

## ЭЛЕМЕНТЫ ТЕОРИИ АССОЦИАТИВНОЙ СЕМАНТИКИ\*)

**В.Н. Кучуганов<sup>1</sup>**

(ФГБОУ ВПО "Ижевский государственный технический университет", Ижевск)

*В статье предлагается инструментарий, призванный уменьшить неоднозначность семантических моделей высказываний и, тем самым, повысить толерантность человеко-машинного диалога и эффективность систем извлечения знаний. Семантическая модель ситуации, – сюжета описывается плекс-грамматикой, содержащей 4 сорта символов с различным количеством точек примыкания. Процесс синтеза семантической модели текста, рассматривается как процесс вывода корректной означающей конструкции, адекватной человеческому восприятию текста. Предлагаются операции сопоставления и сшивания графов образов. Оценка адекватности семантической модели осуществляется путем проверки грамматической корректности модели, установления соответствия между элементами новой модели и имеющимися в базе системы, а также выявления отличий в свойствах, отношениях и связях новых элементов.*

Ключевые слова: семантическая модель, плекс-грамматика, граф образа, ассоциативная семантика, вектор виртуальности, оценка адекватности.

---

\*) Работа выполнена при поддержке Федеральной целевой программы "Научные и научно-педагогические кадры инновационной России" Правительства Российской Федерации (проект №16.740.11.0423).

<sup>1</sup> Валерий Никонорович Кучуганов, доктор технических наук, профессор ([kuchuganov@istu.ru](mailto:kuchuganov@istu.ru)).

## 1. Введение

Этот материал следует рассматривать скорее не как теорию, а как очередную попытку формализовать такое многогранное понятие, как семантика контекстно-зависимых высказываний. Проблема в том, что как только эксперт и программист в пределах некоторого класса задач пришли к соглашению между собой, практика подбрасывает новые вариации и эксперт не понимает, для чего нужно что-то еще допрограммировать или почему он (эксперт) должен изучать какую-то новую методику, когда он и так профессионал в своей области. При этом далеко не факт, что компьютер решит проблему быстрее и лучше, чем он.

С точки зрения компьютера, если можно так выразиться, *семантика* понятия, определения – это константа или функция, запрос к базе данных, метод, которые выдают значения экземпляров, удовлетворяющих данному понятию.

Назначение текста – описать реальный или виртуальный (вымышленный) сюжет, продукт, изделие. Чем отличаются технические тексты от произведений искусства, "физика от лирики"? Утрировано, ценность последней возрастает по мере роста смысловой многозначности, тогда как в математике и технике это серьезнейший недостаток.

Герменевтика – теория понимания смысла доказала, что нельзя добиться однозначности истолкования естественно-языкового текста. Полифония высказывания заключается в том, что истина – это не конечное высказывание, а процесс. Советский философ Михаил Бахтин (1895-1975) открыл полифонию художественного высказывания, развил заимствованное у Эйнштейна и Германа Минковского понятие "пространство-время" и назвал его "хронотоп" (от греческого хронос – время и тоπος - пространство). Бахтин утверждает, что слово есть хронотоп, оно многомерное и говорит о неисчислимости смыслов, которые содержатся в любом высказывании.

Под семантикой понятия будем подразумевать метод реализации экземпляров, удовлетворяющих понятию.

Так же, как при общении между людьми, человеко-машинному диалогу необходим толковый словарь. Тем не менее, из-за

того, что одни и те же слова встречаются в определениях разных сущностей или используются как обозначения экземпляров, принадлежащих различным сущностям, возникает неоднозначность восприятия информации.

Другой причиной недопонимания является человеческая привычка пропускать факты, которые кажутся очевидными, или о которых он говорил совсем недавно.

Ассоциативная семантика, как фактор взаимопонимания, возникает по причине того, что источники и носители (хранители) экземпляров сущности имеют различную физическую природу и способы реализации. Чем сложнее *структура сущности*, тем больше разночтений.

Инструментарий теории ассоциативной семантики призван уменьшить неоднозначность семантических моделей высказываний и, тем самым, повысить толерантность человеко-машинного диалога и эффективность систем извлечения знаний.

