

# Модель “смысл—текст”: тридцать лет спустя

И. А. БОЛЬШАКОВ,  
А. Ф. ГЕЛЬБУХ

Центр компьютерных исследований  
Национального политехнического  
института (IPN)  
г. Мехико, Мексика

*Оцениваются положение и перспективы модели “смысл—текст” (МСТ), выдвинутой в начале 70-х гг. российским лингвистом И. А. Мельчуком, в сравнении с основной западной парадигмой вычислительной лингвистики, предложенной Н. Хомским. Обсуждаются наиболее важные черты сходства и различия между МСТ и западным “главным течением” в его нынешнем состоянии. Лишь немногие черты МСТ признаются не реализованными, не реализуемыми либо устаревшими по своим формулировкам. Наоборот, значительная часть этих черт оказываются достоинствами МСТ и не имеют западного аналога. Однако западная парадигма развивается без всякого учета достижений МСТ, оставляя последней роль аутсайдера. Многие черты сходства между парадигмами были развиты Западом по своим внутренним причинам, и сходство в целом растет. В общем перспективы для МСТ довольно пессимистичны: она остается аутсайдерской и ее сторонники вынуждены представлять свои результаты на научном языке Запада. Высказывается уверенность, что после подобающего включения в современные западные теории многие результаты, полученные в рамках МСТ, послужат дальнейшему развитию прикладной лингвистики в целом.*

**Ключевые слова:** модель смысл—текст, генеративные грамматики, синтаксис, семантика, лингвистические формализмы

## ВВЕДЕНИЕ

Просмотрим труды конференций COLING, ACL, ANLP, LISA1 (последние — только в части лингвистики) за последние 10 лет и после этого мысленно зададим рядовому западному специалисту по вычислительной и прикладной лингвистике вопрос: “Что Вы знаете про модель “смысл—текст”?”. Скорее всего, в ответ мы услышим: “What do you mean?”. Если даже разъяснить ему, что это лингвистическая теория, которая была 35 лет назад предложена российским лингвистом И. А. Мельчуком, а затем развита усилиями самого Мельчука в Канаде и группы Ю. Д. Апресяна в России, наиболее начитанный из опрашиваемых может вспомнить: “А, это что-то из грамматик зависимостей. К английскому и ко мне это отношения не имеет”. Но скорее всего, ответом нам будет молчание.

Можно признать, что одна из первых по времени возникновения и наиболее продвинутая в России прикладная лингвистическая теория, автор и проповедник которой собирал на свои лекции в Москве в начале 70-х гг. сотни слушателей с самыми разными интересами, оказалась по существу невостребованной на Западе. В итоге даже вполне сочувствующий издатель вынужден назвать Мельчука в предисловии к его книге [1] “великим аутсайдером”.

Но вернемся в Россию, где табель о рангах отдельных ученых можно приблизительно составить при личном общении или опираясь на труды Международного семинара “Диалог” за 1995–1999 гг. Применительно к личности Мельчука

и его теории от прикладных лингвистов старшего поколения мы услышим воспоминания о ранней, романтической эпохе развития вычислительной лингвистики в России, хотя можем услышать и резкую оценку вроде “я уже двадцать лет говорю, что модель устарела”. Младшему же поколению российских лингвистов имя Мельчука мало что говорит. Для них он просто русскоязычный иностранец, которого раз в три года можно лицезреть в виде лектора в одном из ведущих институтов Москвы.

Настоящая статья написана в порядке дискуссии и призвана ответить на вопрос, поставленный в заголовке: так устарела ли МСТ? Забегая вперед, попробуем сформулировать наш ответ примерно так. Нет, эта лингвистическая теория идеологически не устарела, она продолжает развиваться как в прикладном [2,3], так и особенно в теоретическом плане [1,4,5], хотя число прикладных публикаций по ней на английском языке за весь период развития было мало. Но в рамках современного сообщества западных вычислительных лингвистов ничто, видимо, не спасет ее от вечного аутсайдерства. Отдельные ее черты уже воспроизведены независимо западной прикладной лингвистикой, иные же черты будут, скорее всего, переоткрываться в рамках других моделей. При всем этом никто может и не понять, что речь идет о переоткрытии, поскольку будут изобретены новые термины, формализмы и алгоритмы.

