

Язык, мышление и современные системы распознавания речи

А.Н.Соловьев

Санкт-Петербургский Государственный Университет

Лучшим кажется то, к чему он имеет охоту.

Козьма Прутков

Охота бывает разная: кому охота, а кому и неохота.

Ком.взвода ВДВ

Что такое язык? Как язык соотносится с мышлением? Возможно ли учитывать когнитивные механизмы в современных системах распознавания речи?

Доклад точно с таким же названием был озвучен на второй международной конференции по когнитивным наукам, прошедшим в июне месяце 2006 года в Санкт-Петербургском Университете. Инициатором этого доклада был Валерий Иванович Галунов. К сожалению, в декабре 2005 случилось трагическое событие: Валерия Ивановича не стало...

После доклада, во время дискуссии мне был задан вопрос: «А как Вы думаете, возможно ли сделать систему, которая была бы неотличима от человеческого восприятия / понимания?». Я ответил, что, разумеется, нет, но ведь и нет предела совершенству и нашим стремлениям, то есть охота есть...

В настоящем докладе я хотел бы более подробно, более развернуто поговорить на эту тему.

В настоящее время рабочей гипотезой нейролингвистов является то, что наш мозг представляет собой не хранилище неких символов, в котором происходит сравнение входящего сигнала с эталонными данными, а сложную динамическую систему, в которой вся информация заключена во взаимосвязях и развертывании динамических процессов в каждый момент времени [Ламб С. 2003]. Как показывают многочисленные эксперименты (MEG, fMRI и др.), чем сложнее воспринимаемый испытуемым входной сигнал, тем сложнее и разнообразнее протекают процессы в мозге человека. Например, при исследовании акустического или фонетического восприятия достаточно четко прослеживается категоричность восприятия в зависимости от входного сигнала. При исследовании семантической или синтаксической структуры речи пока все достаточно сложно и неоднозначно интерпретируемо. Маппирование коры мозга при изучении творческой деятельности (при решении различных тестовых заданий) показывает, что в

этом процессе участвуют большое количество различных областей мозга: «...физиологические корреляты самых разных видов психической активности могут быть обнаружены почти в каждой точке мозга» [Н.Бехтерева. 2006].

Из многочисленных исследований известно, что одним из главных управляющих механизмов мозга является детектор ошибок. Он является своего рода управляющим, отвечающим не только за правильность принятия решений и адекватности реакции на раздражители, но и ответственным за продуктивность творческих процессов.

В принципе даже если предположить, что детектор ошибок – один из главных механизмов задающих доминанту (доминирование тех или иных физиологических процессов в мозге, приводящих к проявлению индивидуальных характеристик, проявлению личности), то, не смотря на знание местоположения и некоторых принципах работы этого механизма, пока совершенно не понятно причины его «поведения».

Таким образом, в настоящее время с помощью современных прецизионных приборов мы можем подробно узнать, как происходят те или иные когнитивные процессы в нашем мозге (например, техника MEG (магнитоэнцефалография) дает прекрасное временное разрешение (до миллисекунд) активности нейронов коры головного мозга, а техника fMRI (ядерный магнитный резонанс) точное пространственное разрешение), но все же не можем понять, как работает наш мозг, как происходит процесс понимания и осознания действительности. Изучая разные результаты экспериментов создается впечатление того, что мозг представляет собой оркестр, «местоположение дирижера которого неизвестно и нестабильно, а возможно и не заполнено вообще, так как оркестр самоорганизуется с учетом множества факторов» [Черниговская, 2006].

Возможно учитывать когнитивные процессы происходящие при понимании речи в современных системах распознавания речи? Рассмотрим некоторые современные модели автоматического распознавания речи.

Автоматические системы распознавания речи, как правило, состоят из двух основных модулей: 1) так называемый нижний уровень распознавания – собственно само распознавание акустических событий и 2) верхний уровень или семантический, отвечающий за понимание речи.

