

The phonemes of language are themselves systematically organized structures of features. In the 1920s and 1930s, following Saussure's lead, Jakobson and N. S. Trubetzkoy isolated the "distinctive features" of phonemes. These features are based upon the physiological structure of the speech organs – tongue, teeth, vocal chords, and so on – that Saussure mentions in the Course and that Harris describes as "physiological phonetics" ([1983] 39; Baskin's earlier translation uses the term "phonology" [(1959) 38]) – and they combine in "bundles" of binary oppositions to form phonemes. For instance, in English the difference between /t/ and /d/ is the presence or absence of "voice" (the engagement of the vocal chords), and on the level of voicing these phonemes reciprocally define one another. In this way, phonology is a specific example of a general rule of language described by Saussure: In language there are only differences. Even more important: a difference generally implies positive terms between which the difference is set up; but in language there are only differences without positive terms. Whether we take the signified or the signifier, language has neither ideas nor sounds that existed before the linguistic system. ([1959] 120)

In this framework, linguistic identities are determined not by inherent qualities but by systemic ("structural") relationships.

I have said that phonology "followed the lead" of Saussure, because even though his analysis of the physiology of language production "would nowadays," as Harris says, "be called 'physical,' as opposed to either 'psychological' or 'functional'" (Reading 49), nevertheless in the Course he articulated the direction and outlines of a functional analysis of language. Similarly, his only extended published work, Mémoire sur le système primitif des voyelles dans les langues indo-européennes (Memoir on the primitive system of vowels in Indo-European languages), which appeared in 1878, was fully situated within the project of nineteenth-century historical linguistics. Nevertheless, within this work, as Jonathan Culler has argued, Saussure demonstrated "the fecundity of thinking of language as a system of purely relational items, even when working at the task of historical reconstruction" (Saussure 66). By analyzing the systematic structural relationships among phonemes to account for patterns of vowel alternation in existing Indo-European languages, Saussure suggested that in addition to several different phonemes /a/, there must have been another phoneme that could be described formally. "What makes Saussure's work so very impressive," Culler concludes, "is the fact that nearly fifty years later, when cuneiform Hittite was discovered and deciphered, it was found to contain a phoneme, written h, which behaved as Saussure had predicted. He had discovered, by a purely formal analysis, what are now known as the laryngeals of Indo-European" (66).

This conception of the relational or diacritical determination of the elements of signification, which is both implicit and explicit in the Course, suggests a third assumption governing structural linguistics, what Saussure calls "the arbitrary nature of the sign." By this he means that the relationship between the signifier and signified in language is never necessary (or "motivated"): one could just as easily find the sound signifier arbre as the signifier tree to unite with the concept 'tree'. But more than this, it means that the signified is arbitrary as well: one could as easily define the concept 'tree' by its woody quality (which would exclude palm trees) as by its size (which excludes the "low woody plants" we call shrubs). This should make clear that the numbering of assumptions I have been presenting does not represent an order of priority: each assumption – the systemic nature of signification (best apprehended by studying language "synchronically"), the relational or "diacritical" nature of the elements of signification, the arbitrary nature of signs – derives its value from the others.

That is, Saussurean linguistics understands the phenomena it studies in overarching relationships of combination and contrast in language. In this conception, language is both the process of articulating meaning (signification) and its product (communication), and these two functions of language are neither identical nor fully congruent (see Schleifer, "Deconstruction"). Here, we can see the alternation between form and content that Greimas and Courtés describe in modernist interpretation: language presents contrasts that formally define its units, and these units combine on succeeding levels to create the signifying content. Since the elements of language are arbitrary, moreover, neither contrast nor combination can be said to be basic. Thus, in language distinctive features combine to form contrasting phonemes on another level of apprehension, phonemes combine to form contrasting morphemes, morphemes combine to form words, words combine to form sentences, and so on. In each instance, the whole phoneme, or word, or sentence, and so on, is greater than the sum of its parts (just as water, H2O, in Saussure's example [(1959) 103] is more than the mechanical agglomeration of hydrogen and oxygen).