Ниже мы сначала перечислим основные черты МСТ, как мы их понимаем, — вначале те, которые стали уже общими для многих лингвистических теорий, а затем те, которые

остаются особенными для МСТ и поныне. Далее мы сформулируем причины, по которым модель была практически проигнорирована западной научной общественностью. Затем мы более развернуто повторим основной вывод и предложим нечто вроде рекомендаций.

Мы излагаем свое сугубо личное мнение, ни с кем детально не обсуждавшееся, так что ответственность за расставляемые акценты и точность сопоставления разных лингвистических теорий целиком лежит на нас как авторах сообщения.

## **ЧЕРТЫ СХОДСТВА МСТ С ИНЫМИ МОДЕЛЯМИ**

Перечислим кратко те черты МСТ, которые теперь, по прошествии многих лет, считаются общими чертами — или общими местами — большинства современных лингвистических моделей.

### **Функциональность модели**

Практически все известные языковые модели теперь функциональны, т. е. пытаются воспроизвести функции языка без воспроизведения черт активности мозга — носителя языка у человека.

### **Противопоставление текстовой/звуковой формы языка его семантическому представлению**

Западное учебное пособие [6,7], характеризующее три разные широко известные синтаксические теории (включая новый вариант теории Хомского), замечает: “В конечном счете язык выражает отношение между звуком на одном конце лингвистического спектра и значением на другом”. Как-то уточнив неопределенное понятие спектр, имеем то же определение, что и в МСТ.

### **Обобщающий характер языка**

Язык — теоретическое обобщение всей открытой и потому бесконечной совокупности высказываний. Обобщение оперирует чертами, типами, структурами, уровнями, правилами и т. п., непосредственно не наблюдаемыми. Эти теоретические конструкты являются плодами интуиции лингвиста, многократно проверяемыми на новых высказываниях.

### **Динамический характер модели**

Функциональная модель не просто предлагает совокупность понятий, но и показывает (с помощью правил), как эти понятия могут быть использованы при обработке языковых сообщений. Этим достигается соответствие между моделью и моделируемой языковой активностью человека.

### **Формальный характер модели**

Цель модели языка — сконструировать совокупность понятий и правил настолько строгих, что ими можно оперировать применительно к любому тексту или смыслу чисто формально, без участия автора модели или иного лица. При этом применение правил к фиксированной языковой информации всегда дает один и тот же результат (или одно и то же множество результатов). Любая часть функциональной модели может быть принципиально выражена и в строго

математической форме. Если готового математического средства на данный момент не оказалось, приходится развивать новое.

### **Непорождающий характер модели**

Информация не возникает при функционировании языковой модели, она просто принимает иную форму при переходе от текстового представления к смысловому или в обратную сторону. Хотя внутри модели форма сообщения меняется, применение для таких изменений термина Хомского трансформация несколько дезориентирует. Мельчук предложил для них термин эквативные преобразования.

### **Независимость модели от направления преобразований**

Правила языковых преобразований должны быть двусторонними или принципиально допускать обращение.

### **Независимость алгоритмов от данных**

Описание языковой модели отделяется от алгоритмов, которые это описание используют. Точные знания о языке не детерминируют какой-либо конкретный алгоритм. Наоборот, во многих ситуациях воплощающий правила алгоритм может иметь множество вариантов. В тех случаях, когда выделение декларативных знаний о языке удается провести наиболее последовательно, воплощающий алгоритм оказывается достаточно универсальным, применимым к нескольким языкам сразу. Подобная языковая универсальность перекликается с универсальной грамматикой, которую стремится ныне построить Хомский. МСТ на полный универсализм не претендует, но фактически все к нему готовит.

### **Упор на подробные словари**

Главная часть описания любого языка связана с описанием различных его слов. Поэтому словари считаются главной частью строгого лингвистического описания. Только очень общие свойства обширных классов и подклассов лексем выделяются из словарей, образуя то, что называется машинными грамматиками.

## **ОСОБЕННОСТИ МСТ**

Перейдем теперь к тем особенностям МСТ, которые она сохраняет по отношению к иным лингвистическим моделям. Мы сформулируем их так, как они представляются сейчас, через 25 лет после выхода основополагающего труда [8]. При этом не определяется, считать ли их достоинствами модели или некоторые — скорее недостатками.

### **Ориентированность на синтез**

При заявленной равнозначности направлений синтеза и анализа, первичным и более важным в модели считается синтез. Именно синтез привлекает все знания о языке, в то время как анализ возможен и на основе частичных знаний. В западной же традиции упор делается на анализ языка, в прикладном отношении считающийся более важным. В итоге для столь важных при анализе текстов унификационных алгоритмов нам пока не известны обратные. Остается неясным, как в рамках какой-либо из ветвей генеративной теории

превратить некую семантическую сеть в последовательность деревьев зависимостей или составляющих, отвечающих отдельным предложениям.

