ВОПРОСЫ ЯЗЫКОЗНАНИЯ

№ 2 1993

КРИТИКА И БИБЛИОГРАФИЯ

ОБЗОРЫ

© 1993 г. ПИОТРОВСКИЙ Р.Г., ПОПЕСКУЛ А.Н., СОВПЕЛЬ И.В.

КАК СТРОИТСЯ И РАБОТАЕТ ЛИНГВИСТИЧЕСКИЙ АВТОМАТ

1. Информационно-семиотические основы построения лингвистического автомата.

Исследовательская группа "Статистика речи" (СтР), основные коллективы которой работают в Санкт-Петербурге, Минске, Кишиневе, а также в Средней Азии [1], разрабатывает в течение более чем тридцати лет лингвистическую проблематику искусственного интеллекта [2]. Одновременно в группе создавались экспериментальные и промышленные системы автоматической переработки текста (АПТ)[1]. Логика этих теоретических и прикладных исследований привела СтР в середине 70-х годов к идее многоцелевого лингвистического автомата (ЛА)[2], представляющего собой реально функционирующую модель речемыслительной деятельности человека. На Западе многофункциональные системы этого типа стали строиться лишь в конце 80-х годов [3—4].

Теоретической основой концепции ЛА и одновременно базой для построения образующих его инженерно-лингвистических модулей должна была стать некоторая семиотическая схема порождения, передачи и восприятия текста. Поиски этой схемы шли, с одной стороны, по линии проработки и адаптации к нуждам человеко-машинной коммуникации уже известных лингвистических, психолингвистических и когнитивных моделей [5, с. 221—237; 6, с. 187—200, 217, 250; 7, с. 124—128, 131—132], а с другой — по линии организации ориентированных на экологический подход самостоятельных исследований в области психолингвистики и психиатрического языкознания [8].

В итоге этих поисков была построена расширенная соссюрианская модель лингвистического знака [9, с. 22—23; 10, с. 6—29], на основе которой и была развернута та семиотическая схема, которая явилась психолингвистической основой для построения ЛА. Эта схема описывает формирование сообщения, начиная от его денотативного замысла (Dn_1), отражающего некоторый факт объективной действительности, через тематический десигнативный план, вплоть до

---

[1] В настоящее время группа СтР реализует на внутреннем и внешнем рынке системы англо-русского, русско-английского, французско-русского, испанско-русского, португальско-русского лексического и лексико-грамматического МП по различным общественно-политическим, деловым и научно-техническим тематикам. Частично реализована комбинированная система TUTSY, осуществляющая поддержку обучения переводу французских и английских текстов. Наконец, разрабатываются промышленные системы МП с восточных языков.

[2] Под лингвистическим автоматом понимается комплекс вычислительных и программных средств, объединяющий: а) достаточно мощную специализированную или универсальную ЭВМ, снабженную необходимой для оперативной переработки текста периферией, например, читающим устройством (сканнером) и автоматическим редактором (спеллером); б) лингвистическую информационную базу данных ЛИБД, включающую большие словари и "сильные" грамматики; в) лингвистическое программное обеспечение (ЛПО); 2) набор системных и сервисных средств, представляемых операционной средой, вместе с программами ведения ЛИБД и ЛПО.

его лексико-грамматического и графемно-фонетического кодирования. Все развертывание сообщения происходит под управлением коммуникативно-прагматического оператора — КПО [11, с. 108], который обеспечивая выбор необходимой информации из тезауруса (θ) и лингвистической компетенции (ЛК), регулирует переход с одного уровня на другой при формировании сообщения. Что касается восприятия и расшифровки сообщения, то СтР в своих исследованиях ориентируется на две схемы. Согласно первой гипотетической схеме [12], принятые абонентом звуковые или визуальные (графические) сигналы сопоставляются с заложенными в ЛК сенсорными (фонемно-фонетическими или графемными) образцами. Если такое сопоставление дает положительный результат, включается поверхностный лексико-грамматический анализ предложения, а также составляющих его с/с и с/у (словосочетаний и словоупотреблений). Затем на десигнативном уровне производится глубинный рема-тематический анализ, опирающийся на семантико-синтаксическую информацию, извлекаемую из энциклопедического словаря, ЛК, и анализа контекста. Наконец, на заключительном денотативном уровне происходит обобщающая интерпретация информации, извлеченной из сообщения на предыдущих уровнях. Эти операции, осуществляющиеся абонентом исходя из его личной прагматики, пресуппозиции и предварительного знакомства с ситуацией, должны привести абонента к формированию денотата (Dn_2), т.е. обобщенного симультанного образа факта, о котором повествует принятое сообщение.

