

3. *Пойа Дж.* Математика и правдоподобные рассуждения. – М.: Наука, 1975.
4. *В.М. Лохин, И.М. Макаров, С.В. Манько, М.П. Романов.* Методические основы аналитического конструирования регуляторов нечеткого управления // Известия РАН. ТиСУ. – 2000, № 1, с. 56–69.
5. *Дюбуа Д., Прад А.* Теория возможностей. Приложения к представлению знаний в информатике. – М.: Радио и связь, 1990.
6. *Малышев Н.Г., Берштейн Л.С., Боженьюк А.В.* Нечеткие модели для экспертных систем в САПР. – М.: Энергоатомиздат, 1991.
7. *D. Dubois, H. Prade.* Fuzzy sets in approximate reasoning, Part 1: Inference with possibility distributions. // Fuzzy Sets and Systems № 100 (1999) pp. 73–132.
8. *I. Iancu.* Propagation of uncertainty and imprecision in knowledge-based systems. // Fuzzy Sets and Systems. № 94 (1998) pp. 29–43.
9. *Берштейн Л.С., Боженьюк А.В.* Нечеткие модели принятия решений: дедукция, индукция, аналогия. – Таганрог: ТРТУ, 2001.

М.Н. Жуков

МОДЕЛЬ ЕСТЕСТВЕННОГО ЯЗЫКА КАК УНИВЕРСАЛЬНАЯ МОДЕЛЬ ПРЕДМЕТНОЙ ОБЛАСТИ

Прогресс, достигнутый в последние десятилетия в области аппаратуры и средств программирования ЭВМ, несоизмерим с фактическим топтанием на месте в областях, связанных с интерфейсом человек-машина. Достижения в этой области, вошедшие в обиход, можно пересчитать по пальцам. Во-первых, это средства автоматизации проверки орфографии, синтаксиса, стиля и т. п., вошедшие в состав популярных текстовых процессоров. Во-вторых, это средства распознавания сканированных, в том числе рукописных, текстов. В-третьих, это системы распознавания голосовых сообщений и команд, пока еще далеко не совершенные. И в-четвертых, это средства синтеза речи из текста. Каждое из перечисленных направлений составляет отдельный объект исследований, сектор рынка коммерческих продуктов и имеет несколько независимых реализаций.

Между тем, многие направления исследований в области человеко-машинного интерфейса, заложенные в семидесятых-восемидесятых годах, так и не покинули стен лабораторий. Ключевым элементом здесь является метод представления знаний об объектах взаимодействия, т. е. модель предметной области. Все модели предметной области мы разделяем на две группы – экстенсиональные и интенсиональные.

Модели первого типа – к ним, в частности, относятся семантические сети – видоизменяются в зависимости от содержания конкретной предметной области. Вместе с увеличением количества информации о предметной области модель экстенсивно “растет” за счет добавления элементов ее описания. Таким же свойством обладают модели представления знаний, основанные на предикатах, используемые в экспертных системах, и т. п.

Модели второго типа не требуют настройки на конкретную предметную область. Информация из предметной области изменяет лишь количество экземпляров, т. е. интенционал модели. Удобство использования интенциональных моделей заключается в том, что в каждом конкретном случае не требуется настройки на предметную область и система, использующая такую модель, сразу готова к работе.

Интенциональная модель в чистом виде является идеалом, к которому реальные модели могут лишь приближаться. Например, любая мо-

дель предметной области, используемая в базах данных, в процессе эксплуатации базы данных является интенциональной – манипуляции с данными осуществляются на уровне добавления, изменения и удаления записей. Вместе с тем, любые изменения в предметной области требуют изменения в составе полей, т. е. изменения экстенционала модели.

Целью данной работы является изложение и обоснование одного из подходов к построению интенциональных моделей предметной области.