Когда идет речь о синтезе, на Западе обычно строятся другие теории, отличные от теорий анализа. Причина кроется в том, что конкретные генеративные грамматики описывают все предложения языка  $L$ , но при этом лишь те из возможных семантических структур, которые этому множеству  $L$  соответствуют. В то же время любая теория синтеза должна была бы описать все возможные семантические сети и соответствующие им предложения языка. Насколько нам известно, синтез текста по произвольно заданной семантической сети серьезно продумывался именно в рамках МСТ.

### Многоуровневый характер модели

Согласно МСТ, в языке есть несколько уровней (текстовый, два морфологических, два синтаксических, семантический), причем представление одного уровня считается полностью эквивалентным представлению любого иного уровня. Эквивалентный преобразователь "смысла в текст" и обратный ему преобразователь "текста в смысл" распадаются на несколько парциальных преобразователей с одного уровня на соседний. Разложение на уровни в модели призвано упростить правила межуровневых преобразований.

### Разнообразие структур и формализмов

Каждому парциальному преобразователю соответствуют свои правила и формализмы ввиду существенного разнообразия структур, которыми отображаются данные разных уровней (строки, деревья, сети). На каждом уровне в МСТ привлекается только минимально необходимое количество изобразительных средств. (Хотя при этом не считается обязательным, чтобы в прикладной системе парциальные преобразователи алгоритмически всегда воплощались строго в той же последовательности, в которой они стоят в модели.) Напротив, современная западная научная мысль старается найти практически для всех уровней языка единый формализм.

### Различение глубинного и поверхностного синтаксических представлений

Четко разделяются объекты и синтаксические черты, которые свойственны двум этим уровням. Вспомогательные и служебные слова текста на глубине исчезают. Аналогично, одни характеристики словоформ оказываются чисто грамматическими и остаются на поверхности (например, грамматические падежи и согласовательные показатели прилагательных), другие, определяемые семантикой, сохраняются и на глубинных уровнях. Такое разделение способствует минимизации изобразительных средств на каждом уровне. Понятие глубинного и поверхностного синтаксических представлений есть и у Хомского, но они определяются им по-иному.

### Независимость состава слов и порядка их следования в предложении

Вообще говоря, это свойство целой группы синтаксических теорий, а не только МСТ. Полной независимости порядка слов от их состава в модели не постулируется. Это две стороны, определяемые разными факторами. С формальных позиций различение двух этих сторон ведет к последовательному использованию на синтаксическом уровне грамматик

зависимостей, а не грамматик составляющих, как в большинстве западных теорий. В итоге основные правила межуровневых преобразований оказываются в МСТ иными, чем в западной парадигме. Основное преимущество грамматик зависимостей усматривается в том, что именно связи между (полнозначными) словами сохраняются на семантическом уровне, а для грамматик составляющих их обычно приходится выявлять на семантическом уровне отдельным механизмом.

### Учет коммуникативной структуры текста

Состав слов зависит от содержания высказывания, а порядок зависит как от этого состава (грамматически детерминированная расстановка), так и от коммуникативной структуры текста (влияние деления на тему/рему, старое/новое). В западной прикладной лингвистике зависимость порядка слов в предложении от коммуникативной структуры текста долго просто не замечалась.

### Ориентированность на языки иного строя, чем английский

В известной мере противопоставленность грамматик составляющих и грамматик зависимостей связала с разными типами языков. Грамматики зависимостей особо хороши для языков со свободным порядком слов типа русского и латыни, в то время как грамматики составляющих — для языков с жестким порядком типа английского. Впрочем, МСТ практически доказала свою способность описывать и языки типа английского, французского или немецкого. Уже накоплен большой опыт работы с деревьями зависимостей для любых языков. Генеративная традиция (например, HPSG) тоже приходит к тем же деревьям зависимостей, но постепенно и в имплицитной манере.