## **2. Описание подхода**

Предлагаемая ниже методология ассоциативной семантики служит для сокращения семантической многозначности путем формирования семантической модели, позволяющей восстановить общий смысл текста в условиях многозначности его фрагментов, вычислить и оценить степень ее адекватности той цели, с которой текст написан.

**Толковый словарь** информационной системы дает определение сущности, тем самым устанавливая соответствие между термином (словом, словосочетанием) и его толкованием (пониманием) в виде *функции реализации* (получения, воспроизведения) экземпляров (образцов, денотатов) этой сущности. Значения конкретных экземпляров, выдаваемых функцией реализации, зависят от времени, текущего состояния информационной системы и ее параметров настройки.

Представим семантическую модель (СеМ) текста, описывающего ситуацию, в виде нечеткого графа ситуации, все вершины и ребра которого связаны с записями онтологического толкового словаря (ОТС), описывающими методы их интерпретации.

*Теорию ассоциативной семантики* (ТАС) определим как теорию

$TAS(O(OV, \{G\}, \{Pattern\}, \{R\}), \{P\}, V)$ ,

где  $O$  – онтологическая модель предметной области;

$OV$  – онтологический толковый словарь, в котором каждое слово или устойчивое словосочетание имеет несколько толкований со ссылками на онтологию, где эти понятия реализуются;

$\{G\}$  – множество грамматик, задающих правила построения сложных понятий из простых;

$\{Pattern\}$  – множество графов образов субъектов (экземпляров понятий);

$\{R\}$  – множество отношений;

$\{P\}$  – множество операций с образами;

$V$  – вектор виртуальности системы управления, реализующей ТАС – совокупность параметров настройки, которые задают прагматику системы.

Логический вывод путем моделирования рассуждений – это исчисление истинности высказывания (вопроса) по заданному описанию ситуации. При этом рассуждениями считается цепочка подстановок в процессе поиска ответа.

Технология ассоциативной семантики формализует процесс логического синтеза конструкции семантической модели ситуации/сюжета. Здесь процесс рассуждений можно интерпретировать как процесс ассоциативного определения сущностей, о которых идет речь, поиска подходящих экземпляров этих сущностей и согласования их между собой, так, чтобы не нарушить систему отношений конструкции в целом.

Задача (вопрос) ставится следующим образом.

Имеется:

1) онтологическая модель предметной области (ОМПрО);

2) текст, описывающий ситуацию/сюжет – условие задачи.

Требуется найти значение свойства, отношения, сущности.

Решение:

1. Построить адекватную СеМ ситуации/сюжета.

2. Дополнить СеМ информацией из аналогов с помощью ассоциативных связей.

3. Извлечь (вычислить) ответ.

### 3. **Онтологическая модель предметной области**

Онтология ПрО [Кучуганов, 2011] включает:

1. Декларативные знания:

– свойства, предметы, процессы – базовые типы и конструкции из них;  
– сюжеты и ситуации – образцы из предметной области (ПрО).

2. Императивные (процедурные) знания:

– справочники, в т.ч. по функциям и программам, как форма систематизации правил, представляющих однотипные зависимости между параметрами;

– логические операции выявления и установления отношений;

– отношения и правила предметной области;

– правила вывода;

– планы решения задач – это разбиения задач на подзадачи, в т.ч., подготовительные и заключительные действия.

Онтология предметной области содержит шесть семантических категорий концептов: свойство, предмет, процесс, отношение, сюжет, ситуация.

**Концепт-свойство** имеет вид:

*<Свойство> ::= <Имя>, [<Комментарий>], <Тип значения>, [<Метод>],*

где: *Тип значения* – непрерывный, дискретный, денежный, качественный, текстовый, ссылочный;

*Метод* – способ вычисления атрибута.

Свойства используются в качестве атрибутов при описании всех других категорий концептов. Имеется пять семантических разновидностей атрибутов (параметров):

- Имена (напр., ФИО, Адрес);
- Бренды, шифры (Код, Шифр, Ключ);
- Свойства собственные (по областям науки);
- Характеристики относительные (по видам отношений, см. ниже);
- Время.

