

ПИТРОВСКИЙ Р. Г.

ЛИНГВИСТИЧЕСКИЕ УРОКИ МАШИННОГО ПЕРЕВОДА *

Уже почти сорок лет, начиная с первых опытов А. Бута и Р. Риченса, идет спор между лингвистами-«машинборцами» и языковедами, верящими в научно-техническую ценность и перспективность МП и иных форм автоматической переработки текста (АПТ). Аргументация обеих сторон в этой всем изрядно наскучившей дискуссии хорошо известна, да и последнее слово в таких спорах принадлежит обычно истории науки и техники. Поэтому, оставляя в стороне спор о полезности и реализуемости МП и его разновидностей, рассмотрим те результаты и уроки, которые языковеды извлекают или могут извлечь из почти сорокалетнего опыта построения формально-машинных моделей языка и речи и использования этих моделей на ЭВМ.

Одним из наиболее существенных методологических итогов инженерно-лингвистических исследований явилось обнаружение глубокой антиномии, противопоставлявшей фундаментальные лингвистические теории и их приложения к языковому материалу. Ее существо заключалось в том, что языковеды, не располагавшие методическим аппаратом для оперативной проверки онтологичности используемой ими теории, т. е. ее соответствия объективной действительности, вынуждены были опираться одновременно на несколько, часто взаимоисключающих друг друга гипотез. При этом преимущество отдавалось в лучшем случае той теории, которая отвечала критериям непротиворечивости, полноты и простоты. Бывало, что эти теории оценивались, исходя из таких интуитивно-романтических критериев, как сенсационная новизна или красноречие автора теории. Еще недавно языковед, приступавший к решению конкретной лингвистической задачи или к разработке воспроизводящей инженерно-лингвистической модели (ВИЛМ), был лишен возможности осознанно выбрать в качестве методологического фундамента ту теорию, которая лучше всего отражает онтологию изучаемого лингвистического объекта. Эта ситуация напоминает до некоторой степени положение в географической теории и мореплавании эпохи Великих географических открытий, когда многие прославленные путешественники планировали свои экспедиции, исходя из таких произвольных гипотез, как уменьшенные размеры земного шара и существование только одного океана, отделяющего Европу от Дальнего Востока, наличие в Южном полушарии обширной Terra Australis, уравнивающей Евразию, или существование сказочных Эльдорадо и Маноя. Параллели можно найти и в судьбе некоторых фундаментальных лингвистических гипотез XX века. Оставляя в стороне «анализ по четырем элементам» Н. Я. Марра, вспомним очень популярную в 50-х и 60-х годах идею построения машинного перевода на основе хомскианских концепций глубинных структур и их трансформаций, которая была принята *ad hoc* без предварительной экспериментальной проверки. Понадобилось не менее пятнадцати лет безуспешных поисков с гигантскими затратами времени и средств, чтобы прийти к выводу о том, что «лингвисты, занимающиеся трансформационным анализом, которых общественное мнение обычно связывает с работой на компьютерах, практически не имеют отношения к машинному переводу» [1].

* Настоящая статья представляет собой обобщение и дальнейшую разработку идей и результатов, изложенных автором на Международном семинаре по машинному переводу, проходившем в Москве 5—10 декабря 1983 г., и на III симпозиуме по лингвистическим аспектам искусственного интеллекта (Ленинград, 2—4 апреля 1984 г.).

Еще более красноречивый пример того, как недостаточно обоснованная онтологически гипотеза направляет прикладника или лингвиста-практика по тупиковому пути, дает другая отрасль инженерной лингвистики (ИЛ) — автоматическое распознавание и понимание устной речи. Развернувшиеся в 50—60-х гг. работы по созданию устройств для автоматического анализа речи безоговорочно опирались на гипотезу, согласно которой звучащая речь представляет собой цепочку дискретных единиц-фонем. Альтернативная гипотеза, подсказанная материалом тоновых языков Восточной и Юго-Восточной Азии, по которой реальной единицей звучащей речи является силлабофонема [2], полностью игнорировалась. При этом фонема понималась то как типизированный гомогенный звук, то как нечленимая в синтагматике единица, функционирующая в составе морфемы, то как пучок смыслоразличительных признаков.

Эти концепции фонемы принимались многими фонетистами и инженерами-акустиком на веру, без предварительной экспериментальной проверки. И только неудачи при построении действующих систем по распознаванию слитной речи обнаружили, что названные теории не учитывали, по крайней мере, двух онтологических свойств звука: 1) акустической изменчивости каждого звука в ходе его произнесения; 2) размытости границ звука, приводящей к «взаимонашляванию» соседних звуков в потоке речи.