### Средства синонимического варьирования и лексические функции

Только в МСТ даны исчисление лексических функций и правила внутриуровневых древесных преобразований с их использованием. Это неотъемлемое и очень важное для модели средство синонимического варьирования предложений. Быть может, оно является наиболее важной особенностью модели, играющей ключевую роль в ее механизме синтеза (генерации текста), по глубине проработки не имеющем аналога в генеративной традиции. Именно с помощью синонимического варьирования производится поиск реализуемых на поверхности синтаксических вариантов данного семантического представления при переводе с одного языка на другой. Лексические функции позволяют также стандартизовать семантическое представление, уменьшив разнообразие узлов в нем. С прикладной стороны синонимическое варьирование разработано на таком же уровне строгости, что и западные формализмы, и практически — в программной реализации — проверено на разных языках. В рамках западной парадигмы подобная задача, похоже, и не ставилась.

### Разметка синтаксических отношений между словами

Известно, что если внутри каждого правила контекстно-свободной грамматики формально выделить главную составляющую (head), как это делается в HPSD, то дерево состава

вляющих легко переводится в дерево зависимостей, эквивалентное по содержащейся информации о связях между словами и группами слов. Но в МСТ имеется дополнительное свойство деревьев зависимостей — размеченность всех их дуг. При этом обнаружено, что в конкретных языках могут существовать изоморфные деревья с разными метками на дугах, и это различие связано с различием смысла.

### Модели управления

В отличие от субкатегориальных фреймов генеративной лингвистики, внешне несущих ту же информацию, модели управления (МУ) связывают семантические и синтаксические валентности лексем, притом не только у глаголов, но и у прочих частей речи. В итоге МУ позволяют явным образом указать, какими вариантами оформляется каждая данная семантическая валентность на поверхностном уровне: существительным беспредложно, существительным с конкретным предлогом или с несколькими разными предлогами по выбору, инфинитивом и т. п. Субкатегоризация же обычно сводится к простому перечислению всех возможных сочетаний синтаксическим валентностей при данном фиксированном порядке их расположения во фразе. В языках со сравнительно свободным порядком слов количество таких фреймов для отдельных глаголов может достигать нескольких десятков, и они затуманивают единую картину семантических валентностей. Да и разнообразие групп глаголов, имеющих одинаковое сочетание фреймов, бывает сопоставимо с числом глаголов в языке.

### Сохранение традиций и терминологии классической лингвистики

МСТ относится к наследию классической лингвистики значительно бережнее, чем западная вычислительная лингвистика. Своим многолетним развитием МСТ показала, что в значительном большинстве случаев даже повышенная точность описания и необходимость в строгом формализме позволяет сохранить уже известную терминологию, быть может, дав терминам более четкое определение. Сохранены понятия лексем, грамем, морфем, морфа, подлежащего, сказуемого, дополнений, обстоятельств и пр. В рамках же генеративной лингвистики теория строится каждый раз как бы с нуля, без попыток объяснить релевантные явления в терминах, известных в лингвистике ранее. Дополнительную строгость это дает не всегда, а чаще ведет к терминологической путанице и отрыву прикладных исследований от наследия и современного состояния теоретической лингвистики, поскольку специалисты близких областей просто не понимают друг друга.

### НЕОСТОРОЖНО СФОРМУЛИРОВАННОЕ, ТРУДНО РЕАЛИЗУЕМОЕ, СУЩЕСТВЕННО НЕДОДЕЛАННОЕ

#### Выдача всех синонимических вариантов сразу

В ранних работах И. Мельчук заявлял, что его модель предназначена для порождения параллельно сразу всех синонимических вариантов одного смысла (и выявления всех возможных разборов данного предложения, если оно омонимично). Как теоретическая формулировка соотношения смысла с текстами это было хорошо, но было понято некоторыми

исследователями как необходимость в таких программах и обещание строить именно их. Хотя в принципе МСТ такую возможность дает, практически при генерации текстов обычно нужно выдавать только один правильный вариант.

### Лозунг атомизации семантики

Неосторожным было прозвучавшее как обязательство реализовать в рамках данной модели систему семантических атомов (сем), через которые можно представить любые смыслы. Это пока не получилось ни в этой, ни в какой-либо иной теории. До сих пор идут споры, сколько таких атомов должно быть. Все благополучно сошлись на идее, что достаточно разлагать смыслы слов до разумного предела, диктуемого прикладной задачей. Так, перевод с одного близкого языка на другой может вообще не нуждаться в подобном разложении.

Ограниченное же и равно приемлемое для разных исследователей разложение смыслов применительно, например, к индоевропейским языкам, потенциально осуществимо, но требует для больших словарей десятилетия упорной словарной работы. Одним из направлений такой работы является начатая под руководством Ю. Д. Апресяна разработка синонимических словарей особого типа — с выделением всех семантических сущностей, различающих неабсолютные синонимы [9]. Без подобной проработки, пока не сопровождающейся формализацией, невозможно представить себе будущие системы “понимания” текста.