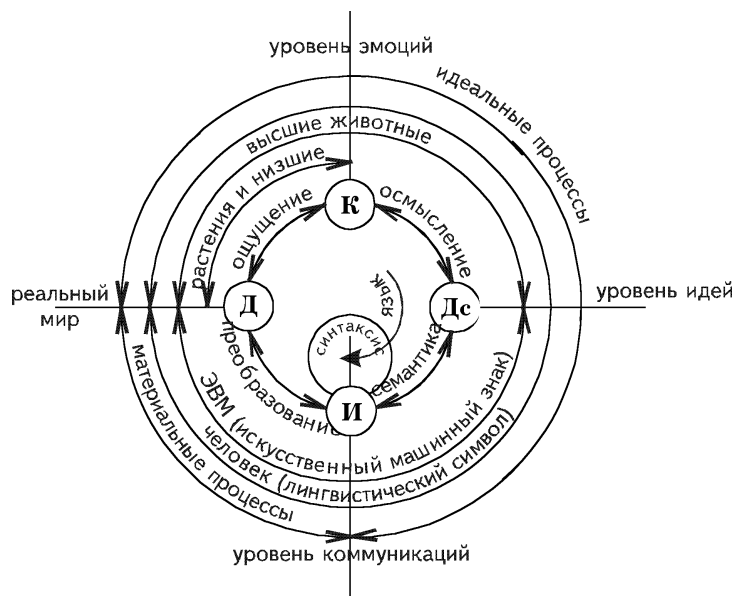


Рис. 1. Модель семиозиса в искусственных и естественных системах

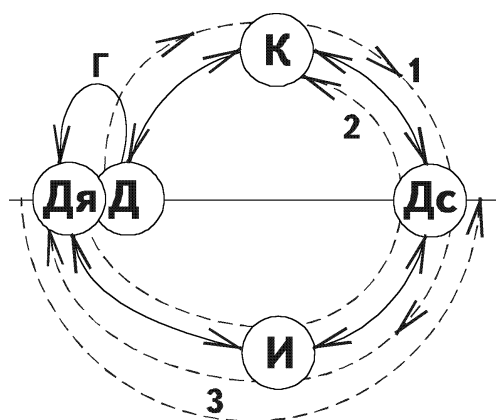


Рис. 2. Информационные потоки в искусственных и естественных системах

Модель семиозиса в естественных и искусственных системах

Прежде всего рассмотрим модель семиозиса в естественных и искусственных системах. При этом под семиозисом понимается процесс означивания, т. е. преобразования информации из реального мира в знаковую форму. Показанная на рис. 1 модель содержит четыре компонента так называемого лингвистического символа [1. С. 9]: И – имя, знак (материальный носитель информации), Д – денотат (предмет, явление действительности, обозначаемое именем), Дс – десигнат или концепт (т. е. смысл, понятие о предмете или явлении), К – коннотат, охватывающий дополнительные экспрессивно-оценочные значения. Эти же компоненты,

за исключением коннотата, образуют в своей совокупности искусственный машинный знак.

На рис. 2 стрелкой 1 показан информационный поток, возникающий при описании (вербализации) человеком объектов и явлений реального мира. Своими началом и концом он имеет денотаты Д и Дя соответственно. При этом Д – суть описываемые объекты и явления, а Дя – материализованная в оптический (текст), акустический (речь) или иной сигнал знаковая информация об этих объектах и явлениях. Эти сигналы становятся самостоятельными семиотическими объектами реального мира, т. е. денотатами для своих перцепиентов. Отображение Д в Дя, обозначенное на рис. 2 символом Г, представляет собой не что иное, как естественный язык (ЕЯ), а Дя – выражения естественного языка. Действительно, познание окружающего мира (стрелка 1 на рис. 2) является прерогативой человека – носителя ЕЯ – и его непосредственными результатами, естественно, являются выражения ЕЯ. Этот процесс осуществляется как на начальном этапе проектирования любой информационной системы (ИС) при создании модели предметной области, так и в процессе функционирования ИС при накоплении в ней новых данных. В подтверждение этому не столь очевидному, когда мы имеем дело с искусственно создаваемыми ИС, факту приведем далее несколько цитат:

ЭНТУИСТЛ, цит. по [2]:

“Любое знание вербализовано и нет знаний без слов ...”;

ГУМБОЛЬДТ, цит. по [2]:

“... языки являются ... средством выражения ... познанной действительности ... и средством познания ... неизвестной”;

ПОТЕБНЯ, цит. по [2]:

“... Язык есть средство не выражать уже готовую мысль, а создавать ее ... он не отражение уже сложившегося мирозерцания, а слагающая его деятельность”;

КОПНИН [3]:

“... язык – форма существования знания в виде системы знаков. Отсюда и само знание всегда выступает в виде какого-то языка”;

СЕПИР, цит. по [2]:

“Сущность языка заключается в соотношении условных ... звуков или их эквивалентов к различным элементам опыта”;

СОВЕТСКАЯ “ФИЛОСОФСКАЯ ЭНЦИКЛОПЕДИЯ” [4]:

“... С помощью языка осуществляется познание мира ...”.

Процессы восприятия знаковой информации (в данном случае выражений ЕЯ) человеком и искусственной ИС показаны на рис. 2 стрелками 2 и 3 соответственно. Что касается процесса 3, то о том, что он достигает уровня “десигнат”, можно говорить лишь для ограниченного числа ИС. К ним следует отнести такие системы обработки естественного языка, как “Программа, понимающая естественный язык” Винограда [5], ПОЭТ [6]. Однако и эти системы следует упоминать с большими оговорками: они обрабатывают ограниченный ЕЯ, функционируют в сильно ограниченной предметной области и поэтому по своей сути являются экспериментальными. Информационные потоки в остальных системах достигают лишь уровня “имя” (по стрелке 3 на рис. 2); к ним относятся информационно-поисковые системы, а также так называемые системы с текстовой основой [5]. По этому поводу заметим также, что предложенная нами модель укладывается в философскую теорию отражения: от живого созерцания к абстрактному мышлению и от него к практике – таков диалектический путь познания истины, познания объективной реальности [Ленин В.И. Полн. собр. соч. Изд. 5, т. 29, с. 152–153] – (стрелка 1 на

рис. 2). Что касается коннотата, то его философской основой является тезис: “Без человеческих эмоций никогда не бывало, нет и быть не может человеческого искания истины” [там же, т. 25, с. 112].

Информационные процессы в современных БД иллюстрирует рис. 3. Здесь стрелкой 1 показан информационный поток, имеющий место при построении концептуальной схемы (КС) БД и ее экземпляров. При этом разработчик, проектирующий КС, либо администратор БД, вносящий изменения в КС, либо, наконец, пользователь БД, создающий новый экземпляр КС, осуществляет отображение E денотатов $Дя$ – выражения $ЕЯ$ – в денотаты $Дк$, которые представляют собой выражения либо так называемого языка описания схемы, либо языка описания данных. На этом пути осуществляется концептуализация знания, заключенного в выражениях $ЕЯ$, что находит свое отражение в индексе “к” при обозначении элементов лингвистического символа.