Равенство $Dn_1 = Dn_2$ указывает на то, что сообщение понято собеседником в точном соответствии с начальным замыслом отправителя. При $Dn_1 \neq Dn_2$ расшифровка сообщения неадекватна смыслу, вложенному в него автором.

Согласно второй гипотезе [6, с. 217—250], поиск Dn_2 идет уже в рамках сенсорного и лексико-грамматического декодирования сообщения. На этом начальном этапе идет выявление ключевых вех высказывания (отдельных слов, словосочетаний, простых семантико-синтаксических схем). Сам поиск осуществляется абонентом на основе его прагматической установки и ожидания, включающих ориентировку в референциальной среде и пресуппозицию. Затем формируются гипотезы о смысле принятого сообщения. После этого на базе прагматической установки абонента, его пресуппозиции (при необходимости с привлечением лексико-грамматического анализа и, наконец, путем сравнения получаемой информации с семантико-синтаксическими фреймами, заложенными в тезаурусе и ЛК) происходит выбор наиболее правдоподобной гипотезы, которая и кладется в основу денотата (смыслового образа) принятого сообщения. Все операции по распознаванию текста осуществляются под управлением КПО.

Приведенные схемы будут рассматриваться в качестве натурных объектов (оригиналов) для компьютерных моделей анализа и синтеза текста, объединенных в ЛА.

2. Лингвистический автомат.

Разработка лингвистической стратегии построения ЛА потребовала принятия ряда альтернативных решений.

Первое предусматривало выбор либо лексической, либо грамматической приоритетности в построении общего алгоритма ЛА. При решении этой задачи СтР руководствовалась двумя соображениями:

1) лексика, как показали информационные исследования текста, несет в себе львиную долю содержащейся в тексте информации [13, с. 169—190];

2) машинный анализ и синтез отдельных лексических единиц в гораздо меньшей степени, чем грамматический анализ входного и генерация синтаксической структуры выходного предложений, подвержены действию "генетических" парадоксов АПТ [14, с. 168—169]. Поэтому было решено начать построение ЛА не с разработки грамматических алгоритмов, как это делало большинство

неофитов АПТ, но с создания словарной базы ЛА и процедур смысловой обработки лексических единиц (ЛЕ) текста.

Второе решение было связано с выбором между хомскианской, жестко дедуктивной традицией и вероятностной функциональной грамматикой речи. Первая традиция сохраняется до сих пор в таких применяемых в современных системах МП формализмах, как грамматика управления и связывания (government and binding grammar) [15], грамматика присоединяемых деревьев (tree-adjoining grammar) [16] и грамматика фразовых структур (GPSG) [17, с. 521—587]. Основные положения второго подхода начали, как известно, вырисовываться уже в 60-х годах в работах Дж. Гринберга [18, с. 114—162], Ч. Филлмора [19, с. 369--469] и М. Холлидея [20, с. 179--215].

Информационно-статистические измерения текста [13, с. 120—207] настойчиво указывали на то, что порождение речи представляет собой сложный марковский процесс, в котором жесткое планирование возможно лишь на коротких участках текста, а между достаточно удаленными друг от друга лингвистическими единицами обнаруживаются лишь быстро затухающие стохастические связи. Поэтому если моделирование лексико-грамматической системы языка можно было вести на традиционной основе алгебры отношений и четких множеств, полученных путем приложения операции интенсификации к реальным нечетким лингвистическим множествам [21, с. 16--17], то при моделировании в ЛА процессов анализа и синтеза текста пришлось ориентироваться на функциональную лингвистику речи, опирающуюся на валентностные модели типовых ситуаций (фреймы), на вероятностные модели устранения неоднозначности и, наконец, на формальное распознание смыслового образа текста [14, с. 192—236]. Итак, лингвистическая стратегия СтР, использующаяся при построении ЛА и отличающая ее от большинства коллективов, занимающихся созданием систем АПТ, характеризуется приоритетом лексических разработок, адаптированных к задачам потребителя, вместе с ориентацией на вероятностную функциональную лингвистику речи.