В результате многие фонетисты и инженеры-акустики стали склоняться к тому, чтобы считать фонему более или менее удобной гносеологической гипотезой-фикцией [3], не имеющей под собой достаточно глубокого онтологического фундамента. Одновременно в работах по распознаванию вместо фонемы начинают применяться такие понятия, как «дифон», «транзема» или «звуковой сегмент» [4, 5]. Эти понятия, учитывающие онтологию звуковой речи, представляют звуковые единицы меньшего размера, чем фонема.

Разумеется, периодический пересмотр фундаментальных концепций — это явление, свойственное всем наукам. Мощным средством оперативной проверки теоретических гипотез является хорошо спланированный и теоретически обоснованный эксперимент. Если говорить о языкознании, то здесь из-за недостатка экспериментальных методик фундаментальные концепции и гипотезы проверяются обычно не с помощью предварительного эксперимента, но путем многолетнего испытания в практических исследованиях языка и речи. Более того, слабость экспериментальных методик часто не дает возможности проверить до конца онтологическую ценность той или иной лингвистической теории. Поэтому многие старые, иногда даже скомпрометировавшие себя теории продолжают существовать рядом с современными взглядами на язык и его природу. В основе этой странной ситуации лежат субъективные причины: в отличие от наук о неживой природе, языкознанию, как, впрочем, и другим наукам о человеке, приходится иметь дело с «мягкой» (нечеткой) структурой исследуемого ею объекта и нестационарными процессами в текстообразовании и диахронии. Все это затрудняет применение традиционных и развитие новых экспериментальных методик.

Тем не менее сорокалетний опыт развития МП и ИЛ показывает, что предварительная проверка онтологической состоятельности лингвистической модели является разрешимой задачей. В качестве примера вспомним историю реализации лексико-фразеологического (пословно-пооборотного) МП. При всей внешней простоте и, казалось бы, примитивности указанной схемы по сравнению с концепцией грамматического, или «высококачественного» семантико-синтаксического автоматического перевода (АП), онтологическая состоятельность идеи лексико-фразеологического МП была предварительно подтверждена информационным экспериментом. Этот эксперимент показал, что от 67% до 85% информации текста заложено в лексических единицах (словоупотреблениях и словосочетаниях) [6]. Поэтому центральной задачей при построении действующей системы лексико-фразеологического МП было составление большого автоматического словаря (АС), который стал не просто информационно-лингвистической базой системы, но и ее алгоритмическим ядром. Основной ячейкой здесь

является словарная статья, которая благодаря открытости ее построения и заданию в ней функциональной (синтагматической) информации дает возможность лексико-фразеологическому МП при всей примитивности выдаваемого им перевода обойти некоторые препятствия, создаваемые парадоксами формализации и барьером отторжения [29], который возникает между компьютером и естественным языком [6, с. 44—54, 358—366].

Что касается альтернативных концепций, то их онтологическая состоятельность оставалась неясной вплоть до того момента, когда их авторы

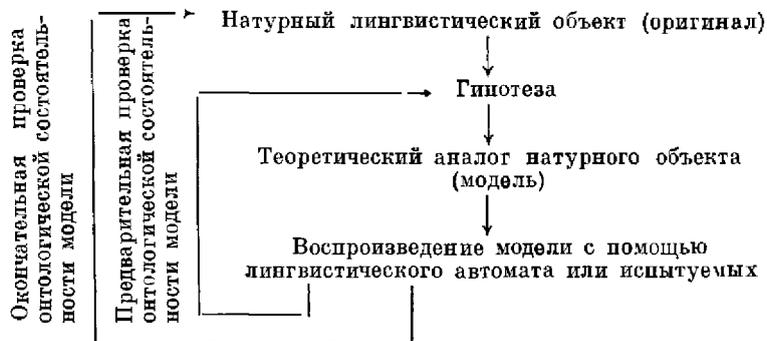


Рис. 1. Схема воспроизводящей лингвистической или инженерно-лингвистической модели

приступили к построению действующих систем МП. Именно в ходе машинной реализации этих систем стало ясно, что заманчивые умозрительные идеи «высококачественного» АП, не проверенные промежуточным экспериментом, парализуются барьером отторжения и парадоксами формализации.

Итак, опыт разработки МП показал, что наиболее эффективной формой реализации фундаментальных лингвистических теорий является представление их в виде структурных или динамических воспроизводящих моделей (рис. 1). В отличие от широко используемых в настоящее время теоретических моделей воспроизводящие модели должны обеспечить проверку онтологической состоятельности лингвистических теорий. Несмотря на слабое по сравнению с другими науками развитие теории и практики эксперимента в языкознании эту проверку уже сейчас можно проводить с помощью таких приемов, как информационно-статистический и просто квантитативный эксперименты, глоттохронологический и диалектный опыты, нейро- и психолингвистический, а также дидактический эксперименты и, наконец, машинно-грамматический и машинно-лексикологический опыты [6, с. 77—267; 7—9].