**Предмет** имеет состав (детали), схему соединения и геометрические модели (кинематическая схема, чертеж, карта, 3D-модель):

$\langle \text{Предмет} \rangle ::= \langle \text{Имя} \rangle, \langle \text{Комментарий} \rangle, \langle \text{Список атрибутов} \rangle, [\langle \text{Список компонент} \rangle, [\langle \text{Список отношений между компонентами} \rangle,]] \langle \text{Геометрич. модель} \rangle]$

$\langle \text{Атрибут} \rangle ::= [\langle \text{Имя} \rangle], \langle \text{Концепт-свойство} \rangle, \langle \text{Диапазон изменения значений} \rangle, \langle \text{Ожидаемое значение} \rangle, [\langle \text{Единица измерения} \rangle,] [\langle \text{Экземпляр-метод} \rangle]$

Список отношений между компонентами определяет относительное положение, типы соединений, степени свободы и т.п.

Состав предмета – это дерево входимости деталей в сборочные узлы.

Граф-схема предмета:  $GSh = G(V, E)$ ,

где:  $V$  – множество деталей предмета;

$E$  – множество соединений (отношений предметной области).

**Процессы** имеют состав и алгоритм, т.е. их описание содержит вычислительные модели:

$\langle \text{Процесс} \rangle ::= \langle \text{Имя} \rangle, \langle \text{Комментарий} \rangle, \langle \text{Список атрибутов} \rangle,$

$[\langle \text{Длительность} \rangle,] \langle \text{Список входных концептов} \rangle, \langle \text{Список выходных концептов} \rangle, [\langle \text{Вычислительная модель} \rangle,] [\langle \text{Состав подпроцессов} \rangle, \langle \text{Граф-схема} \rangle].$

Роли участников действия (входных/выходных концептов): агент (исполнитель); бенефициант (заказчик); реципиент (приемник действия); предмет воздействия; сцена действия; инструмент; коагент; эффект и т.д.

### **Отношение:**

$\langle \text{Отношение} \rangle ::= \langle \text{Имя} \rangle, \langle \text{Тип} \rangle, [\langle \text{Комментарий} \rangle,] \langle \text{Список атрибутов} \rangle,]$

$\langle \text{Имя объекта 1} \rangle, \langle \text{Имя объекта 2} \rangle, \langle \text{Тип значения} \rangle, \langle \text{Ожидаемое значение} \rangle,$

$[\langle \text{Единица измерения} \rangle,] \langle \text{Экземпляр-метод} \rangle.$

Здесь экземпляр-метод – это конкретный метод из некоторого класса процессов или запрос на выборку из БД.

Семантически непересекающиеся категории отношений (значения которых принадлежат разным множествам):

1) *принадлежности* предмету, процессу, классу, множеству;

2) *родства* – степень удаленности друг от друга двух субъектов в родовидовом дереве;

3) *положения* субъекта в пространстве предмета, сцены относительно другого субъекта;

4) *следования* во времени относительно события, процесса;

5) *каузальные* – причинно-следственные связи;

6) *сравнения/сопоставления* значений свойств или математических моделей, в частности, графов;

7) *коммуникации* (толерантности) субъекта к другому субъекту или классу;

8) *сопутствия* – совместного существования разных сущностей;

9) *процессуальные* – ролевые отношения между участниками процесса;

10) *системные* – отношения на множестве субъектов, взаимодействующих между собой и оказывающих прямое или косвенное влияние друг на друга, например, в экологической системе.

Отношения 1 - 7 бинарные, 8 - 10 *n* – арные.

**Отношения двузначные** (принимающие булевские значения):

1) сравнение:  $\ll, < \leq, =, \geq, >, \gg$ ;

2) принадлежность предмету, процессу: *Part of*;

3) принадлежность классу, множеству: *ISA (is an instance)*;

4) родство: *AKO (a Kind of)*;

5) следование (за): *(A Follow B)*;

6) процессуальные:

- исполнитель: *(x Actor P)*;

- бенефициант (заказчик): *(x Customer P)*;

- реципиент (приемник действия): *(x Recipient P)*;

- субъект воздействия: *(x Subject P)*;

- место действия: *(x PlaceOf P)*.

Пример. *Преподаватель читает лекции студентам.*

(*x Actor P*) & (*y Recipient P*).

В конкретной ПрО роли могут быть более конкретными и вводиться экспертами. Например, Охрана труда. Несчастный случай. Роли: *пострадавший, свидетель, член комиссии*. Покупка: *продавец, покупатель, товар, контролер*.