2.1. Архитектура ЛА.

Прежде чем обратиться к описанию архитектуры ЛА, укажем на два ее конструктивных принципа, которые обусловлены только что описанной лингвистической стратегией.

Первый заключается в открытой стратификационной (модульно-уровневой) организации, предусматривающей, с одной стороны, возможность устранения из ЛА одних и включения других модулей, а с другой, соотнесенность каждого модуля с определенным уровнем порождения и восприятия сообщения.

Второй принцип состоит в постоянном взаимодействии человека и машины при конструировании, функционировании и совершенствовании ЛА.

Это значит, что при построении машинных словарей и грамматик, а также при "дообучении" ЛА должны, с одной стороны, использоваться не только традиционные "человеческие" знания о ЕЯ, но также широко применяться результаты машинного обследования больших массивов (корпусов) современных текстов определенного стиля или подъязыка. Каждый такой корпус виртуальных текстов (КВТ)[3] следует рассматривать в качестве базы знаний, на основе которой должна строиться функциональная машинная грамматика данной разновидности языка.

С другой стороны, различные формы переработки текста в ЛА должны осуществляться не только в независимом от человека пакетном режиме, но также в

---

[3] Развернутая концепция КВТ изложена в работе [22]. В ней описывается полный технологический цикл разработки и использования КВТ, от формулирования информационно-системных принципов его формирования и построения необходимых классификаторов для различных уровней ЕЯ до создания практических лингво-математических моделей и алгоритмов эффективного решения многих задач АПТ (рис. 3--4).

условиях человеко-машинного диалога, так называемом интерактивном режиме. Этот режим следует применять также при обучении и дообучении ЛА.

2.2. Две схемы представления ЛА.

ЛА является сложной системой. Поэтому для ее описания приходится использовать многоаспектное представление, включающее модели и схемы, построенные на аппаратных (hardware), системно-сервисных (software), лингвистических (lingware) и других подходах. Для нас наибольший интерес представляют две схемы описания — структурно-функциональная и схема управления и решений (децизивная).

2.2.1. Структурно-функциональное описание.

Это описание, строящееся в отвлечении от физического субстрата ЛА, представляет собой уровневую систему, имеющую следующие четыре плана (страта).

1. Нижний страт. Его занимает ЛИБД, которая, выполняя роль аналога тезауруса и лингвистической компетенции в речемыслительном аппарате человека, включает входные и выходные словники лексических единиц (основ, с/ф, исходных форм слова, словосочетаний), списки морфем и другой грамматический инвентарь.

2. Средний страт, который охватывает множество функциональных модулей, каждый из которых выполняет лингвистическое задание, моделирующее определенную функцию речемыслительной деятельности человека. Это множество распадается на два подмножества. Первое включает следующие анализирующие модули:

— модуль декодирования текста (d),
— модуль корректировки (c),
— модуль лексического анализа ключевых ЛЕ текста (l_k),
— модуль пословно-пооборотного (лексического) анализа всех ЛЕ текста (l),
— модуль автономного морфологического анализа с/у текста (q),
— модуль лексико-морфологического анализа ключевых ЛЕ текста (λ_k),
— модуль лексико-морфологического анализа всех ЛЕ текста (λ),
— модуль анализа поверхностной структуры текста (g),
— модуль анализа глубинной (тема-рематической) структуры текста (s_1),
— модуль семантико-прагматического анализа текста (s_2).