В настоящее время большинство существующих систем автоматического распознавания речи используют в семантическом модуле вероятностные модели (аналогично как и на нижнем уровне). Это так называемые Марковские цепи, строящиеся

на статистических языковых особенностях, где каждое последующее звено цепи стохастически зависит от предыдущего. При расчете условной вероятности последующего события, учитывая 2-4 предыдущих события и модель называют 2 или 4 граммной соответственно. Так же к этой группе относятся вероятностные модели, основанные на деревьях решений. Этот класс моделей использует деревья решений для оценки распределения вероятностей очередного слова по известной истории. Под деревом решений понимается бинарное дерево, каждой листовой вершине которого приписывается распределение вероятностей на словаре, а остальным вершинам приписываются предикаты, определённые на множестве историй.

Другие, менее распространенные в виду своей трудоемкости модели, это модели, использующие развитый аппарат теории формальных языков для представления лингвистической информации. При этом подходе естественный язык описывается при помощи систем правил, обычно представленных в виде некоего метаязыка. Обычно правила для таких языковых моделей строятся «вручную» исследователем, что сопряжено со значительными трудностями. Однако точность такой модели оказывается существенно выше, чем точность вероятностных грамматик, типа n-грамм.

Этот класс моделей (вероятностные и основанные теории формальных языков) можно назвать статичными моделями, так как их обучение происходит на конкретно заданной обучающей выборке и появление новых слов или изменение значений приводит к резкому увеличению ошибок в процессе машинного понимания. Не останавливаясь на других недостатках этих моделей, можно сказать, что это главный недостаток моделей понимания.

Существует другой класс моделей понимания: так называемые фреймовые модели (или их подмножество – ситуационные модели), реализованные на нейросетевых алгоритмах и использующие реляционные базы знаний. Вообще, понятие фрейма и фреймовой модели пришло в лингвистику из психологии, где так и осталось на умозрительном описании. Попытки практически использовать фреймы было сложно из-за необходимости обязательного наличия некой базы знаний о мире, участвующей в работе долговременной и кратковременной памяти. В настоящее время все большую популярность находят реляционные базы знаний, суть которых – существование неких внутренних латентных связей, которые можно использовать в зависимости от запроса. Для их успешной реализации в моделях понимания используют нейросетевые алгоритмы. Эти базы представляют собой сложные системы, имеющие скрытые структуры, с помощью которых можно интерпретировать тот или иной фрейм. Например, такая база

может содержать некие правила, описывающие ситуацию, ассоциативные или семантические связи и пр.

Такой класс моделей понимания более перспективен, ибо реализует неявные взаимосвязи между элементами базы знаний. Появление новой информации приводит не к остановке алгоритма, а поиску ассоциативных связей за счет исследования контекста и может привести к образованию нового фрейма.

Один из примеров таких моделей – это использование латентно-семантического анализа для представления знаний и обработки входящей информации. В основе метода лежат принципы факторного анализа, в частности, выявление латентной связей изучаемых явлений или объектов. Он основывается на идее, что совокупность всех контекстов, в которых встречается и не встречается данная лексическая единица, задает множество обоюдных ограничений, которые в значительной степени позволяют определить близость смысловых значений слов (ассоциативность, синонимичность).

Но, несмотря на современные достижения в современных речевых технологиях, остается один существенный и совершенно незатронутый вопрос: как учитывать в автоматических системах индивидуальное восприятие и понимание. Преобладание того или иного состояния приводит к разному пониманию действительности. Одна и та же фраза может вызвать совершенно разные ассоциации, и, соответственно, реакцию.

А.А. Ухтомский говорил, что понимание – это не обмен знаками и не передача информации, а активно уважительное постижение доминант другого [Ухтомский. А. 2002]. Знаки возникают при нашем восприятии семиозиса. Задача коммуникантов – проявление доминирующего знака языковыми средствами у адресанта и выявление доминирующего знака у адресата. Лингвистические аспекты выполняют вторичную или вспомогательную роль при понимании, а семантический уровень предопределяется доминантой. Понимание – это не только нейрофизиологическая реакция речевых отделов мозга на внешние стимулы, но и функция состояния нейрофизиологической модели в каждый момент времени.

Таким образом, пока рано еще говорить даже о возможности создания систем коммуникации подобных человеческой. Практически все современные автоматические системы распознавания строятся на статистической информации о языковых единицах, где принципиально невозможно ввести каким-либо образом понятие доминанты.

«Вероятно, следует возлагать надежды не на ещё большее усложнение разрешающей способности техники, а на методологический и даже философский прорыв,

который должен привести к возникновению новой мульти-дисциплинарной научной парадигмы» [Черниговская Т. В., 2006].