**Отношения многозначные** (принимające значения соответствующих лингвистических переменных):

1) сопутствие (*accompanies*):

$x @ Z$ ,

в отличие от  $x \in Z$  –  $x$  принадлежит множеству  $Z$  однотипных объектов, означает, что  $x$  является атрибутом (свойством, параметром, признаком) объекта  $Z$ , т.е. является одним из множества *различных* атрибутов всех субъектов класса  $Z$ . Запись  $a @ b$ , где  $a \in A$ ,  $b \in B$  можно также применять, когда  $A$ ,  $B$  – являются разными сущностями и имеют разные наборы атрибутов, например *стул @ стол*;

2) положение: (*X LocationOf Y*);

3) сопоставление (графов): (*x Pattern y*);

4) толерантность: (*x ToleranceTo y*);

5) системные (*зависит от*): (*x DependsOn Y*).

**Ситуация** – это совокупность состояний предметов, процессов и отношений между ними, где состояние – это совокупность значений параметров предмета, процесса, отношения в некоторый момент или отрезок времени.

**Сюжет** – это временная диаграмма совокупности процессов. Всякий сюжет может быть раскадрован на ситуации с произвольным шагом по времени, в частности путем логического деления в моменты событий начала или конца любого из рассматриваемых процессов.

С точки зрения сюжета, предметы подразделяются на:

- исполнителей;
- других участников;
- статистов.

Процессы бывают:

- внешние, например, (шел) дождь;
- внутренние, например, Иванов гриппует;

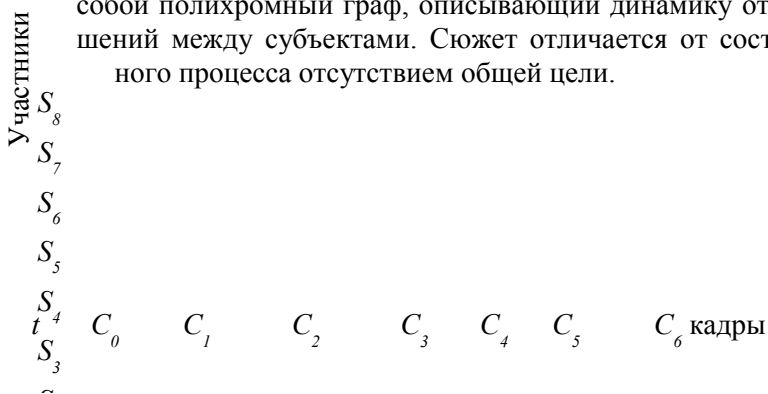


- деятельностные, например, Петров пишет программу.

Соответственно отношения:

- внешние;
- внутренние;
- межсубъектные.

Сюжет как способ запоминания опыта представляет собой полихромный граф, описывающий динамику отношений между субъектами. Сюжет отличается от составного процесса отсутствием общей цели.



*Рис. 1. Временная диаграмма сюжета*  
 пассивный субъект  
 активный  
 выполняющий  
 несколько действий

Граф сюжета можно представить как микрофильм, содержащий последовательность кадров – ситуаций, связанных между собой процессами и участниками (рис. 1):

1) каждый кадр – это плоский относительно времени граф, множество вершин которого отображает процессы и участников, а множество ребер – связи между ними в заданный момент или интервал времени;

2) каждые два соседних кадра связаны ребрами, соединяющими одноименные процессы и одноименных участников (двудольный граф).

Промежутки между так называемыми "ключевыми" кадрами могут быть различными. Если, например, граф сюжета отображает некий видеоклип, то промежуточные кадры генерируются автоматически с заданным шагом по времени.

#### 4. Семантическая модель ситуации

Для формализации семантической модели текста, описывающего ситуацию (СеМС), воспользуемся плекс-грамматикой [Feder, 1971], [Фу, 1974], где символы грамматических конструкций имеют не две (слева, справа), а  $N$  "точек примыкания".

Выделим 4 сорта элементов (деталей), из которых составляются модели ситуаций, и представим их в виде символов (рис. 2).