Второе подмножество включает такие синтезирующие модули, как:

— модуль графического или фонетического представления (кодирования) текста (k),
— модуль корректировки (c),
— модуль лексического синтеза (выбор из АС лексических эквивалентов для входных с/у и с/с) (l'),
— модуль автономного морфологического синтеза (q'),
— модуль лексико-морфологического синтеза с/у и с/с (λ'),
— модуль синтеза поверхностной структуры выходного текста (g'),
— модуль синтеза тема-рематической структуры выходного текста (s_1'),
— модуль синтеза семантико-прагматического образа текста (s_2').

При необходимости в ЛА могут включаться самые разнообразные модули, начиная с блоков вычисления темпов развития отдельных грамматических и лексических категорий и определения этимологии слова и с/с [23, 24] и кончая дидактическим блоком. В первом случае ЛА становится инструментом историко-лингвистических исследований, во втором он превращается в обучающий лингвистический автомат [25, с. 26].

3. Верхний страт. Его образует комплекс генерирующих программ-функций (F), которые, подобно КПО, должны генерироваться функциональными модулями среднего страта, а также той заложенной в ЛИБД лексической (L) и грам-

матической (G) информацией, которая необходима для переработки текста в одноязычной или двуязычной ситуации.

В настоящее время еще не удалось создать такие компонующие функции, которые могли бы построить полный ЛА, включая все указанные модули. Однако уже сейчас строятся "малые" ЛА, состоящие из части перечисленных модулей.

Так, например, с помощью функции $F_1(E)$ можно сформировать интерактивную систему коррекции и редактирования английского "дефектного" текста (спеллер)

$$Sp(E) = \{d, c, k, L, G\},$$

в которой пользователю предоставляется возможность самому выбрать правильную с его точки зрения форму из альтернативных с/ф, предлагаемых ЛА. Эта система может быть преобразована в автоматическую систему, работающую в пакетном режиме. В этом случае система выбирает среди реальных альтернатив "испорченного" с/у наиболее частотную словоформу.

Функция $F_2(P \rightarrow R)$ генерирует систему

$$S_2(P \rightarrow R) = \{d, \lambda, \lambda', k, L, G\},$$

осуществляющую первичный лексико-морфологический анализ португальских (P) информационных сообщений и их "грубый" русский (R) перевод.

Функция $F_3(F \rightarrow R)$ строит более сложную систему

$$S_3(F \rightarrow R) = \{d, lk, l, \lambda_k, \lambda, \lambda', k, L, G\},$$

производящую фрагментирование, составление поискового образа и "грубый" перевод на русский язык французского (F) патента, автоматически фрагментированного по рубрикам [2; 30, с. 141—154].

С помощью концептуально-сетевой (к/с) функции $F_8(R \rightarrow к/с \rightarrow G)$, выступающей в роли языка-посредника, строится система

$$S_8 (R \rightarrow к/с \rightarrow G) = \{d, \lambda, g, s_1, s_1', g', \lambda', k, L, G\},$$

превращающая русское (R) предложение в концептуальный граф [26, с. 22—29]. Затем на базе этого графа, представляющего глубинную семантико-синтаксическую структуру входного предложения, синтезируется иноязычный (немецкий — G) перевод.

Высший страт реализуется в настоящее время в форме человеко-машинного взаимодействия. Это взаимодействие можно условно считать аналогом мотива и отчасти КПО в схеме речемыслительной деятельности человека

2.2.2. Децизивная схема ЛА.

Подобно речемыслительной деятельности человека любая система АПТ всегда связана с операциями распознавания, которые осуществляются в условиях неопределенности. Эта неопределенность задается в машинном словаре и грамматике в виде множества альтернатив, из которых ЛА, обладающий элементами искусственного интеллекта, должен выбрать правильное решение. Поэтому архитектура ЛА должна описываться не только со структурно-функциональной, но и с точки зрения принятия решений, т.е. по децизивной схеме.

По аналогии с системами управления децизивную схему ЛА можно представить в виде иерархии следующих планов (стратов):

1) самоорганизации;

2) адаптации ЛА к обрабатываемым текстам;

3) выбора способа решения для поставленной задачи.

На первом уровне, обычно в режиме человеко-машинного общения, разрабатывается стратегия решения общей задачи $\bar{P}$. В соответствии с этой стратегией вырабатываются и компонуются те подсистемы и модули, которые необходимы автомату для решения этой задачи.