The three assumptions of the Course in General Linguistics led Saussure to call for a new science of the twentieth century that would go beyond linguistic science to study "the life of signs within society." Saussure named this science "semiology (from Greek semeîon 'sign')" (16). The "science" of semiotics, as it came to be practiced in Eastern Europe in the 1920s and 1930s and Paris in the 1950s and 1960s, widened the study of language and linguistic structures to literary artifacts constituted (or articulated) by those structures. Throughout the late part of his career, moreover, even while he was offering the courses in general linguistics, Saussure pursued his own "semiotic" analysis of late Latin poetry in an attempt to discover deliberately concealed anagrams of proper names. The method of study was in many ways the opposite of the functional rationalism of his linguistic analyses: it attempted, as Saussure mentions in one of the 99 notebooks in which he pursued this study, to examine systematically the problem of "chance," which "becomes the inevitable foundation of everything" (cited in Starobinski 101). Such a study, as Saussure himself says, focuses on "the material fact" of chance and meaning (cited 101), so that the "theme-word" whose anagram Saussure is seeking, as Jean Starobinski argues, "is, for the poet, an instrument, and not a vital germ of the poem. The poem is obliged to re-employ the phonic materials of the theme-word" (45). In this analysis, Starobinski says, "Saussure did not lose himself in a search for hidden meanings." Instead, his work seems to demonstrate a desire to evade all the problems arising from consciousness: "Since poetry is not only realized in words but is something born from words, it escapes the arbitrary control of consciousness to depend solely on a kind of linguistic legality" (121).

That is, Saussure's attempt to discover proper names in late Latin poetry – what Tzvetan Todorov calls the reduction of a "word . . . to its signifier" (266) – emphasizes one of the elements that governed his linguistic analysis, the arbitrary nature of the sign. (It also emphasizes the formal nature of Saussurean linguistics – "Language," he asserts, "is a form and not a substance" [Course (1959) 122] – which effectively eliminates semantics as a major object of analysis.) As Todorov concludes, Saussure's work appears remarkably homogeneous today in its refusal to accept symbolic phenomena [phenomena that have intentional meaning]. . . . In his research on anagrams, he pays attention only to the phenomena of repetition, not to those of evocation. . . . In his studies of the Nibelungen, he recognizes symbols only in order to attribute them to mistaken readings: since they are not intentional, symbols do not exist. Finally in his courses on general linguistics, he contemplates the existence of semiology, and thus of signs other than linguistic ones; but this affirmation is at once limited by the fact that semiology is devoted to a single type of sign: those which are arbitrary. (269-70)

If this is true, it is because Saussure could not conceive of "intention" without a subject; he could not quite escape the opposition between form and content his work did so much to call into question. Instead, he resorted to "linguistic legality." Situated between, on the one hand, nineteenth-century conceptions of history, subjectivity, and the mode of causal interpretation governed by these conceptions and, on the other hand, twentieth-century "structuralist" conceptions of what Lévi-Strauss called "Kantianism without a transcendental subject" (cited in Connerton 23) – conceptions that erase the opposition between form and content (or subject and object) and the hierarchy of foreground and background in full-blown structuralism, psychoanalysis, and even quantum mechanics – the work of Ferdinand de Saussure in linguistics and semiotics circumscribes a signal moment in the study of meaning and culture.

Ronald Schleifer

Приложение 2

Фердинанд де Соссюр (перевод)

Швейцарский языковед Фердинанд де Соссюр (1857-1913) считается основателем современной лингвистики – благодаря своим попыткам описать структуру языка, а не историю отдельных языков и словоформ. По большому счёту, основы структурных методов в лингвистике и литературоведении и, в значительной мере, семиотики были заложены в его работах в самом начале двадцатого века. Доказано, что методы и концепции так называемого "постструктурализма", развитые в работах Жака Деррида, Мишеля Фуко, Жака Лакана, Юлии Кристевой, Ролана Барта и других, восходят к лингвистическим трудам Соссюра и анаграмматическим прочтениям поздней римской поэзии. Следует заметить, что работы Соссюра по лингвистике и языковой интерпретации помогает связать широкий круг интеллектуальных дисциплин – от физики до литературных новшеств, психоанализа и философии начала двадцатого века. А. Дж. Греймас и Ж. Курте пишут в «Семиотике и языке»: «Аналитический словарь с заголовком «Интерпретация» как новый вид интерпретации появился в начале ХХ века вместе с лингвистикой Соссюра, феноменологией Гуссерля и психоанализом Фрейда. В таком случае, "интерпретация – это не приписывание данного содержания к форме, которая иначе испытала бы недостаток в том; скорее это - пересказ, который формулирует другим способом то же содержание значимого элемента в пределах данной семиотической системы" (159). В таком понимании «интерпретации», форма и содержание неразрывны; напротив, каждая форма наполнена семантическим значением («значимая форма»), поэтому интерпретация предлагает новый, аналогичный пересказ чего-то, значимого в другой знаковой системе.