Возможно, это должен быть не только «методологический и даже философский прорыв», а и еще, что более важно, другой взгляд на такие вещи как сознание, наблюдение, восприятие.

Наше знание ограничено тем семиозисом (более обще - семиосферой), который возникает при нашем восприятии (не важно чего: внешнего мира, внутреннего – дело всего лишь в смене кода, но не смене «обозначенности»). Увеличение знания – всего лишь иллюзия. Знание не увеличивается (по объему), происходит более мелкое «дробление» семиозиса, с возникновением все более мелкозернистой структуры. Мысль, которая сейчас в моей голове уже когда-то была и будет в следующий момент в другом месте, другой голове. Все зависит от моего наблюдения и ее интерпретации мной. Но что такое интерпретация? Это сравнение данного текущего элемента (полученного дифференсацией сегмента) со сходными единицами, которые могли бы его замещать, как по сходству, так и по смежности, входя в ту же парадигму, что и он.

Вот тут мы подходим к понятию языка.

Язык – механизм отражения функции состояния мозга в каждый момент времени.

Моя интерпретация некой мысли языковыми средствами зависит от лично моего восприятия действительности. Но от чего же зависит восприятие? Восприятие обусловлено сложными когнитивными процессами, происходящими в моем мозгу. Образно можно представить, что мозг является своего рода измерительным прибором, одним концом стрелки указывающей на внешний (внутренний – без разницы) мир (раздробленное семиотическое пространство), а другой на нечто, что пока мы назовем сознанием. То, что заставляет его определенным образом реагировать на нечто.

Чтобы понять, что это такое, нам нужно это померить, то есть провести процесс наблюдения. Но мы не можем наблюдать наше мышление. Я говорю не о физиологических процессах или нашей интерпретации нами нашего мышления – языке (внутреннем или озвученном). Наблюдая мышление, мы вносим в него изменения тем, что мы называем восприятием. Этот эффект хорошо известен в квантовой механике: любое измерение приводит к изменению квантового состояния. Например, известно, что эффект интерференции возможен от одной элементарной частицы, и в какой бы из отверстий решетки мы измерили наличие частицы (если быть точным – квантовую вероятность), то ее там и обнаружим. Частица как бы находится одновременно во всех отверстиях решетки

одновременно. «Как бы», потому что наше понимание ограничено и не может этого понять. Это лишь один из множества примеров квантовой механики. Тем не менее, не смотря на «непредставимость» квантового мира, ученые нашли законы, достаточно хорошо описывающие этот мир. Это значит, не представляя и не понимая микромира, мы можем говорить о его свойствах, описывать неким выбранным метаязыком.

Аналогичное рассуждение применимо к тому, что выше было определено как сознание. Это не сознание «я» или «он», это нечто надопытное, метафизичное в философской терминологии. То, что наблюдает мозг, который наблюдает нечто. Оно (то что было названо сознанием) проявляет себя в языке в момент его (языка) реализации. И это то, с чем уже можно работать.

Разумеется, пока эти спекулятивные рассуждения малонаучны, ибо научная ненаблюдаемость того, что здесь подразумевалось под сознанием, не позволяет прийти к нему логическими доводами. В связи с этим возникает очень сложный вопрос о методе исследования такого объекта, о его наблюдении и мышлении о таком объекте. Сложность здесь заключается в том, что это нечто, что здесь названо сознанием – «объект такого мышления, в котором нет мыслящего об этом объекте» [Пятигорский А. 2002]. А это приводит к тому, что необходим совершенно иной подход (не психологический или биологический, и скорее всего не философский) к изучению такого рода надопытных, выходящих за рамки нашего понимания объектов.

ЛИТЕРАТУРА

1. Ламб С. О нейрокогнитивной лингвистике. Лекции по когнитивным наукам. Под ред. Соловьева В.Д. Казань. Вып.6. 2002.
2. Бехтерева Н.П. Магия творчества и психофизиология. СПб. 2006.
3. Черниговская Т.В. Зеркальный мозг, концепты и язык: цена антропогенеза. Физиологический журнал им. И.М.Сеченова, 2006, т.92, №1, с. 84-99.
4. Ухтомский. А.А., Доминанта, СПб., 2002.
5. Пятигорский А., Мышление и наблюдение. Рига, 2002.