Важнейшей задачей языкознания является развитие теории и практики лингвистического эксперимента, который, став основой для теоретических выкладок, строящихся сейчас больше «на вере, чем на познании» [10], дал бы возможность нашей науке преодолеть барьеры, отделяющие ее теорию от языкового материала. Превращение языкознания из науки преимущественно описательно-спекулятивной в науку экспериментально-теоретическую является, на мой взгляд, основным условием того, чтобы лингвистика стала равноправным членом в системе общественных и естественных наук эпохи научно-технической революции. Таков первый гносеологический урок, который мы извлекаем из сорокалетней истории машинного перевода и инженерной лингвистики.

С момента своего зарождения системный подход к изучению языка был связан с теорией знака, которая реализовалась в виде таких конкурирующих гипотез, как прагматическая концепция Ч. Морриса, билатеральная психическая схема Ф. де Соссюра, «треугольник» Г. Фреге и Р. Карнапа вплоть до современных более сложных многокомпонентных построений Д. Науты, Г. П. Мельникова [11] и др.

Различия в этих семиотических концепциях объясняются, в первую очередь, разным пониманием их авторами природы языка, в том числе со-

отношения основных (конститутивных), — т. е. коммуникативной и экспрессивной (отражательной), — функций и дополнительных (прагматической, поэтической, магической, фатической, метаязыковой, референционной, контекстуальной и т. д.) функций. Так, для Ч. Морриса доминирующей оказывалась коммуникативная и прагматическая, а для Г. Фреге, Р. Карнапа и их последователей — экспрессивная функция языка. Последняя функция оказывается ведущей и в семиотической схеме Г. П. Мельникова. Что же касается Ф. де Соссюра, то он при построении своей билатеральной схемы знака исходил из постулата психичности и гештальтного (целостного, неанализируемого) отражения в нашем сознании объекта внешнего мира. В ходе построения действующих систем МП, в том числе и при неудачных попытках реализовать высококачественный АП, возникла необходимость построить эффективные модели словесного знака и коммуникации на естественном языке. Реализация этих задач становится мощным средством проверки онтологической состоятельности отдельных семиотических гипотез. Одновременно (и это самое главное) в ходе построения реальных систем МП и других видов АПТ постепенно выкристаллизуются те схемы словесного знака и речевой коммуникации, которые полнее всего отвечают коммуникативно-экспрессивной природе естественного языка.

Вклад ИЛ и МП в построение онтологически содержательных семиотических и коммуникативных схем можно свести к следующим пунктам.

1. В ходе создания реальных систем МП еще раз подтвердился высказываемый психолингвистами тезис о том, что построить универсальную систему АПТ, которая была бы способна перерабатывать текст без учета его принадлежности к определенной тематической или жанровой разновидности языка и прагматики адресата, невозможно. И дело здесь не столько в кризисе размерности, парализующем построение и функционирование такой универсальной системы, сколько в трудноразрешимых парадоксах самой коммуникации, создающих барьер на пути создания систем АПТ, обладающих универсальной прагматикой. Сущность основного парадокса состоит в том, что реальное содержательное сообщение включает в себе три смысла — независимый от коммуникантов универсальный смысл, авторский смысл и, наиболее важный с точки зрения целей коммуникации, перцептивный смысл, извлекаемый из сообщения самим адресатом [13] (к этому вопросу мы еще вернемся ниже). Для того, чтобы преодолеть или хотя бы ослабить указанный парадокс, в машинную модель словесного знака (точнее — в словарную статью лексической единицы АС) потребовалось ввести информацию о потенциальном адресате и той жанрово-тематической среде (ситуации), в которой функционирует данный знак. Можно предполагать, что такая информация как-то фиксируется и в реально существующей «записи» языкового знака в мозгу человека.

Все созданные до настоящего времени реально функционирующие системы МП используют два типа лексико-семантических алгоритмов: 1) автоматические словари словоформ, слов и словосочетаний, осуществляющие грубый «денотативный» перевод по схеме «входная лексическая единица (ЛЕ) — выходная ЛЕ» [6, с. 30—32; 358—366]; 2) алгоритмы распознавания смыслового образа, заключенного в тексте и его фрагментах, устранения языковой неоднозначности ЛЕ с помощью компонентного анализа их десигнатов, учета значимостей и реализованных валентностей [6, с. 255—307; 314—353; 367—376; 14]; при этом предполагается, что «десигнативные» алгоритмы станут центральным звеном «интеллектуальных» блоков ЭВМ пятого поколения [15].

Из сказанного следует, что в модели словесного знака, используемой как в теоретическом, так и в прикладном языкознании, должны быть учтены не только противопоставление денотата и десигната, но также противопоставление значимости и валентности знака. Ниже будет показано, что одной из сложнейших и пока нерешенных задач МП является моделирование вторичного означивания (семиозиса), в котором принимает участие коннотат знака.