Подобное понимание формы и содержания, представляемое Клодом Леви-Строссом в одной из программных работ структурализма, ("Структура и Форма: Размышления над трудами Владимира Проппа") – можно увидеть в посмертно вышедшей книге Соссюра «Курс общей лингвистики» (1916, пер., 1959, 1983). При жизни Соссюр мало публиковался, «Курс» – его основная работа – был собран по конспектам студентов, посещавших его лекции по общей лингвистике в 1907-11 гг. В «Курсе» Соссюр призывал к «научному» исследованию языка, противопоставляя его сравнительно-историческому языкознанию девятнадцатого века. Эту работу можно считать одним из величайших достижений западной мысли: беря за основу отдельные слова как структурные элементы языка, историческое (или «диахроническое») языкознание доказывало происхождение и развитие западноевропейских языков от общего, индоевропейского языка– и более раннего праиндоевропейского.

Это - точно это исследование уникальных возникновений слов, с сопутствующим предположением, что основная "единица" языка, фактически, положительное существование этих "элементов слова", что Соссюр подверг сомнению. Его работа была попыткой сократить множество фактов о языке, вскользь изученных сравнительной лингвистикой, до небольшого числа теорем. Сравнительная филологическая школа XIX века, пишет Соссюр, «не преуспела в создании настоящей школы лингвистики», так как «она не поняла сущности объекта изучения» ([1959] 3). Эта «сущность», утверждает он, заключается не только в отдельных словах – «позитивных субстанциях» языка – но и в формальных связях, помогающих этим субстанциям существовать.

Соссюровская «проверка» языка основана на трёх предположениях. Первое: научное понимание языка основано не на историческом, а на структурном феномене. Поэтому он различал отдельные явления языка –«события речи», которые он определяет как «parole» – и надлежащий, по его мнению, объект изучения лингвистики, систему (код, структуру), управляющую этими событиями ( «langue»). Подобное систематическое изучение, кроме того, требует «синхронной» концепции отношений между элементами языка в данный момент, а не «диахронического» исследования развития языка через его историю.

Эта гипотеза стала предтечей того, что Роман Якобсон в 1929 назовёт «структурализмом» – теории, где "любой набор явлений, исследованный современной наукой, рассматривается не как механическое скопление, а как структурное целое, в котором конструктивная составляющая соотносится с функцией" ("Romantic" 711). В этом отрывке Якобсон сформулировал соссюровскую идею определения языка как структуры в противовес «машинальному» перечислению исторических событий. Кроме того, Якобсон развивает и другое соссюровское предположение, ставшее предтечей стркутурной лингвистики: базовые элементы языка должны изучаться в связи не столько со своими причинами, сколько со своими функциями. Отдельные явления и события (скажем, история происхождения отдельных индоевропейских слов) должны изучаться не сами по себе, а в системе, в котрой они соотнесены с подобными же составляющими. Это был радикальный поворот в сопоставлении явлений с окружающей действительностью, значимость которого философ Эрнст Кассирер сравнил с «наукой Галилея, перевернувшей в семнадцатом веке представления о материальном мире". Такой поворот, как замечают Греймас и Курте, меняет представление о «интерпретации», а, седовательно, и сами объяснения. Явления стали трактоваться не относительно причин их возникновения, а относительно того эффекта, который они могут оакзать в настоящем и будущем. Толкование перестало быть независимым от намерений человека(несмотря на то, что намерения могут быть безличными, «бессознательными» во фрейдистском понимании этого слова).