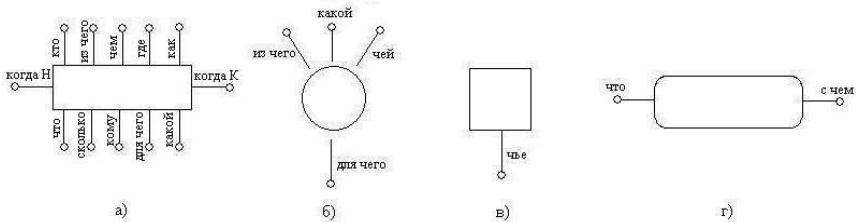


Рис. 2. Символы деталей семантической модели ситуации: а) процесс; б) предмет; в) свойство; г) отношение

**Процессы** в точках примыкания соединяются с элементами, отвечающими на вопросы:

(кто), {из чего}, {чем}, (где), (как), (что), (сколько), (кому),  
(для чего), (когда).

Фигурные скобки обозначают, что несколько символов одного сорта может быть связано с данным символом. Например, для процесса может потребоваться несколько наименований материалов (из чего), инструментов (чем).

**Предметы** в точках примыкания связываются с другими элементами в соответствии с вопросами:

(из чего), {какой}, {чей}, {для чего}.

Предмет характеризуется свойствами, различными по физической природе (какой) и имеет несколько предназначений (для чего) – процессов, в которых он может быть использован.

**Свойство** связывается с одной сущностью, которую оно характеризует.

**Отношение** связывает формулой две или более сущности.

Плекс-грамматику, порождающую множество семантических моделей ситуаций, представим как шестерку  $\langle V_T, V_N, R_S, R_O, Q, q_0 \rangle$ , где  $V_T$  – множество основных символов;  $V_N$  – множество производных (вспомогательных) символов;  $R_S$  – множество системных правил подстановки;  $R_O$  – множество предметных правил подстановки;  $Q$  – множество идентификаторов – меток точек примыкания;  $q_0$  – специальной (пустой) идентификатор.

Основные элементы изображены на рис. 2.

Производные элементы накапливаются в ПрО как устойчивые сочетания сущностей. Например: *человек – паспорт – квартира; театр – гардероб – фойе – зрительный зал.*

Идентификаторы точек примыкания для одного и того же элемента должны быть различными. В какой-то ПрО это множество может быть сокращено или расширено. Здесь множество идентификаторов точек примыкания приближено к вопросительным словам русского языка.

Идентификатор  $q_0$  служит для заполнения места, не связанного ни с какой сущностью, т.е. является признаком неполноты СеМС.

Правила подстановки – это правила построения конструкций из символов и ранее определенных конструкций, т.е. из основных элементов и производных.

Свойства присоединяются к процессам (какой) и предметам (какой).

Предметы присоединяются к процессам в точках примыкания *кто, из чего, чем, где, что, кому.*

Отношения (зависимости) связывают формулами свойства или предметы, или процессы.

Минимальной конфигурацией ситуации можно считать  $Sit \rightarrow \langle p \rangle \langle s \rangle$ , где  $p$  – процесс;  $s$  – предмет – субъект, выполняющий процесс.

Например: *Имеется стол.*

**Системные правила подстановки** – это правила, универсальные для всех предметных областей, в то время как предметные правила задает эксперт или конечный пользователь (менеджер).

Системные правила подстановки очевидны. Они выделяют детали, исходя из синтаксиса языка описания ситуаций. Например: что есть подлежащее, что – сказуемое, прилагательное и т.п., и, соответственно, определяют их место в семантической модели.

Вторую группу системных правил составляют правила плекс-грамматики. Они рекурсивно собирают модель ситуации из деталей.