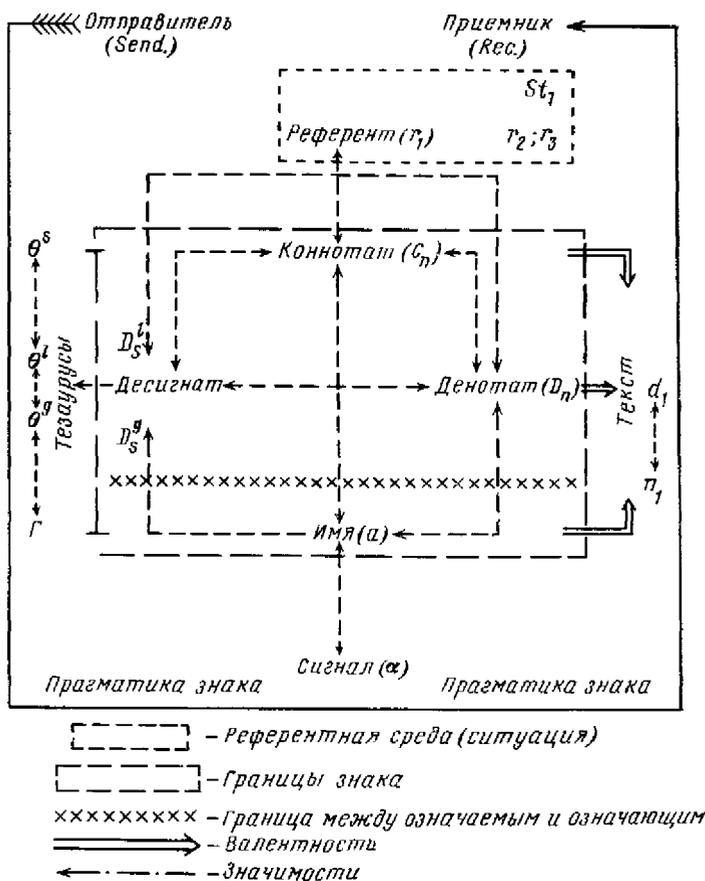


Рис. 2. Лингвистический знак и его окружения

Итак, учитывая опыт машинного моделирования языка и речи, можно предложить в качестве исходной модели ту схему словесного знака и его окружения, которые показаны на рис. 2. Поскольку эта схема понадобится нам при рассмотрении других лингвистических результатов МП, ее необходимо прокомментировать.

Знак, являясь психической сущностью, включает: 1) денотат D_n , т. е. целостный образ реального или воображаемого объекта внешнего мира в мозгу носителя языка; 2) десигнат с его лексическим (D_s^l) и грамматическим (D_s^g) аспектами, т. е. тот элемент системы социального опыта, который соответствует денотату; 3) коннотат, представляющий собой совокупность эмоциональных, оценочных и смысловых ассоциаций, содержащихся в означаемом знаке, — ассоциаций, служащих базой для вторичного семиозиса (переозначивания) знака; 4) имя (форму) a , т. е. внутренний психический образ сигнала α .

Окружение лингвистического знака охватывает: а) референт r , т. е. объект реальной действительности; б) сигнал (материальный носитель информации) α , представляющий собой физическое состояние или процесс, который выступает как средство обозначения объекта r ; в) прагматическую систему языка (человеческая «база данных»), которая включает стилистический — θ^s , концептуальный (лексический — θ^l , грамматический — θ^g) тезаурусы, множество формальных средств выражения (Γ), а также лингвистическую компетенцию отправителя и получателя сообщения; г) множество коммуникативных ситуаций (S_1^t, S_2^t, \dots и т. д.); д) тексты, которые содержат смысловые (d_1, d_2, \dots и т. д.) и формальные ($\Gamma_1, \Gamma_2, \dots$ и т. д.) элементы; е) прагматические намерения отправителя (Send.) и получателя (Rec.).