В своей лингвистике Соссюр особенно показывает этот поворот в изменении понятия слова в лингвистике, которое он определяет как знак и описывает с точки зрения его функций. Знак для него –соединение звучания и смысла, «обозначаемого [signifié] и обозначения [signifiant]» (66-67; в английском переводе 1983 года авторства Роя Харриса – «signification» и "signal" [67]). Природа этого соединения – «функциональная» (ни тот, ни другой элемент не могут существовать друг без друга); более того, "одно заимствует качества у другого" (8). Таким образом Соссюр определяет основной структурный элемент языка – знак – и делает основой исторического языкознания идентичность знаков словам, что требует особо строгого анализа. Поэтому мы можем понять разные значения, скажем, одного и того же слова «дерево» – не потому что слово представляет собой лишь набор определённых качеств, а потому что оно определено как элемент в знаковой системе, в «структурном целом», в языке.

Подобное относительное («диакритическое») понятие единства лежит в основе представления о всех элементах языка в структурной лингвистике. Это особенно ясно видно в наиболее оригинальной находкесоссюровского языкознания, в развитии концепции «фонем» и «отличительных особенностей» языка. Фонемы – самые мелкие из произносимых и значимых языковых единиц. Они являются не только звуками, встречающимися в языке, но «звуковыми образами», замечает Соссюр, которые воспринимаются носителями языка как обладающие значением. (Следует заметить, что Элмар Холенштейн называет лингвистику Якобсона, по основным положениям продолжающего идеи и концепции Соссюра, «феноменологическим структурализмом»). Именно поэтому ведущий докладчик пражской школы структурализма, Ян Мукаровский, заметил в 1937 году, что «структура . . . не эмпирическое, а феноменологическое понятие; это не сам результат, а набор значимых отношений коллективного сознания (поколения, окружающих и т.д.)». Похожую мысль высказал в 1960 году Леви-Стросс, лидер французского структурализма: «У структуры нет определённого содержания; она сама по себе содержательна, и логическая конструкция, в которую она заключена, представляет собой отпечаток реальности».

В свою очередь, фонемы, как самые мелкие языковые элементы, приемлемые для восприятия, представляют собой отдельную цельную «феноменологическую реальность». Например, в английском языке звук «т» может произноситься по-разному, но во всех случаях человек, владеющий английским, будет воспринимать его как «т». Произнесённый с придыханием, с высоким или низким подъёмом языка, долгий звук «т» и т.п будет одинаково различать значение слов «to» и «do». Более того – различия между языками таковы, что разновидности одного звука в одном языке могут соответствовать разным фонемам в другом; например «л» и «р» в английском различны, в то время как в других языках это разновидности одной фонемы (подобно английскому «т», произнесённому с придыханием и без). Обширный словарный запас любого естественного языка представляет собой набор сочетаний гораздо меньшего количества фонем. В английском, например, для произнесения и написания около миллиона слов используется всего 40 фонем.

Звуки языка представляют собой системно организованный набор черт. В 1920е –1930е, вслед за Соссюром, Якобсон и Н.С.Трубецкой выделили «отличительные черты» фонем. Эти черты основаны на строении органов речи – языка, зубов, голосовых связок – Соссюр замечает это в «Курсе общей лингвистики», а Харрис называет «физиологической фонетикой» (в более раннем переводе Баскина используется термин «фонология») – они соединяются в «узлы» дург против друга, чтобы издавать звуки. Скажем, in в английском разница между «т» и «д» заключается в наличии или отсутствии «голоса» (напряжении голосовых связок), и в уровне голоса, отличающем одну фонему от другой. Таким образом, фонологию можно считать примером общего языкового праивла, описанного Соссюром: «В языке есть только различия». Более важно даже не это: различие обычно подразумевает точные условия, между которыми оно и находится; но в языке существуют только различия без точных условий. Рассматриваем ли мы «обозначение» или «обозначаемое» – в языке не существует ни понятий, ни звуков, которые существовали бы до того, как развилась языковая система.

В подобной структуре, языковые аналогии определены не с помощью изначально присущих им качеств, а с помощью системных («структурных») отношений.