### Отсутствие правил расстановки слов

Пока отсутствуют правила расстановки слов, связанные с коммуникативной организацией фраз. Дело в том, что до сих пор не существует адекватной теории коммуникативной организации предложений и текстов в целом. В языках типа английского, где коммуникативная организация внутри предложения внешне обычно не выражена (порядок слов строго фиксирован), эту проблематику многие просто игнорируют. Но в МСТ, разделившей состав слов и их порядок, такое игнорирование выглядит как введение лишней сущности без формальных средств оперирования ею.

### Технология составления толково-комбинаторных словарей осталась не разработанной

Их составление оказывается до сих пор по силе только тем, кто осваивал модель много лет, а по существу создавал и совершенствовал ее. Вероятно, именно отсутствие ясной и массовой технологии разработки словарей явилось одной из основных исторических причин отставания модели от западных “конкурентов”, о чем мы еще поговорим подробнее ниже.

### ПРИЧИНЫ ИГНОРИРОВАНИЯ МОДЕЛИ

Попробуем теперь понять, почему же модель “смысл—текст” не нашла заметной поддержки на Западе, несмотря на то, что автор модели живет в Канаде уже более 20 лет и уделил немало внимания популяризации своих идей. На наш взгляд, причины здесь кроются в следующем.

#### Преждевременность

Исторически МСТ была предложена и проработана в деталях учеными с огромной эрудицией, в стране с глубокими

традициями гуманитарных исследований, но задолго до реальной возможности ее программной реализации.

С самого начала создателям МСТ была ясна невероятная объективная сложность задачи. Ими было детально проработано в теории множество вариантов, исключений, редких и трудных случаев. Была предложена сложная структура словарей, призванных содержать гигантский объем лексической информации. В результате первые же попытки практической реализации модели просто захлебнулись в сложности и объемах работы, особенно при отсутствии государственной поддержки в Советском Союзе.

Напротив, на Западе первые простейшие грамматики из десяти строк были восторженно подхвачены целой армией специалистов — и неспециалистов, — имеющих доступ к реальным компьютерам, корпусам текстов, а с ними — и к реальным заказчикам и коммерческим проектам. Для реализации этих проектов годились даже эти простейшие грамматики. Возможность легкого успеха, а с ней приток интереса и капитала вызвал бурное и, главное, последовательное развитие теории, создание массовой научной школы.

К сегодняшнему же пониманию того, что настоящий научный успех не легок и не скор, западная традиция пришла, пусть на 30 лет позже, но с багажом по крохам накопленных формализмов и их программных реализаций, опыта как разработки теории, так и ее преподавания, а главное, — с армией обученных профессионалов и “кредитом доверия” у научных и финансирующих организаций.

#### Англоязычный изоляционизм

Еще в начале XX века было предложено описывать английский язык с помощью непосредственных составляющих. Деление предложений на составляющие казалось непосредственно наблюдаемым в тексте, т. е. единственно “объективным”, и каждое новое наблюдение, производимое на английском материале, только подтверждало применимость такого деления. В результате метод составляющих стал в глазах англоязычных исследователей чем-то абсолютно универсальным, даже единственно возможным.

Научная Европа не спешила примкнуть к этому методу. Именно в 30-х гг. Л. Тенбер ввел в научный оборот деревья зависимости, которые задолго до этого были известны на неформальном уровне в практике описания различных языков, в частности русского. Когда же в Америке появились генеративные грамматики Н. Хомского, среди которых контекстно-свободные грамматики внешне столь хорошо соответствовали методу составляющих, это решение показалось единственно возможным и с формальной стороны. Иначе говоря, описание естественных языков с помощью генеративных грамматик на этом этапе казалось вполне адекватным.

Дальнейшее развитие генеративной традиции велось в распространении методов, уходящих корнями в английский синтаксис, на другие языки и на сложные случаи самого английского. Отдельные успехи на этом пути создавали иллюзию того, что метод этот универсален. Действительно, ведь описывает же он теперь падежи в исландском и клитики во французском, чего же больше? На деле до самого последнего времени это было, скорее, приспособление одной структуры для описания другой [10]. Впрочем, справедливости ради надо сказать, что МСТ уходит корнями в русский язык, а свойственный этому языку морфологический способ кодирования синтаксических отношений для западных языков не столь характерен и потому не особо интересен.