$$\langle \textit{Situation} \rangle \rightarrow \langle p \rangle \langle s \rangle \textit{ (кто – для чего)}, \quad (1)$$

$$\langle \textit{Situation} \rangle \rightarrow \langle a \rangle \langle \textit{Situation} \rangle \textit{ (где – какой)} \mid \langle a \rangle \langle \textit{Situation} \rangle \textit{ (чье – как)}, \quad (2)$$

$$\langle \textit{Situation} \rangle \rightarrow \langle s \rangle \langle \textit{Situation} \rangle \textit{ (для чего – что)} \mid \langle s \rangle \langle \textit{Situation} \rangle \textit{ (для чего – из чего)} \mid \langle s \rangle \langle \textit{Situation} \rangle \textit{ (для чего – чем)} \mid \langle s \rangle \langle \textit{Situation} \rangle \textit{ (для чего – где)} \mid \langle s \rangle \langle \textit{Situation} \rangle \textit{ (для чего – кому)} \mid \langle s \rangle \langle \textit{Situation} \rangle \textit{ (для чего – из чего)} \mid \langle s \rangle \langle \textit{Situation} \rangle \textit{ (для чего – чей)} \mid \quad (3)$$

$$\langle \textit{Situation} \rangle \rightarrow \langle r \rangle \langle \textit{Situation} \rangle \textit{ (что, с чем – чье}_{ij}, \text{ чье}_{kl})} \mid \langle r \rangle \langle \textit{Situation} \rangle \textit{ (что, с чем – s}_j, s_l, j \neq l} \mid \langle r \rangle \langle \textit{Situation} \rangle \textit{ (что, с чем – p}_j, p_l, j \neq l} \mid \langle r \rangle \langle \textit{Situation} \rangle \textit{ (что, с чем – s}_j, p_l), \quad (4)$$

$$\langle \textit{Situation} \rangle \rightarrow \langle p \rangle \langle r \rangle \langle \textit{Situation} \rangle \textit{ (когда H, когда K – что, с чем – p}_i, p_s), \quad (5)$$

где  $p$  – процесс,  $s$  – предмет;  $a$  – свойство;  $r$  – отношение.

Первое правило соединяет процесс и предмет через их точки примыкания "кто" (делает), "для чего" (предназначен) соответственно.

Второе правило добавляет к ситуации свойство, соединяя точку примыкания "чье" свойства с точкой примыкания "какой" предмета или процесса.

Третье правило присоединяет к ситуации новый предмет с помощью его точки примыкания "для чего" и точки примыкания процесса "что" или "из чего", или "чем", или "где", или "кому", или точек примыкания другого предмета "из чего", или "чей".

Отношение как метод вычисления значений связывает значение неизвестного параметра одного объекта относительно значения однотипного параметра другого объекта. Напомним, что формулой любого отношения может быть константа.

Четвертое правило показывает, как отношение связывает  $i$ -тое свойство  $j$ -того объекта с  $k$ -тым свойством  $l$ -того объекта (рис. 3а) или связывает объекты в целом в случае, если это отношения родства, толерантности, сопутствия, сопоставления (сходства) и некоторые другие (рис. 3б, 3в).

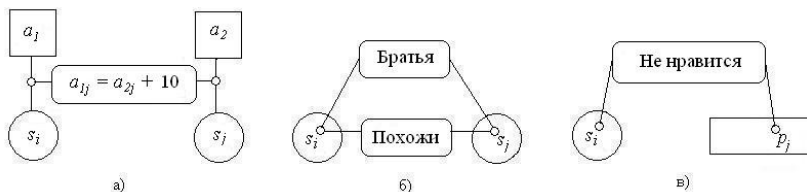


Рис. 3. Примеры подстановки отношений

Пятое правило присоединяет к ситуации новый процесс. Правило показывает, что вначале нужно связать его во времени с какими-то имеющимися процессами с помощью отношения следования. Другие связи нового процесса построятся рекурсивно с помощью правил (2), (3), (4).

**Предметные правила подстановки** возникают следующим образом. В базе знаний ПрО для всех точек соединения каждого процесса имеются множества взаимозаменяемых элементов (это знания экспертов, которые можно извлекать из них постепенно). Тогда везде, где вместо идентификатора записан  $q_0$ , можно подставить элемент из соответствующего множества взаимозаменяемых. Например, *экскаватор*  $\rightarrow$  *лопата*.

Пусть имеется некоторая обобщенная (производственная) ситуация *SituationO*. Обозначим  $\mathcal{S}_{qi(p)}$  нечеткое множество взаи-

мозаменяемых участников процесса  $p$  в точке примыкания  $q_i \in Q$ .