Обращаясь к описанию второго уровня, следует иметь в виду, что решая лингвистическую задачу, ЛА обычно оказывается в условиях неопределенности, порождаемой, с одной стороны, полисемией словарных ЛЕ, многозначностью морфологических форм и синтаксических схем, содержащихся в тексте, а также

недостатком лингвистических и энциклопедических знаний в ЛИБД. Поэтому децизивная архитектура должна обладать средствами сокращения этой неопределенности. Они включают, во-первых, фильтрующие алгоритмы, частично снимающие эту неопределенность [21, с. 91—145], а во-вторых — приемы адаптации ЛА к обрабатываемым текстам. К последним относятся в первую очередь пополнение АС географическими названиями, именами собственными, а также терминологическими с/ф и с/с, характеризующими данный подъязык, а затем создание новых и перестройка уже работающих алгоритмов. Это дообучение ЛА осуществляется как в процессе человеко-машинного общения, так и в автономном автоматическом режиме путем отбора автоматом наиболее частых альтернативных решений [13, с. 296—298] из КТВ [22]. Комплекс всех этих приемов и образует адаптационный уровень децизивной схемы ЛА.

Наиболее важной для развития концепции ЛА является проблема организации и функционирования механизмов третьего страта.

Рассмотрим эту проблему в деталях.

2.2.3. Поиск способа наилучшего решения задачи.

К настоящему времени в СтР для поиска оптимальных решений выработан ряд приемов, учитывающих инженерно-лингвистические ограничения, накладывающиеся на конкретную систему АПТ. Часть из них уже алгоритмизирована. Наиболее интересны два приема этого поиска.

Первый заключается в иерархической организации работы подсистем и модулей. Этот способ, сочетающийся со способностью последних к автономному функционированию, реализуется в следующих правилах:

— высшим уровнем управления является человеко-машинный блок принятия решения,

— подсистемы и модули верхних уровней обусловливают целенаправленную работу соответствующих блоков нижних уровней;

— если подсистемы и модули нижнего уровня, работая в автономном режиме, оказываются неспособными принять решение или принимают несколько альтернативных решений, то полученные здесь результаты переработки текста передаются на верхний уровень иерархии для выработки там окончательного решения.

Рассмотрим эту процедуру на примере системы, осуществляющей тема-рематический перевод немецких заголовков статей, относящихся к предметной области (ПО) "Сточные воды" [27]. Лексико-грамматическая переработка заголовка начинается на низшем уровне ЛА (шаг 1) и осуществляется с помощью немецко-русского словаря и машинной морфологии, содержащейся в ЛИБД. Она представляет не только лексико-морфологические заголовки для последующего перевода, но включает также сведения о морфологических граничных сигналах и семантическую информацию, которые будут использованы для принятия решений на более высоких уровнях.

Затем (шаг 2) с помощью морфологических индикаторов, полученных на предыдущем шаге, и синтаксических граничных сигналов, извлеченных из ЛИБД, осуществляется членение заголовка на семантико-синтаксические сегменты. Для определения коммуникативной (тематической или рематической) природы этих сегментов применяется цепочка фильтров, с помощью которых и принимается одно из альтернативных решений.

Первый фильтр имеет вероятностно-синтаксическую природу. Дело в том, что предварительные статистические исследования немецкого заголовка показали, что примерно в 90% случаев первый сегмент целиком совпадает с его ремой или входит в нее. Что касается конечных сегментов заголовка, то они примерно в 70% случаев входят в его тематическую часть. Еще менее четкую принадлежность к реме или теме дают второй и третий сегменты. Таким образом, позиционное сегментирование заголовка часто не дает однозначного решения. Поэтому полученные на шагах 1 и 2 результаты приходится передавать на более

высокий уровень, где осуществляется дальнейший коммуникативный анализ текста и одновременно проверяются ранее полученные результаты сегментов (шаг 1). Здесь работает фильтр, осуществляющий проверку всех слов и основ заголовка на их совпадение с ЛЕ, которые находятся в списках рематических и тематических индикаторов, содержащихся в ЛИБД. Результаты этого отождествления сопоставляются с информацией, переданной с нижнего уровня (шаг 2). Эта операция дает обычно один из следующих исходов.