#### Финансовая мощь Соединенных Штатов

Граматики Хомского появились в 50-х гг., а конец 50-х — начало 60-х гг. характеризуется в США беспрецедентной поддержкой фундаментальных наук государством. Проиграв России первый раунд гонки за спутник, Америка сочла за благо поддерживать точные науки, сразу все и везде, тем более что денег хватало. Среди прочего в выгодном положении оказались и формальные грамматики. Количество научных (как правило, чисто математических) работ в этой области быстро росло, математическая лингвистика сформировалась как отдельная наука и стала преподаваться в крупных университетах страны.

Дальше наступило отрезвление в части универсальной приложимости контекстно-свободных грамматик к естественным языкам и одновременно — их триумф в приложении к языкам программирования. Появились иные лингвистические модели, но на базе уже изученных. В процесс познания втянулись молодые математики и программисты. Сформировались коллективы, научное направление которых стало называться вычислительной лингвистикой. И финансовая мощь США явилась основой быстрого развития таких коллективов.

Когда же Европа стала достаточно богатой, чтобы шире поддерживать движение, методы уже сложились. При этом более модными оказались те из них, которым изначально сопутствовала большая материальная поддержка.

#### Инерция высшей школы

Представим себе циклический процесс подготовки докторов в высшей школе на Западе. Доктор в области вычислительной лингвистики готовит себе смену, год за годом порождая новых докторов по той же специальности. Выпускники, которые идут затем в фирмы, разрабатывающие лингвистические программы, знают только ту теорию, которая им преподавалась. И даже если она не очень годится для совершенствования их программ, искать какую-либо иную просто нет времени, нужно программировать. Те же, кто остался в высшей школе, передают свои знания очередному поколению докторов в рамках уже сложившейся традиции. Нужно быть достаточно смелым и хорошо подготовленным в области общей лингвистики, чтобы внести в уже сложившуюся традицию какой-либо существенно новый элемент. Инерция развитой высшей школы сопротивляется нововведениям, особенно, когда они обещают что-то лучшее не для родного (и такого естественного) языка, а для экзотических, вроде исландского или русского, и притом в неопределенном будущем.

#### Незыблемая вера в возможности машин, а не человека

На наших глазах вычислительные машины увеличили свою производительность и объемы памяти на несколько порядков. Для хорошо знакомых с программированием, но весьма посредственно — с лингвистикой это выглядит прямым доказательством того, что на каком-то уровне своего развития машины смогут решать задачу “сами”, грубой электронной силой, путем громадных переборов различных вариантов. Разве не подтвердила эту тенденцию новейшая шахматная программа, когда выиграла одну из партий у Каспарова?

Но если это верно, зачем тогда пытаться понять тонкие механизмы языка самому или вручную переводить на “человеческий” (т. е. программистский) язык то, что преподносят лингвисты-теоретики вообще и их аутсайдерские теории

в частности? А здесь уж недалеко от приговора проигравшим: "Раз данный подход не стал популярным, значит, он не достиг нужного уровня зрелости, а потому его не стоит ни знать, ни тем более, преподавать". (Авторы получили такую формулировку в одной из рецензий.) Де, пусть поверженные сами доказывают свою необходимость для мира.

### Требование немедленного успеха

Любые новые теории в вычислительной лингвистике начинаются ныне с формализмов, затем следуют новые алгоритмы или способы погружения нового формализма в старые алгоритмы (скажем, унификационные). На этом докторские диссертации обычно и заканчиваются. Затем возможно некоторое отрезвление у тех, кто пытается новую теорию применить. В ее рамках нужно описать не менее нескольких тысяч (лучше — десятков тысяч) лексем, т. е. создать вручную или полуавтоматически (чисто автоматически пока не удастся никому) какой-то новый тип словаря. Для этого нужна технология, причем такая, чтобы могла использоваться теми, кому за эту работу докторской степени не обещают.

Если теория очень хорошо объясняет что-то новое и сулит существенные практические преимущества, а ее автор и/или руководитель проекта достаточно энергичен, чтобы достать деньги на создание крупного "лексического ресурса", новый словарь может появиться. Но против крупных проектов действует установка западных чиновников финансировать только краткосрочные проекты, максимум на 2-3 года, а за этот срок невозможно создать крупный словарь, будь то печатный или машинный. (Чиновников тоже можно понять: стоимость ученого на Западе слишком велика, как и степень подотчетности самих чиновников в расходовании госсредств.) Так что в лучшем случае словарь будет создаваться и расти по частям. И, естественно, это будет словарь собственный по идеологии и технологии создания, а не построенный по чей-то чужой теории, за которую академические лавры получить трудно, а, тем более, по такой, для которой и технология-то по существу отсутствует.