Знак формируется в результате семиозиса (знакообразования), или

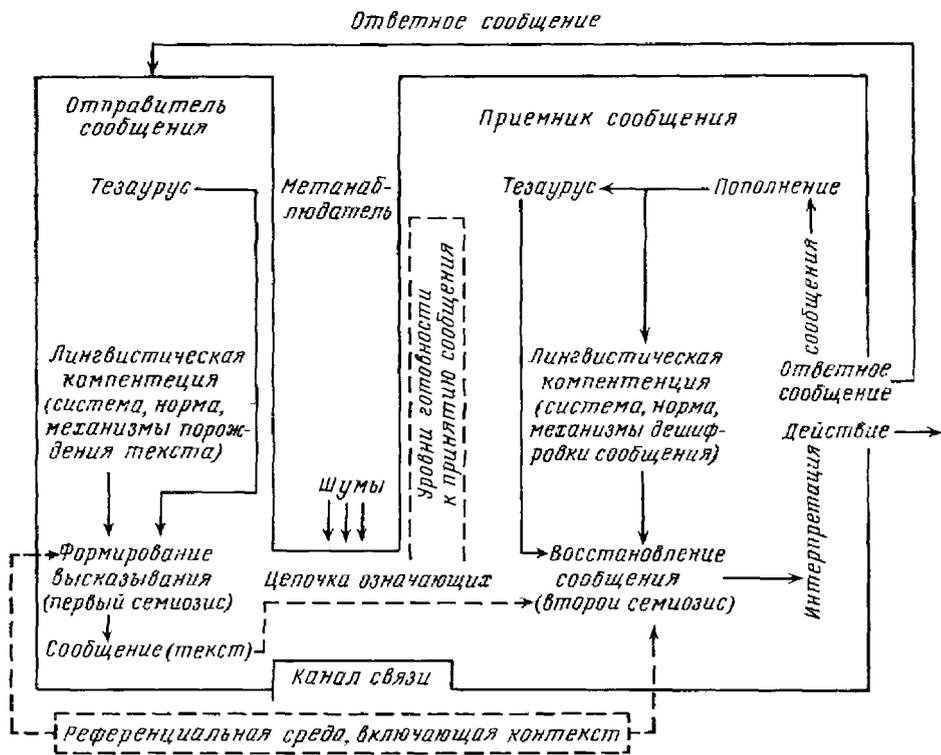


Рис. 3. Схема речевой коммуникации

означивания, которое представляет собой процесс, описываемый пятичленным отношением $R_s (\alpha \rightarrow a; r \rightarrow D_n \rightarrow D_s; C_n; \text{Send, Rec})$.

2. Действующие системы машинного перевода и других форм АПТ ценны тем, что они дают возможность наблюдать знаково-коммуникативный механизм порождения, передачи и приема сообщения. В частности, эти модели со всей очевидностью показывают, что в ходе речевой коммуникации знакообразование осуществляется дважды: первый раз при порождении сообщения его отправителем и второй раз при приеме и дешифровке сообщения абонентом (рис. 3). Из этого, казалось бы, самоочевидного факта (действительно, по каналу связи от отправителя к приемнику передаются не сами знаки, но воплощающие их сигналы) вытекает важное следствие. Оно состоит в том, каждый из указанных семиозисов регулируется не только коллективным стереотипом системы и нормы языка, но также интенциями и «видениями мира» отправителя и/или получателя (получателей) сообщения. Именно этим и объясняется присутствие в сообщении трех смыслов («общезыкового», авторского и смысла, извлекаемого приемником), о которых мы говорили выше.

Одновременно в ходе речевой коммуникации появляются условия для возникновения вторичного означивания (вторичного семиозиса), опирающегося как на коллективные коннотативные потенции знаков, так и на индивидуальные смысловые (и формальные) намерения коммуникантов. Напоминаем, что в ходе вторичного семиозиса узаконенный системой и нормой языка знак сам становится означающим для вновь созданного в процессе данной коммуникации или уже бывшего в ходу концепта (смысла).

Обычно считают, что вторичный семиозис, лежащий в основе каждого тропа, используется главным образом в литературно-художественной и отчасти разговорной речи. Однако, как выясняется, научно-технические, деловые и публицистические тексты также не могут обойтись без вторичного означивания, которое служит средством выражения новых понятий и выводов, а также применяется для воплощения «рематических» фраг-

ментов повествования¹. О масштабах вторичного семиозиса в указанных текстах можно судить по массовым машинным переводам наугад взятых и не подвергавшихся пост- и предредактированию (т. е. подгонке под алгоритм) текстов. Здесь более половины лексико-семантических ошибок обычно падает на случаи вторичного семиозиса, который в настоящее время не может быть распознан и правильно переработан лингвистическим автоматом. Ср. примеры, приводимые в [6, с. 360, 366].

Итак, инженерно-лингвистические исследования и их массовая реализация помогают выявить скрытые от прямого наблюдения, сложные механизмы порождения и восприятия речевого сообщения, а также эксплицитировать сопровождающие их семиотические процессы актуализации языковых значений в текстовые смыслы с помощью вторичного семиозиса. Одновременно МП помогает лингвистам осознать парадокс трех смыслов текста, — парадокс, в котором реализуется гумбольдтовская антиномия языка и идиолекта, а также противопоставление системы языка (тезауруса) и текста. Эти результаты составляют второй лингвистический урок МП.