1. ЛЕ, опознанные в качестве рематических индикаторов, попадают в начальный, а тематические индикаторы — в конечные сегменты. Таким образом, результаты анализа на обоих уровнях совпадают, и ЛА принимает следующее решение: конечный сегмент представляет тему, а начальный — рему. Примыкающие к тематическому и рематическому сегментам атрибутивные с/с могут быть детерминативами темы или ремы.

2. Информация, добытая на шаге 2, противоречит информации, выработанной на шаге 3. В этом случае приоритетным является тема-рематическое разбиение заголовка, полученное на третьем шаге.

3. Шаг 3 не дает однозначного решения по тема-рематическому разбиению заголовка. В этом случае параметры заголовка, полученные на шагах 2 и 3, передаются на следующий шаг, основная задача которого состоит в соотнесении заголовка с подобластями (ППо) исследуемой предметной области. Для решения этой задачи используются индексы принадлежности терминологических с/ф к ППо "Сточные воды" и другим смежным ПО, затрагиваемым в исследуемых текстах (эти индексы указываются в словарных статьях входных терминов).

Предварительные статистические исследования показали, что термины, принадлежащие к подобластям "Сточные воды", чаще всего встречаются в тематической части заголовка, в то время как термины других ПО — в рематическом его сегменте. Таким образом наличие в сегменте с/ф, принадлежащей к той или иной ПО, может быть использовано в качестве дополнительного индикатора рематичности или тематичности сегмента. Проиллюстрируем эту ситуацию на примере темарематического анализа немецкого заголовка статьи из журнала Wasserwirtschaft — Wassertechnik (1985, N1, S. 4). Обработка заголовка на шагах 1—3 не дает окончательного решения по опознанию темы и ремы. Поэтому все параметры передаются на следующий шаг, где путем обработки индексов принадлежности терминов происходит отнесение обрабатываемого заголовка к одному из ментальных пространств (подобластей) исследуемой ПО. При этом термины, имеющие индексы указанных ментальных пространств, рассматриваются в качестве слабых индикаторов темы, а с/ф и с/с с индексом какой-либо другой, отличной от области "Сточные воды" ПО, выступают в роли показателей ремы. Исходя из этого правила, сегмент eines Auswerterrechners включается в рематическую часть заголовка. Что касается сегмента Trinkwasseraufbereitung, то его предметный индекс подтверждает принадлежность этой с/ф к тематическому сегменту. Коммуникативная природа сегмента bei Verfahrensuntersuchungen остается невыясненной и должна быть решена в ходе общения ЛА* с потребителем.

Второй прием поиска оптимального решения состоит в способности ЛА к декомпозиции или упрощающей модификации общей задачи $\hat{P}$ в том случае, когда решить ее невозможно или это требует таких временных затрат и ресурсов памяти, которыми система в настоящий момент не располагает.

В случае декомпозиции общая задача $\hat{P}$ представляется в виде множества частных задач P_1, P_2 и т.д. В качестве примера рассмотрим ситуацию, возникшую при построении экспериментального турецко-русского МП. Неизоморфность именной и глагольной парадигм турецкого языка и соответствующих им русских парадигм крайне велика. Поэтому лексико-морфологические модули λ и λ' оказываются неспособными без взаимодействия с модулями анализа и синтеза поверхностной и глубинной структур предложения (g/g', s_1/s_2') выдавать русские с/ф с/с, которые морфологически точно соответствовали бы входным турецким

с/у. Поскольку модули g/g', s_1/s_2^1 для турецко-русского МП пока не созданы, лексико-морфологическую задачу λ/λ' приходится разбить на следующие независимые подзадачи:

P_1 — анализ турецкого с/у, в результате которого оно расчленяется на основу (исходную форму) и составляющие ее аффиксы (ср. модуль q),

P_2 — определение грамматической природы каждого аффикса (модуль q'),

P_3 — перевод основы (модули l/l').