Я уже упомянул, что фонология в своём развитии опиралась на идеи Соссюра. Несмотря на то, что его анализ языковой физиологии в наше время по словам Харриса, «был бы назван «физическим», в противовес «психологическому» или «функциональному», в «Курсе» он отчётливо сформулировал направление и основные принципы функционального анализа языка. Его единственная изданная при жизни работа, «Mémoire sur le système primitif des voyelles dans les langues indo-européennes» (Записки о первоначальной системе гласных в индоевропейских языках), изданная в 1878, полностью находилась в русле сравнительно-исторического языкознания XIX века. Тем не менее этим трудом, как говорит Джонатан Каллер, Соссюр показал «плодотворность представления о языке как о системе взаимосвязанных явлений, даже при его исторической реконструкции». Анализируя взаимосвязи между фонемами, объясняя чередования гласных в современных языках индоевропейской группы, Соссюр предположил, что кроме нескольких разных звуков «а», должны быть и другие фонемы, описываемые формально. «Что производит особое впечатление в труде Соссюра, – делает вывод Каллер, – то, что почти 50 лет спустя, при открытии и расшифровке хеттской клинописи, была найдена фонема, на письме обозначаемая «h», которая вела себя так, как предсказывал Соссюр. С помощью формального анализа он открыл то, что сейчас известно как гортанный звук в индоевропейских языках.

В концепция относительного (диакритического) определения знаков, как явно выраженной, так и подразумеваемой в «Курсе», существует и третье ключевое предположение структурной лингвистики, названное Соссюром «произвольной природой знака». Под этим подразумевается, что отношение между звучанием и значением в языке ничем не мотивировано: с одинаковой лёгкостью можно соединить слово «arbre» и слово «tree» с понятием «дерево». Более того, это значит, что звучание тоже произвольно: можно определить понятие «дерево» по наличию у него коры (кроме пальм) и по размеру (кроме «низких древесных растений» - кустарников). Из этого должно быть понятно, что все представляемые мною предположения не делятся на более и менее важные: каждое из них – системный характер знаков (более всего понятный при «синхронном» изучении языка), их относительная (диакритическая) сущность, произвольная природа знаков – исходит из остальных.

Таким образом, в соссюровском языкознании изучаемый феномен понимается как свод сопоставлений и противопоставлений языка. Язык – это и выражение значения слов (обозначение), и их результат (общение) – и эти две функции никогда не совпадают (см. «Деконструкцию языка» Шлейфера). Мы можем заметить чередование формы и содержания, которое Греймас и Курте описывают в новейшем варианте интерпретации: языковые контрасты определяют его структурные единицы, и эти единицы взаимодействуют на сменяющих друг друга уровнях, чтобы создать определённое значимое содержание. Так как элементы языка случайны, ни контраст, ни сочетание не могут быть основой. Значит, в языке отличительные признаки формируют фонетический контраст на другом уровне понимания, фонемы соединяются в контрастные морфемы, морфемы – в слова, слова – в предложения и т.д. В любом случае, целая фонема, слово, предложение и т.д. представляет собой нечто большее, чем сумма составляющих (так же как вода, в соссюровском примере – больше, чем сочетание водорода и кислорода).

Три предположения «Курса общей лингвистики» привели Соссюра к идее новой науки двадцатого века, отдельно от лингвистики изучающей «жизнь знаков в обществе». Соссюр назвал эту науку семиологией (от греческого «semeîon» - знак). «Наука» семиотики, развивавшаяся в Восточной Европе в 1920е –1930е и в Париже в 1950е and 1960е, расширила изучение языка и лингвистических структур до литературных находок, составленных (или сформулированных) с помощью этих структур. Кроме того, на закате своей карьеры, параллельно своему курсу общей лингвистики, Соссюр занялся «семиотическим» анализом поздней римской поэзии, пытаясь открыть умышленно составленные анаграммы имён собственных. Этот метод был во многом противоположен рационализму в его лингвистическом анализе: он был попыткой, как пишет Соссюр в одной из 99 записных книжек, изучить в системе проблему «вероятности», которая «становится основой всего». Такое исследование, как утверждает сам Соссюр, помогает сосредоточиться на «вещественной стороне» вероятности; «ключевое слово», анаграмму которого ищет Соссюр, как утверждает Жан Старобинский, «инструмент для поэта, а не источник жизни стихотворения. Стихотворение служит для того, чтобы поменять местами звуки ключевого слова». По словам Старобинского, в этом анализе «Соссюр не углубляется в поиски скрытых смыслов». Напротив, в его работах заметно желание избежать вопросов, связанных с сознанием: «так как поэзия выражается не только в словах, но и в том, что порождают эти слова, она выходит из-под контроля сознания и зависит только от законов языка».