По мере развития структуральных теорий языковеды вынуждены были приступить к их воплощению в виде модельных схем, призванных описывать сначала отдельные аспекты языка, а затем и язык в целом. Эти модели строились, как правило, путем введения фиксированного числа отношений, на основе которых с помощью определенного исчисления развевталась закрытая система с конечным числом элементов. Каждый из этих элементов должен был выступать в качестве аналога реального лингвистического объекта. После пионерских работ Л. Ельмслева и И. Шахтера, пытавшихся еще в 20-е и 30-е годы строить закрытые пространства логико-грамматических универсалий, исследования в этой области стали приобретать все более широкий размах. В 60-х и 70-х годах эти работы, «поддержанные» интересами МП с одного языка на другой и переводами с естественного языка на искусственный, достигли наибольшего размаха. В этот период появляются десятки построенных на традициях лейбнищевского универсализма логических языков, призванных по замыслу их авторов давать с помощью лексико-семантических кодов однозначные и непротиворечивые десигнативные описания объектов и фактов внешнего мира.

Вопрос о соотношении естественных и логических языков и возможностях описания естественного языка с помощью универсальных логико-семантических кодов является сложнейшей эпистемологической проблемой, которая не может быть детально рассмотрена и оценена в рамках статьи. Поэтому, несколько упрощая ситуацию, напомним, что с одной стороны, существуют многочисленные, выполненные в неопозитивистском плане исследования, авторы которых пытаются обосновать правомерность и перспективность построения универсальных логико-семантических пространств. С другой стороны, высказывались и высказываются принципиальные сомнения по поводу возможности построения жестких закрытых логико-семантических схем, способных дать исчерпывающее метаописание объектов и фактов внешнего мира и их отражений в языковом сознании носителей языка. Эти сомнения, высказываемые представителями так называемого научного реализма, основываются на восходящем к постулатам о языке Л. Витгенштейна предположении о том, что высказывание на ЕЯ является первичным и изначальным по отношению к логическому высказыванию, а не наоборот [17].

Сомнения по поводу возможности получать с помощью универсального логико-семантического кода адекватные описания объектов ЕЯ еще более возрастают при знакомстве с реальными примерами этих метаописаний, которые, принимая достаточно произвольную форму, не соответствуют языковому вкусу и здравому смыслу носителя языка (см. табл. 4).

¹ Здесь уместно вспомнить немецкое название известной работы З. Фрейда «Das Ich und das Es» и его английский перевод «The Ego and the Id», где местоимение *Es* (*Id*), переозначиваясь, становится выражением «подсознательной, анимальной инстинктивной области импульсов и потребностей», а *Ich* (*Ego*) обозначает сознательную или по большей части способную к сознанию область [16].

Слова	Интерпретация на Универсальном семантическом коде [63]		Интерпретация в „Словаре русского языка“ С.И. Ожегова [19]
	Формульная	Словесная	
Сидеть	$SS\bar{\downarrow} \leftarrow_a A\bar{\bar{0}}\bar{\bar{0}}$	Агент покоится посредством нижнего и тыльного медиатора [18, с.155]	Находится в неподвижном положении, при котором туловище опирается на что-то нижней своей частью, а ноги согнуты или вытянуты
Отец	$IS/\bar{S}_\chi \square ASS_\chi A\bar{\bar{0}}\bar{\bar{0}}$	Агент, призванный посредством неоптчуждаемого ффектора сделать так, что другой агент посредством неоптчуждаемого медиатора делает третьего агента существующим [18, с.159]	Мужчина по отношению к своим детям
Мальчик	$SS_\chi \{A\} \bar{\bar{0}}\bar{\bar{0}} //$ $IS/\bar{S}_\chi \diamond A/S/\bar{S}_\chi$ $ASS_\chi A\bar{\bar{0}}\bar{\bar{0}}$	Агент, который недолго существует и способен посредством неоптчуждаемого ффектора сделать так, что другой агент посредством неоптчуждаемого ффектора делает третьего агента существующим [18 с.155]	Ребенок мужского пола, подросток

Если вернуться к инженерной лингвистике, то здесь проблема соотношения естественных и искусственных языков оказывается теснейшим образом связанной с задачами устранения неоднозначности и актуализации словарных значений ЛЕ. Одновременно в этот круг эпистемологических идей и прикладных задач вовлекаются вероятностно-статистические приемы, информационные измерения языка и вопросы структурной организации конкретного ЕЯ.

Снятие многозначности и актуализация словарных значений ЛЕ производятся в современных лингвистических автоматах (ЛА) с помощью блоков двух типов. Блоками первого типа являются АС оборотов и алгоритмы, исследующие контекстное окружение ЛЕ [6, с. 105—115; 367—376]. Эти блоки, оперирующие целостным «денотативно-гештальтным» (рис. 2) представлением ЛЕ, обеспечивают простое по исполнению, быстрое и надежное устранение неоднозначности лексических единиц и их актуализацию. Однако эти преимущества денотативных моделей обеспечиваются их слишком большим размером. Действительно, как показывают информационные измерения содержащейся в этих словарях и алгоритмах синтаксической и семантико-прагматической информации, полные комплекты контекстных алгоритмов и АС «машинных» словосочетаний (оборотов) просто не помещаются в памяти серийных ЭВМ средней мощности [6, с. 15, 54]. Поэтому в современных системах МП и искусственного интеллекта денотативные блоки используются как вспомогательное средство для устранения неоднозначности и актуализации значений только у наиболее частых ЛЕ [20].