Используя информацию, полученную в результате решения каждой из указанных подзадач, потребитель сам формирует перевод входного турецкого предложения. Типичным примером замены общей задачи $\bar{P}$ на ее упрощенную модификацию $\hat{P}$ служит переход ЛА на пословно-пооборотный перевод в тех случаях, когда ему не хватает морфологических и семантико-синтаксических ресурсов для построения поверхностной и глубинной структур входного предложения. Позволяя выходить из тупиковых ситуаций, возникающих при отказе автомата от заданной формы переработки текста, прием декомпозиции и упрощения P заметно повышает "живучесть" ЛА[4].

3. Заключение.

Формирование концепции ЛА и ее реализация постоянно наталкивается на невидимый барьер отторжения, разделяющий язык человека и искусственный язык компьютера. Среди образующих этот барьер "генетических парадоксов" [14, с. 168—169] наиболее существенным оказался парадокс человека и робота, обобщивший частные противоречия человеко-машинного интеллектуального и лингвистического взаимодействия. Если рассмотреть его взаимодействие в свете известной теоремы Геделя о неполноте, то лингвистический смысл указанного парадокса можно представить следующим образом.

Пусть на основе имеющихся неформализованных знаний о ЕЯ и его лексико-грамматических описаний создается предельно мощная формализация F (при этом речь идет не только о машинной грамматике, но о любом формальном описании языка). Однако в силу таких особенностей ЕЯ, как идиоматичность, толерантность, нечеткость, F будет содержать выражение S, которое окажется неразрешимым в F. Создание в обозримый срок более мощной формализации F', способной разрешить S, невозможно, поскольку ресурсы нашего формализованного знания в данный момент исчерпаны. Поэтому любая последовательная формализация будет беднее реальной эвристики ЕЯ. Иными словами, нечеткие и толерантные речемыслительные возможности человека оказываются хронологически впереди создаваемой им формализации для лингвистического робота.

С парадоксом человека и робота связаны более частные антиномии, среди которых укажем на:

1) несоответствие между континуальной природой языка, опирающейся на нечеткие толерантные лингвистические множества, и дискретным, оперирующим четкими эквивалентностными множествами, описанием ЕЯ в компьютере;

2) противоречие между открытым, динамическим (диахроническим) характером живого языка и закрытым, статическим (синхронным) его представлением в компьютере;

3) противоречие между единственным смыслом, с которым должен иметь дело компьютер, перерабатывающий текст, и многоаспектностью речевого сообщения, передаваемого от человека к человеку; необходимо иметь в виду, что каждое такое сообщение может нести в себе три смысла: обусловленные

---

[4] Под "живучестью" понимается способность ЛА сохранять свои наиболее существенные свойства в результате воздействия на автомат таких катастрофических факторов, как выход из строя некоторых внешних устройств или участков оперативной памяти, искажение отдельных лексических единиц, грамматических кодов и правил.

прагматикой коммуникантов авторский и перцептивный смыслы и независимый от этой прагматики универсальный смысл [28, с. 27—30].

Большинство из описанных в разд. 2 приемов построения ЛА было выбрано с расчетом, чтобы снизить барьер отторжения и связанные с ним парадоксы. Этой задаче служит информационно-статистическое применение данных, извлекаемых из корпуса виртуальных текстов, отражающих современное состояние и динамику конкретного языка и его разновидностей. Этой цели служат также опирающиеся на лингвостатистику и КВТ приоритет лексических разработок и использование функциональнограмматической тактики. Снижает барьер отторжения способ нежестко модульного построения ЛА. Эти приемы позволяют отразить в искусственном разуме ЛА континуальную толерантность и постоянную динамику речемыслительной деятельности человека.

Однако наиболее сильным средством ослабления всех перечисленных антиномий и парадоксов является организация широкого человека-машинного взаимодействия, предусматривающего оперативную адаптацию ЛА, с одной стороны, к обрабатываему тексту, а с другой – через его перцепционный смысл к прагматике человека-пользователя.

Использование всех этих приемов отражает одну из гумманистических черт современного научно-технического прогресса, состоящую в нашем случае в осознании активной роли субъекта-пользователя в человеко-машинных системах АПТ.