Попытка Соссюра изучить имена собственные в поздней римской поэзии (Цветан Тодоров назвал это сокращением «слова... лишь до его написания») подчёркивает одну из составляющих его лингвистического анализа – произвольную природу знаков, а также формальную сущность соссюровской лингвистики ( «Язык, – утверждает он, «суть форма, а не явление»), что исключает возможность анализа смысла. Тодоров делает вывод, что в наши дни труды Соссюра выглядят на редкость последовательными в нежелании изучать символы [явления, имеющие чётко определённое значение]. . . . Исследуя анаграммы, Соссюр обращает внимание только на повторение, но не на предшествующие варианты. . . . Изучая «Песнь о Нибелунгах», он определяет символы только для того, чтобы присвоить их ошибочным чтениям: если они неумышленны, символов не существует. В конце концов, в своих трудах по общей лингвистике он делает предположение о существовании семиологии, описывающей не только лингвистические знаки; но это предположение ограничивается тем, что семилогоия может описывать только случайные, произвольные знаки.

Раз это действительно так, то только потому, что не мог представить «намерение» без предмета; он не мог до конца преодолеть пропасть между формой и содержанием – в его трудах это превращалось в вопрос. Вместо этого он обращался к «языковой законности». Находясь между, с одной стороны, концепциями девятнадцатого века, основанными на истории и субъективных догадках, и методах случайной интерпретации, основанных на этих концепциях, и, с другой стороны, структуралистскими концепциями, которые Леви-Стросс назвал «кантианством без трансцендентного действующего лица» – стирающими противостояние между формой и содержанием (субъектом и объектом), значением и происхождением в структурализме, психоанализе и даже квантовой механике – труды Ферлинанда де Соссюра по лингвистике и семиотике обозначают поворотный момент в изучении значений в языке и культуре.

# Рональд Шлейфер

Литература

1. Адмони В.Г. Основы теории грамматики / В.Г. Адмони; АН СССР.-М.: Наука, 1964.-104с.

2. Апресян, Ю.Д. Идеи и методы современной структурной лингвистики. М., 1966.

3. Арапов, М.В., Херц, М.М. Математические методы в лингвистике. М., 1974.

4. Арнольд И.В. Семантическая структура слова в современном английском языке и методика её исследования. /И.В. Арнольд– Л.: Просвещение, 1966. – 187 с.

5. Баранов А.Н. Категории искусственного интеллекта в лингвистической семантике. Фреймы и сценарии// А.Н. Баранов – М., 1987 – 300 c.

6.Башлыков А.М. Система автоматизированного перевода. / А.М. Башлыков, А.А. Соколов. – М.: ООО «ФИМА», 1997. – 20 с.

7.Бодуэн де Куртенэ: Теоретическое насле дие и современность: Тезисы докладов международной научной конференции / Ред.И.Г. Кондратьева. – Казань: КГУ, 1995. – 224 с.

8. Гладкий А.В., Элементы математической лигвистики. / . Гладкий А.В., Мельчук И.А. –М., 1969. – 198 с.

9. Головин, Б.Н. Язык и статистика. /Б.Н. Головин –М., 1971. – 210 с.

10. Звегинцев, В.А. Теоретическая и прикладная лингвистика. / В.А. Звегинцев –М., 1969. – 143 с.

11. Касевич, В.Б. Семантика. Синтаксис. Морфология. // В.Б. Касевич –М., 1988. – 292 c.

12. Лекомцев Ю.К. Введение в формальный язык лингвистики/ Ю.К. Лекомцев. – М.: Наука, 1983, 204 с., ил.

13. Лингвистическое наследие Бодуэна де Куртенэ на исходе ХХ столетия: Тезисы докладов международской научно-практическтй конференции 15-18 марта 2000 года. – Красноярск, 2000. – 125 с.