А что, если использовать информационные измерения относительно гипотезы генеративной лингвистики о существовании в языковом сознании носителя языка семантико-синтаксических ядерных структур, из которых путем применения трансформационных правил выводятся реальные предложения? Не рискуем ли мы обнаружить, что размеры полного набора ядерных структур заметно превышают информационные объемы языковой памяти человека?

Недостаточная эффективность АС оборотов и контекстных алгоритмов заставляет нас применять при автоматической переработке текстов бло-

ки второго типа, использующие компонентный анализ десигната (рис. 2) лексического знака. Все «десигнативные» блоки функционируют согласно процедуре формального распознавания смысла, существо которой состоит в том, что перерабатываемый текст и составляющие его ЛЕ отождествляются с предварительно помещенной в памяти ЛА моделью некоторой предметной области. Эта модель представляет собой искусственный язык, организованный в виде семантической сети или древесного тезауруса. После того как значение каждой ЛЕ текста оказывается отождествленным с определенным множеством семантических элементов (сем) сети или тезауруса, строятся декартовы произведения этих множеств. При этом каждое произведение соответствует той или иной последовательности (цепочке) словоупотреблений текста. Затем с помощью фильтрующего алгоритма из всей совокупности полученных цепочек отбирается та цепочка, которая является допустимой с точки зрения правил сочетаемости сем в данном языке. В результате применения этих процедур словарная неоднозначность словоформ, входящих в цепочку, оказывается устраненной [6, с. 367—376; 21].

Теоретический вывод, который можно извлечь из построения и функционирования «десигнативных» процедур в ЛА, состоит в том, что применяемые для формального описания лингвистических объектов искусственные языки должны строиться в виде открытых, способных к постоянному расширению и пополнению систем. Попытки формализовать открытый, непрерывно развивающийся естественный язык с помощью фиксированной номенклатуры [11, с. 243—249] неизменно приводят к возникновению таких ситуаций, при которых некоторые лингвистические объекты не могут быть адекватно описаны в терминах данного закрытого ИЯ. Именно за счет этой слабости закрытого ИЯ следует отнести появление приведенных выше (табл. 1) гротескных метаописаний для отдельных русских слов. Эта же антиномия закрытого языка автомата и открытого ЕЯ помешала использовать надежную номенклатуру Л. Ельмслева в системах МП и автоматического реферирования. Для формального представления в лингвистическом автомате значений английских, французских и русских предлогов пришлось построить открытый, бесконечно ветвящийся предположно-надежный тезаурус [6, с. 269]. Разумеется, открытость вводимого в ЛА искусственного языка следует понимать как потенциальную возможность для человека, работающего с ЛА, легко и беспрепятственно расширять и достраивать этот ИЯ, сообщая о всех его изменениях автомату. Сам же ЛА, не способный распознать вторичный семиозис [22] и преодолеть парадокс трех смыслов текста, практически не может самостоятельно расширять и пополнять введенный в него человеком искусственный язык. Любой ИЯ, включая фреймовые системы, оперирующие конечным числом шаблонов, остается по своему существу фиксированной номенклатурой или исчислением, изменить которую способен только человек.

Опыт разработки и реализации машинных алгоритмов, предназначенных для устранения неоднозначности и актуализации словарных значений ЛЕ, показывает, что оптимальные решения дает взаимодействие в ЛА, с одной стороны, «десигнативных» блоков, использующих компонентный анализ и фильтры, а с другой — «денотативных» блоков, которые функционируют вместе с процедурой вероятностного прогноза, опирающейся на цепи Маркова. Указанные алгоритмы, возможно, являются аналогами десигнативных, денотативных и вероятностно-прогнозирующих механизмов, используемых человеком при распознавании смысла текста. Удельный вес каждого из этих механизмов в процессе расшифровки сообщения зависит, по всей вероятности, от строя языка. Для агглютинирующих языков с их прозрачным, экономным и удобным строением ведущее место занимают «десигнативные» механизмы, в то время как во флективных языках с их «неуловимыми корнями, постоянно меняющими свою огласовку и теряющимися среди префиксов и суффиксов» [23], должны преимущественно использоваться вероятностные «денотативные» процедуры. Это предложение подтверждается практикой МП. При анализе и синтезе тюркского текста удобнее пользоваться сегментирующими словоформу «десиг-

нативными» алгоритмами [24]. Что же касается автоматической переработки текста на флективном или флективно-корневом языке, то здесь приходится либо осуществлять искусственную агглютинизацию флективных словоформ [9, с. 262] с тем, чтобы затем применить «десигнативную» процедуру, либо использовать автоматические словари оборотов и анализирующие контекст «денотативные» алгоритмы [6, с. 105—115; 367—376].