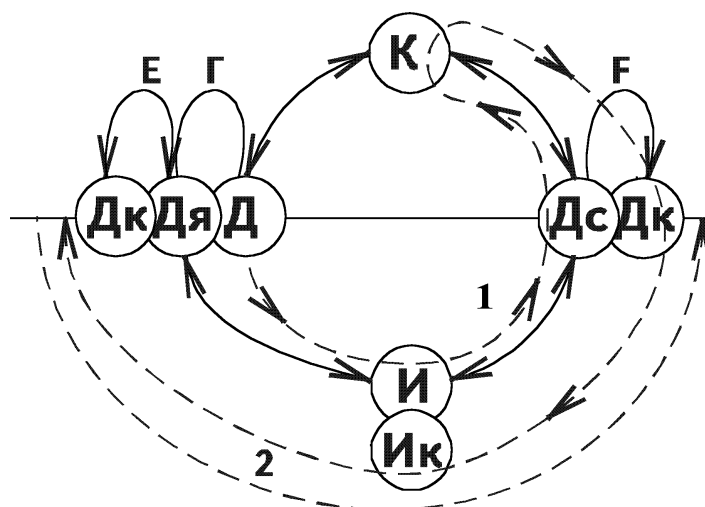


Рис. 3. Информационные потоки
в искусственных и естественных системах

Стрелкой 2 на рис. 3 показано направление потока поступающей в БД структурированной информации, т. е. информации, опосредованной КС. Это может быть запрос на поиск информации на некотором языке запросов, либо новые сведения, сформулированные на языке форматного типа. Простейший вариант последнего случая представляет собой ИС, получающая информацию непосредственно от датчиков, размещенных на исследуемом объекте. Концептуальной моделью здесь является уравнение или система уравнений, связывающих интересующий параметр с показаниями отдельных датчиков и их координатами на объекте.

Вышеприведенный анализ модели позволяет сделать следующие выводы. Информация из реального мира перед тем, как стать объектом хранения и манипулирования в ИС, претерпевает две последовательные концептуализации. Результатом первой из них являются актуализированные каким-либо образом выражения $ЕЯ$; на втором этапе концептуализации подвергаются выражения $ЕЯ$, в результате чего получается КС и ее экземпляры. Другими словами, естественный язык всегда является концептуальной моделью реального мира (ср.: “язык есть абстрактная сущность” [6. С. 21]; “в языке есть только общее” [Ленин В.И. Полн. собр. соч. Изд. 5. Т. 29. С. 249]), а КС суть концептуальная модель $ЕЯ$. Таким образом, в качестве исходного информационного объекта при проектировании ИС мы можем принять не нечто расплывчатое и неопределенное под названием “предметная область”, а вполне определенную сущность –

естественный язык. При этом все модели предметной области суть более или менее адекватные модели ЕЯ.

Предпосылки к разработке модели естественного языка

Что же представляет собой ЕЯ? Как известно, это – “универсальная знаковая система, служащая для передачи ... информации с помощью определенных материальных форм” [8. С. 11]. Известно также, что эти материальные формы традиционно изучаются на трех уровнях: слова, отдельного предложения и дискурса, т. е. связного текста [9]. При этом минимальной коммуникативной единицей ЕЯ является предложение [9. С. 16]. Предложение выражает законченную мысль, “... номинирует события, факты или явления объективной действительности” [10. С. 23]. На основании этого, в качестве фактов из ПО, которой, как мы выяснили, является ЕЯ, мы будем рассматривать предложения ЕЯ.

Известно, что смысл предложения, т. е. потенциально интерпретируемая информация, которую оно содержит, не складывается из значений входящих в него элементов, не является их суммой, а представляет собой качественно иное образование, хотя и формируется на их основе [11. С. 10]. С точки зрения формы предложение представляет собой некоторую линейную последовательность слов – знаков ЕЯ. Слова имеют два статуса: статус слов-названий, или слов-ономатем, и статус синтаксических слов, функционирующих в предложениях, или слов-синтагм [12. С. 30]. Значение слова-ономатемы определяется главным образом внелингвистическими факторами, его соотносительностью с миром вещей и понятий, значение слова-синтагмы в той же мере – факторами внутрилингвистическими, прежде всего контекстом. Традиционно слова-ономатемы изучаются в лексикологии с двух сторон: внешней материальной формы слова-лексем и его внутренней идеальной стороны – семемы. В терминах рассмотренной нами выше модели семиозиса последние суть имя и десигнат соответственно. Обе стороны слова-знака имеют комбинаторный характер, состоят из компонентов, каковыми являются части слова (морфемы) в лексеме, семантические компоненты (семы) – в семеме. Семы (семантические признаки, семантические множители, атомы смысла [7. С. 36], из которых складывается содержание лексических значений слов, соотносятся с признаками соответствующих понятий – основных форм мышления, с помощью которой реальная действительность отражается в нашем сознании. Признаки понятий, в свою очередь, отражают признаки явлений реальной действительности (денотатов), поэтому семы в своей основе имеют внелингвистическую природу. Этот аспект отражен в нашей модели на рис. 1 стрелкой “Язык”.