СПИСОК ЛИТЕРАТУРЫ

1. *Чижаковский В.А., Бектаев К.Б.* Статистика речи. 1957—1985: Библиографический указатель. Кишинев, 1986.
2. *Пиотровский Р.Г.* Лингвистические аспекты "искусственного разума" // ВЯ. 1981. № 3.
3. *Kuhns R.J., Little A.D.* News analysis: A natural language application to text processing // American association for artificial intelligence. Spring symposium series. Text-based intelligence systems. Palo Alto; Stanford, 1990.
4. *Loatman R.B., Post S.D.* A natural language processing system for intelligence message analysis // Signal. 1988. V. 42. № 1
5. *Величковский Б.М.* Современная когнитивная психология М., 1982.
6. *Лурия А.Р.* Язык и сознание. М., 1979.
7. *Gardner H* The Mind's new science. A History of the cognitive revolution. With a new epilogue by the author: Cognitive science after 1984. N.Y., 1987.
8. *Пашковский В.Э., Пиотровская В.Р., Пиотровский Р.Г.* Психиатрическая лингвистика. Л., 1991.
9. *Пиотровский Р.Г.* Лингвистические уроки машинного перевода // ВЯ. 1984. № 4.
10. *Шингарева Е.А.* Семиотические основы лингвистической информатики. Л., 1987.
11. *Piotrowski R.* Linguistic automaton and computer-assisted language learning: a perspective // Language learning via the microcomputer. Proc. of CALL'89. International Conference on computer-assisted language learning at the Institute of applied Linguistics Wilhelm Pieck University Rostock. 15—17 November 1989. Rostock, 1990.
12. *Miller G.A., Johnson-Laird P.N.* Language and perception L., 1976.
13. *Пиотровский Р.Г.* Текст, машина, человек. Л., 1975.
14. *Пиотровский Р.Г., Беляева Л.Н., Попескул А.Н., Шингарева Е.А.* Двуязычное аннотирование и реферирование // Итоги науки и техники. Сер. "Информатика". Т. 7: Автоматизация индексирования и реферирования документов. М., 1983.
15. *Chomsky N.* Some concepts and consequences of the theory of government and binding. Cambridge (Mass.), 1982.
16. *Joshi A.K.* An introduction to tree adjoining grammars // Mathematics of language / Ed. by Manster-Ramer A. Amsterdam, 1987.
17. *Ristad E.S.* Computational structure of GPSG models // Linguistics and philosophy. 1990. V. 13. № 5.
18. *Гринберг Дж.* Некоторые грамматические универсалии, преимущественно касающиеся порядка значимых элементов // Новое в лингвистике. Вып. V: Языковые универсалии. М., 1970.
19. *Филлмор Ч.* Дело о падеже // Новое в лингвистике. Вып. X: Лингвистическая семантика. М., 1981.
20. *Halliday M.A.K.* Notes on transitivity and theme in English // JL. 1984. Pt 1—2. V. 4.
21. *Piotrowski R., Lesohin M., Lukjanenkov K.* Introduction of elements of mathematics to linguistics. Bochum, 1990

22. *Совпель И.В.* Автоматизированная переработка текста на основе воспроизводящего моделирования лингвистических объектов и процессов: Автореф. дис. ...докт. техн. наук. Харьков, 1991.
23. *Маковский М.М.* Английская этимология. М., 1986.
24. *Best K.H., Beöthy E., Altmann G.* Ein methodischer Beitrag zum Piotrowski-Gesetz / Hrsg. von Hammeri R. Bochum, 1990.
25. *Пиотровская К.Р.* Современная компьютерная лингводидактика // НТИ. Сер. 2. 1991.
26. *Попескул А.Н.* Продукционно-сетевой подход к моделированию смысла научно-технического текста: Автореф... докт. техн. наук. Винница, 1991.
27. *Чижаковский В.А.* Семиотико-коммуникативные аспекты автоматической переработки заголовка научно-технического текста: Автореф. дис. ...докт. филол. наук. Л., 1988.
28. *Гончаренко В.В., Шингарева Е.А.* Фреймы для распознавания смысла текста. Кишинев, 1984.