Матвеева Г.Г. Скрытые грамматические значения и идентификация социального лица («портрета») говорящего/ Г.Г. Матвеева. – Ростов, 1999. – 174 с.

14. Мельчук, И.A. Опыт постpоения лингвистических моделей "Смысл <--> Текст"./ И.А. Мельчук. – М., 1974. – 145 c.

15. Нелюбин Л.Л. Перевод и прикладная лингвистика/Л.Л. Нелюбин. – М. : Высшая школа, 1983. – 207 с.

16. О точных методах исследования языка: о так называемой «математической лингвистике»/ О.С. Ахманова, И. А. Мельчук, Е.В. Падучева и др. – М., 1961. – 162 с.

17. Пиотровский Л.Г. Математическая лингвистика: Учебное пособие/ Л.Г. Пиотровский, К.Б. Бектаев, А.А. Пиотровская. – М.: Высшая школа, 1977. – 160 с.

18.Он же . Текст, машина, человек. – Л., 1975. – 213 с.

19. Он же. Прикладное языкознание / Под ред. А.С Герда. – Л., 1986. – 176 с.

20. Ревзин, И.И. Модели языка. М., 1963. Ревзин, И.И. Современная структурная лингвистика. Проблемы и методы. М., 1977. – 239 с.

21. Ревзин, И.И., Розенцвейг, В.Ю. Основы общего и машинного перевода/Ревзин И.И., Розенцвейг, В.Ю. – М., 1964. – 401 с.

22. Слюсарёва Н.А. Теория Ф.де Соссюра в свете современной лингвистики/ Н.А. Слюсарева. – М.:Наука, 1975. – 156 с.

23. Сова, Л.З. Аналитическая лингвистика/ Л.З. Сова – М., 1970. – 192 с.

24. Соссюр Ф. де. Заметки по общей лингвистике/ Ф. де Соссюр; Пер. с фр. – М.: Прогресс, 2000. – 187 с.

25. Он же. Курс общей лингвистики/ Пер. с фр. – Екатеринбург, 1999. –426 с.

26. Статистика речи и автоматический анализ текста / Отв. ред. Р.Г. Пиотровский. Л., 1980. – 223 с.

27. Столл, P. Множествa. Логикa. Aксиомaтические теоpии./ Р. Столл; Пер. с англ. – М., 1968. – 180 с.

28. Теньер, Л. Основы структурного синтаксиса. М., 1988.

29. Убин И.И. Автоматизация переводческой деятельности в СССР/ И.И. Убин, Л.Ю. Коростелёв, Б.Д. Тихомиров. – М., 1989. – 28 с.

30. Фоp, P. , Кофмaн, A., Дени-Пaпен, М. Совpеменнaя мaтемaтикa. М., 1966.

31. Шенк, Р. Обработка концептуальной информации. М., 1980.

32. Шихaнович, Ю.A. Введение в современную мaтемaтику (нaчaльные понятия). М., 1965

33. Щерба Л.В. Русские гласные в качественном и количественном отношении/ Л.В. Щерба – Л.: Наука, 1983. – 159 с.

34. Абдулла-заде Ф. Гражданин мира// Огонёк – 1996. – №5. – С.13

35. В.А. Успенский. Предварение для читателей «Нового литературного обозрения» к семиотическим посланиям Андрея Николаевича Колмогорова. – Новое литературное обозрение. –1997. – № 24. – С. 18-23

36. Перловский Л. Сознание, язык и культура. – Знание – сила. –2000. №4 – С. 20-33

37. Фрумкина Р.М. О нас – наискосок. //Русский Журнал. – 2000. – №1. – С. 12

38. Фитиалов, С.Я. О моделировании синтаксиса в структурной лингвистике // Проблемы структурной лингвистики. М., 1962.

39. Он же. Об эквивалентности грамматики НС и грамматики зависимостей // Проблемы структурной лингвистики. М., 1967.

40. Хомский, Н. Логические основы лингвистической теории // Новое в лингвистике. Вып. 4. М., 1965

41. Schleifer R. Ferdinand de Saussure// press. jhu.ru

42. www.krugosvet.ru

43. www.lenta.ru

44. phil.ru.ru

45. press. jhu.ru

46. ru.wikipedia.org

47. www.smolensk.ru