Нечеткое множество  $\underline{S}$  – это множество пар  $\langle x, \mu(x) \rangle$ , где  $\mu(x): S \rightarrow M$ ,  $M = \{0, 1\}$  – функция, которая отображает  $x$  на единичный отрезок  $M$ , определяя степень принадлежности элемента  $x$  множеству  $S$ . Предметное правило подстановки имеет вид:

$$\text{Situation } O \wedge (q_i(p) = q_0) \wedge (\underline{S}_{q_i(p)} \neq \emptyset) \Rightarrow q_i(p) = \langle x, \mu(x) \rangle,$$

т.е. элемент  $x$  может быть подставлен вместо отсутствующего элемента со степенью адекватности  $\mu(x)$ .

На практике функция  $\mu(x)$  заменяется коэффициентом приоритета, задаваемым экспертом, а количество порождаемых моделей ситуаций ограничивается возможностями компьютера, либо система сама ранжирует СеМС и предоставляет пользователю несколько наилучших по критериям прагматики. Поскольку эти подстановки генерируются как цепочка следствий, то переборы в пределах множеств взаимозаменяемости постепенно сужаются.

Тогда процесс *самообучения* системы можно представить как процесс формирования гиперграфов взаимозаменяемости с нечеткими множествами в качестве ребер.

**Пример 1.** Ёж (Пришвин М. Ёж. – Сб. рассказов "Зеленый шум". – М.: "Правда", 1983). *Раз шел я по берегу нашего ручья и под кустом заметил ежа. Он тоже заметил меня, свернулся и зашипел. Я прикоснулся к нему кончиком сапога, он страшно фыркнул и поддал своими иголками в сапог.*

На рисунке 4 показана семантическая модель (схема) сюжета, составленная с помощью плекс-грамматики из символов процессов, предметов, свойств и отношений. Пунктиром здесь обозначены связи между одноименными (копируемыми) элементами, описанными в разных предложениях. Эти связи устанавливаются правилами языка текста. Неиспользованные точки примыкания не прорисованы, но при желании они могут быть связаны с подходящими экземплярами из аналогичных ситуаций (см. далее).



и другие процессы, выражаемые в естественном языке ментальными глаголами. Например, "Лучше один раз увидеть, чем сто раз услышать" выражает степень достоверности обсуждаемой характеристики некоего предмета.

Поскольку понятие "ментальность" не присуще компьютеру, заменим его на виртуальность – мнимость, искусственность, нереальность.

Виртуальность высказывания запишем следующим образом [Кучуганов, 2011]:

$$V \begin{matrix} \mu \alpha_1 \nu \\ \mu \alpha_2 \nu \\ \dots \\ \mu \alpha_n \nu \end{matrix} (A) \text{ или } \mathfrak{N} \{ \alpha_1, \alpha_2, \dots, \alpha_n \} (A),$$

читается как "постольку, поскольку  $\alpha_1, \alpha_2, \dots$ ", то  $A$ , где  $\alpha_1, \dots, \alpha_n$ , – лингвистические переменные [Zadeh, 1997], [Заде, 2001], принимающие нечеткие значения соответствующих параметров отношения высказывания к действительности. Иначе говоря, виртуальность – это **вектор** в пространстве лингвистических переменных, характеризующий отношение между высказыванием и действительностью.

Вектор виртуальности умножается на значение истинности высказывания и, тем самым, корректирует влияние ментальности эксперта на принятие решений. Естественно-языковое предложение, как правило, описывает несколько фактов. Виртуальность может относиться к отдельному факту, предложению или рассказу в целом.

## 6. Оценка адекватности семантических моделей

Построив семантическую модель мы получаем нечеткий граф, вершинами которого являются нечеткие множества, а параметры вершин описаны лингвистическими переменными.

Если иметь в виду технические тексты, то оценить адекватность семантической модели (СеМ) реальной информации, поступившей на вход системы, можно путем сопоставления СеМ имеющимися в базе опыта. Исходя из предположения, что новая



СеМ будет составлена из известных элементов, в которых изменены какие-то свойства, отношения, взаимосвязи, достаточно установить соответствие с известными, выявить отличия и проверить корректность построенной модели. Заметим, что если хотя бы один параметр отличается от аналогов, то в технике это либо изобретение, либо чрезвычайная ситуация, а в искусстве – это "что-то интересное".