### Престиж и амбиции

Это еще один, чисто человеческий фактор. Приятно и выгодно (а подчас и необходимо для получения должности в университете и финансирования проектов) быть автором новой модной теории. В то же время нужно быть в нашем меркантильном мире немного альтруистом, чтобы сознательно поддерживать чужую теорию и каждодневным трудом способствовать ее продвижению в научный оборот. Проще придумать свои название, терминологию, формализм, не слишком отходя от "главного течения", дабы оставаться понятным теми, кто подобные аппараты понапридумывал раньше.

Представляется, что любого из перечисленных факторов хватило бы для происшедшего отвержения и игнорирования МСТ. В истории развития современной науки бывает, что побеждает не лучшая в каком-то смысле научная парадигма, а та, за которой более или менее случайно оказываются стоящими большие материальные средства, большая энергия и возможности их создателей и пр., т. е. обстоятельства, прямо к науке не относящиеся.

## ЗАКЛЮЧЕНИЕ

Резюмируем кратко, каким образом развивалась западная прикладная лингвистика, не замечая МСТ, и повторим наш основной вывод.

Модель "смысл—текст" явилась практически первой в СССР "кибернетической" теорией в области лингвистики,

ставящей перед собой и прикладные цели — создать двунаправленный лингвистический процессор. Эта теория несла в себе объяснение множества научных фактов и предлагала методы их формального описания. Она была неизмеримо более всеобъемлющей, полной и внятно изложенной, чем существовавшие в то время в СССР, и поэтому привлекла большое число заинтересованных наблюдателей и немалое число последователей.

Но на Западе эта теория первой не была, там уже был свой апостол — Н. Хомский. Основополагающая книга И. А. Мельчука осталась не переведенной на иностранные языки и, будучи широко известной в СССР, не приобрела читателей на Западе.

Запад продолжал идти своим путем, причем в сходном направлении. Было затрачено много интеллектуальных усилий (в том числе — и самого Н. Хомского), чтобы преодолеть недостатки начальных посылок хомскианских теорий. В конце концов, отказались от трансформаций, ввели (другим способом) нечто подобное моделям управления, эволюционно пришли к необходимости учета зависимостей непосредственно между словами и отдельного учета их порядка [7]. Изобретение простых и мощных унификационных алгоритмов упростило решение задач синтаксического анализа.

Последнее десятилетие характеризуется проектами по разработке крупных машинных словарей, ибо самые общие грамматические формализмы еще не решают практических проблем обработки естественного языка. Появились теории синтеза текстов, отдельные от теорий анализа.

Явные практические успехи западной традиции утвердили многих ученых в мысли, что революционные пересмотры основной линии ее развития не нужны, и если в аудсайдерских теориях есть нечто заслуживающее внимания, то в нужное время это неизбежно волеется в основной поток как-то само собой.

Поскольку на развитие МСТ за прошедшие три десятилетия в мире было затрачено неизмеримо меньше усилий, многие стороны этой модели остались слабо развитыми и порой звучат как благие пожелания. В плане же приложений не удалось развить удобных формализмов, например для синтаксического анализа. Осталась также не созданной доступная непосвященным технология создания крупных толково-комбинаторных словарей.

В этих условиях можно констатировать, что МСТ в отдельных своих формулировках стала общим местом или даже устарела. Однако она не устарела в своей общей идеологии. Она остается верной и плодотворной в большинстве своих посылок, и развитие западной парадигмы это подтверждает. Даже если модель будет модернизирована в ближайшем будущем в минимальной степени, она еще сможет послужить источником ряда идей и подходов, новых для западной прикладной лингвистики.

### Вместо рекомендаций

Мы не уверены, что в подобной ситуации стоит рекомендовать что-то конкретное сторонникам западной парадигмы. Наш призыв изучать научное наследие МСТ едва ли найдет отклик, даже если мы будем услышаны.

Что же касается последователей модели, то чтобы быть понятыми (и принятыми в международный журнал или на международную конференцию), им, похоже, всегда придется разъяснять свои идеи на языке западной парадигмы, т. е.