Все эти теоретические и практические результаты, полученные при реализации семантического МП, составляют третий лингвистический урочок машинного перевода.

Машинный перевод прошел трудный путь, на котором было больше неудач, чем реальных успехов. Однако каждый успех и особенно неудача проверяли состоятельность определенной лингвистической концепции и определяли дальнейшие перспективы и пути развития этой концепции. В этом, вероятно, и состоит основная ценность машинного перевода для современного языкознания.

ЛИТЕРАТУРА

1. *Troike R.* The future of MP.— American Journal of Computational linguistics, 1976, v. 13, № 6.
2. *Драгунов А. А.* Грамматическая система современного китайского разговорного языка. Л., 1962, с. 21—22.
3. *Реформатский А. А.* О реальности модели.— В кн.: Проблемы лингвистической типологии и структуры языка. Л., 1977, с. 4.
4. *Шуп Дж. Е.* Фонологические аспекты распознавания речи.— В кн.: Методы автоматического распознавания речи. I. М., 1983, с. 175—176.
5. *Вертаиб А. И., Жидошников А. Ф., Петров А. Н., Пиотровской Р. Г., Туркин В. Н., Чайковская И. И.* Лингвистика для роботов и автоматизированное управление воздушным движением.— В кн.: Инженерная лингвистика и преподавание иностранных языков с помощью ТСО. Л., 1981, с. 22—24.
6. *Piotrowski R.* Text — Computer — Mensch. Bochum, 1984, S. 234—245.
7. *Слама-Казаку Т.* Методика психолингвистического исследования при контрастивном анализе.— В кн.: Психолингвистика. М., 1984.
8. *Алексеев П. М., Герман-Прозорова Л. П., Пиотровский Р. Г., Щенетова О. П.* Основы статистической оптимизации преподавания иностранных языков.— В кн.: Статистика речи и автоматический анализ текста. 1974. Л., 1974.
9. *Пиотровская А. А.* Машинная морфология русского глагола.— В кн.: Статистика речи и автоматический анализ текста. 1972. Л., 1973, с. 260—276.
10. *Fry D. B.* Allocution du président du conseil permanent.— Proceedings of the 7-th International congress of phonetic sciences. The Hague — Paris, 1972, p. 5.
11. *Мельников Г. П.* Системология и языковые аспекты кибернетики. М., 1978, с. 218—266.
12. *Бруннер Дж. С.* Онтогенез речевых актов.— В кн.: Психолингвистика. М., 1984.
13. *Гончаренко В. В., Шингарёва Е. А.* Фреймы для распознавания смысла текста. Кишинев, 1984.
14. *Лесохин М. М., Лукьяненко К. Ф., Пиотровский Р. Г.* Введение в математическую лингвистику. Лингвистическое приложение основ математики. Минск, 1982, с. 92—149.
15. *Walter G.* Intelligent supercomputer: the Japanese computer sputnik.— Journal of information and image management, 1983, v. 16.
16. *Конечный Р., Боузал М.* Психология в медицине. Прага, 1983, с. 176—177.
17. *Хинтиikka Я.* Логико-эпистемологические исследования. Сборник избранных статей. М., 1980, с. 245—280.
18. *Мартынов В. В.* Универсальный семантический код. Грамматика. Словарь. Тексты. Минск, 1977, с. 155, 159.
19. *Ожегов С. И.* Словарь русского языка. М., 1981.
20. *Beliaeva L. N., Piotrowski R. G.* The stratificational approach to modelling of unitary database.— Computers in literary and linguistic research. 11-th international Conference [of the Association for literary and linguistic computing], 2—6 April 1984. Louvain, 1984, p. 27.
21. *Knowles F. E.* Recent Soviet work on computer techniques for representing natural language meaning.— In: The analysis of meaning. Proceedings of a conference held by the Aslib informatics group and the BCS information retrieval specialist group. 26—28 March 1979. The Queen's College, Oxford, 1979, p. 71—73.
22. *Степанов Ю. С.* Семиотика. М., 1971, с. 92—95.
23. *Трубецкой Н. С.* Мысли об индоевропейской проблеме.— ВЯ, 1958, № 1, с. 77.
24. *Бабанаров А.* Разработка принципов построения словарного обеспечения турецко-русского машинного перевода: Автореф. дис. на соискание уч. ст. канд. филол. наук. Л., 1981, с. 13—16.