Семы различаются прежде всего в зависимости от их роли в структуре семемы, в которой они организованы иерархически [8. С. 34]. Самые общие семы, которым подчинены все другие – это семы грамматического характера (граммемы), по которым слова соотносятся друг с другом в рамках частей речи. Менее общими и, соответственно, более конкретными являются собственно лексические семы, примером которых является сема “перемещение” в значениях глаголов “ходить”, “лететь”, “носить” и т. п. При этом “глагол” является содержанием грамматической семы указанных лексем. Таким образом, структура десигнативной части лексических значений слов-ономатем представляет собой единство грамматических и лексических сем, связанных отношениями уточнения.

Однако одна из важнейших языковых функций слова – это функция члена предложения. В рамках предложения слова находятся в отношениях подчинения [12] и отношениях порядка. Если отношения порядка

выражены явно местоположением слов в предложении, то отношения подчинения выражены неявным образом. Во флективных языках, к которым в частности, относится русский, эти отношения актуализируются флексиями, приписываемыми по определенным правилам к основам слов. Отношения подчинения отражают синтаксические связи между словами в предложении, поэтому вполне уместно ввести в рассмотрение третью сторону слова-знака – синтему. Этим словом мы будем обозначать синтаксический компонент значения слова.

Соображения полноты требуют введения еще одного вида значения синтаксического слова, которое актуализировалось бы отношением порядка слов в предложении ЕЯ. Эту сторону слова-знака будем называть секвентемой. Введение в рассмотрение секвентциального значения делает для нас различимыми предложения одинаковые, но с различным порядком слов. Это позволит отражать коннотативный и стилистический аспекты семиозиса и лингвистики соответственно.

Таким образом, каждое слово в предложении ЕЯ может служить объектом исследования с трех сторон: со стороны лексического, синтаксического и секвентциального значений и при этом выступает соответственно как семема, синтема и секвентема.

К этому следует добавить, что указанные три вида значения имеют весьма слабую взаимозависимость внутри предложения ЕЯ. Действительно, изменяя порядок слов, мы практически никак не можем повлиять на структуру подчинений и лексико-семантические отношения. Заменяя лишь семемы, мы оставляем неизменными синтаксическое и секвентциальное значения и т. п.

Возвращаясь к предмету нашего рассмотрения, в качестве вывода сформулируем требования к идеальной модели ЕЯ: ее основным свойствам должна являться способность независимым образом фиксировать три вида значения – лексическое, синтаксическое и секвентциальное. Очевидно, что только при этих условиях можно говорить об адекватном моделировании ЕЯ.

ЛИТЕРАТУРА

1. *Пиотровский Р.Г.* Текст, машина, человек. – Л.: Наука, 1975. – 327 с.
2. *Налимов В.В.* Вероятностная модель языка. – М.: Наука, 1979. – 303 с.
3. *Копнин Л.В.* Философские проблемы языка. – В кн.: *Философия и современность*. – М.: Наука, 1971.
4. *Язык* / *Философская энциклопедия*. – М.: Советская энциклопедия, 1970. т. 5.
5. *Виноград Т.* Программа, понимающая естественный язык. – М.: Мир, 1976. – 294 с.
6. *Попов Э.В.* Обучение с ЭВМ на естественном языке. – М.: Наука, 1982. – 152 с.
7. *Кузнецова Э.В.* Лексикология русского языка. – М.: Высш. школа. – 152 с.
8. *Москальская О.И.* Грамматика текста. – М.: Высш. школа. – 183 с.
9. *Чахоян Л.П.* Синтаксис диалогической речи современного английского языка. – М.: Высш. школа, 1979. – 186 с.
10. *Аспекты общей и частной лингвистической теории текста*. – М.: Наука, 1982. – 192 с.
11. *Валгина Н.С.* Синтаксис современного русского языка. Изд. 2-е. Учебник для вузов. М.: Высш. школа, 1976. – 439 с.