владеть двумя научными языками сразу. Вероятно, придется смириться с тем, что МСТ для Запада всегда останется маргинальной и что основная масса прикладных лингвистов понятия модели не освоит никогда (разве что переоткрыв их в рамках своей традиции).

Сторонникам и последователям остается утешаться тем, что они уже знают кое-что из того, что их западным коллегам еще предстоит открыть. Возможно, вместо безуспешной пропаганды самой модели стоит внимательнее приглядеться к тому, что из багажа российской традиции можно перевести на язык западной и влить в нее на пользу общего дела.

#### Одно методическое замечание

Теперь порассуждаем немного на тему "А нужна ли компьютерщикам обобщающая ("сквозная") модель языка вообще?". Если взглянуть на современную теоретическую лингвистику, то окажется, что одни занимаются в ней фонетикой, другие — морфологией, третьи — синтаксисом, четвертые — семантикой и прагматикой. Внутри фонетики кто-то углубился в акцентуацию, внутри семантики — в речевые акты и т. д. Нет предела дроблению великой науки о языке, как и нет, похоже, необходимости заниматься снова (после древних греков, Ф. де Соссюра и Н. Хомского) общей философией на тему "Что есть язык и какова должна быть его обобщающая модель?".

Критерием истинности теоретических лингвистических исследований остается логичность, последовательность, соответствие интуитивных представлений о данном языковом феномене автора данной теории и остального сообщества лингвистов. Языковые примеры часто приходится признавать доказательством для данной теории лишь под призмой наличия в данном исследовании логики, последовательности и соответствия.

В этом смысле многие работы самого Мельчука (например, по морфологии [5]) и других лингвистов-теоретиков выглядят как важные этапы внутреннего развития науки, и их нет смысла классифицировать по тем научным парадигмам, в рамках которых они возникли.

Положение в прикладной лингвистике несколько иное. Здесь критерием истинности является близость результатов функционирования программ обработки языковых высказываний к тому идеалу, который задается умственными способностями обычного человека — носителя языка. Поскольку процедуру обработки из-за ее сложности приходится разбивать на этапы, нужна некоторая "сквозная" модель, рекомендуемая, какие формальные черты и структуры следует приписать высказываниям и языку в целом на каждом этапе и

как эти черты будут участвовать в отдельных этапах языковых преобразований в машине.

Ю. Д. Апресян говорит о возникновении на базе машин экспериментальной лингвистики. Похоже, что в будущем экспериментальная проверка результатов самых глубоких "узкоотраслевых" теорий в лингвистике станет неизбежным и важным элементом развития лингвистики в целом. В прикладной же науке эксперимент оказывается главным уже сейчас, а он напрямую зависит от того, какие структуры выбраны для описания языка и какая последовательность преобразований рекомендована теорией.

Итак, вполне философская проблема модели языка оказывается для прикладной лингвистики первостепенной. Прикладники и в будущем будут вынуждены выбирать такую модель и при этом отличать ее от чисто математических и алгоритмических формализмов. Модель аккумулирует в себе языковые факты, а формализмы помогают их правильно применить на практике. При этом формализмы меняются быстрее моделей, и эта тенденция отчетливо видна в рамках рассмотренной темы.

#### ЛИТЕРАТУРА

1. Mel'čuk I. A. Dependency syntax: Theory and practice.— NY: SUNY Publ., 1988.
2. Апресян Ю. Д. и др. Лингвистическое обеспечение системы Этал-2.— М.: Наука, 1989.
3. Апресян Ю. Д. и др. Лингвистический процессор для сложных информационных систем.— М.: Наука, 1992.
4. Meaning — Text Theory. Linguistics, Lexicography, and Implications, James Steele (ed.).— University of Ottawa Press, 1990.
5. Мельчук И. А. Курс общей морфологии. тт. 1-4.— Вена, 1997.
6. Sells P. Lectures on contemporary syntactic theories.— Stanford: CSLI Publ., 1985.
7. Sag I. A., Wasow Th. Syntactic theory: A formal introduction.— Stanford:— CSLI Publ., 1999.
8. Мельчук И. А. Опыт теории лингвистических моделей "смысл—текст".— М.: Наука, 1974.
9. Апресян Ю. Д. Лексическая семантика. Синонимические средства языка.— М.: Восточная литература, 1995.
10. Гельбук А. Ф. Между смыслом и текстом // Труды Международного семинара "Диалог-99": компьютерная лингвистика и ее приложения.— Таруса, 1999.