Будем называть *образом объекта* ту часть информации о нем, которая необходима и достаточна для решения поставленной задачи. Граф образа объекта  $Q^u$  порядка  $u$  – это граф

$$G^u = G(V, E, A),$$

где:  $V$  - множество вершин  $g$ , отображающих подобъекты  $q^{u-1}$ ;

$E$  - семейство ребер  $e \in E = (g_i^{u-1} R_{ij}^{u-1} g_j^{u-1})$ ,  $g_i^{u-1}, g_j^{u-1} \in V$ , отображающих отношения между образами подобъектов;

$A = (AH, EH)$  - гиперграф атрибутов  $a_{i1}(g_i), \dots, a_{im}(g_i) \in AH$  вершин и ребер  $eh \in EH$ , которые показывают принадлежность атрибутов из множества  $A$  некоторой вершине  $g_i \in V$ .

Для реализации подхода нам потребуются операции с образами [Кучуганов, 1985], [Кучуганов, 2011]:

1) абстрагирование – отключение параметров, не существенных для задачи;

2) идеализация образа – замена нескольких параметров одним обобщенным;

3) аморфное преобразование образа (масштабирование) с сохранением внутренних и внешних систем отношений. Например, размерная параметризация геометрических моделей в САПР. Например, "Лето – это маленькая жизнь" (Олег Митяев). В технике такой прием используется при "заимствовании", например, киль самолета → киль яхты;

4) сопоставление (сравнение) образов – выявление сходства и отличий;

5) сшивание образов – стыковка подграфов известных образов для получения графа нового образа.

6) визуализация – перевод вербального в образное.

Рассмотрим подробнее некоторые из них.

### **Сравнение образов.**

Пусть  $G_l = G(V_l, E_l, A_l)$  - граф известного (эталонного) образа;

$G_2 = G(V_2, E_2, A_2)$  - граф анализируемого образа;  
 $r = \langle A', A'' \rangle$  - гребенка (набор) анализируемых признаков,

где

$A'$  - множество имен признаков, определяющее *степень абстракции*,

$A''$  - множество коэффициентов, определяющих *точность сравнения* признаков как  $da_i = a_{1i} * a''_k$ .

Будем называть *общей гомоморфной частью*  $\tau(G'_1)$  графов  $G_1$  и  $G_2$  часть  $G'_2$ , определенную на множестве  $V'_2$  и состоящую из всех ребер  $\tau(E_1) = E_2 = (g_{2i}, g_{2j}), g_{2i}, g_{2j} \in V'_2 \subseteq V_2$ , для которых существуют соответствующие ребра

$$E_1 = (g_{1i}, g_{1j}) = \tau^{-1}(E_2) = (\tau^{-1}(g_{2i}), \tau^{-1}(g_{2j})), g_{1i}, g_{1j} \in V_1,$$

в графе  $G_1$ , при этом однозначное отображение  $\tau$  существует тогда и только тогда, когда вершины  $g_{2i}, g_{2j}$  и соответствующие им прообразы  $g_{1i}, g_{1j}$  совпадают в  $n$ -мерном пространстве  $A'$  анализируемых признаков с точностью, заданной  $A''$ .

*Оценка сходства образов*, взвешенная по  $k$ -тому свойству, есть

$$(1) \quad C = \frac{\sum_{j=1}^n L_k(D'_{2j})}{\sum_{i=1}^n L_k(D_{1i})},$$

где  $L_k(D_{1i})$  и  $L_k(D'_{2j})$  – протяженность всех несоприкасающихся цепочек в эталоне и совпавших в анализируемом графе (соответственно), вычисленная с помощью суммируемых по цепочке значений  $k$ -того свойства вершин, что автоматически придает больший вес совпадению тех элементов объектов, которые в соответствии с заданной гребенкой признаков идентификации считаются более существенными.

Таким образом оценка адекватности становится зависимой от прагматики приемника информации и, как показано выше, от вектора виртуальности.

### **Сшивание образов.**

Пусть графы  $G_1$  и  $G_2$  образов имеют некоторую общую гомоморфную часть  $\tau(G'_1)$ , найденную с помощью операции сравнения образов и выражающую их сходство.

Обозначим через  $H_2$  общую гомоморфную часть  $\tau(G'_1)$ , а ее дополнение в графе  $G_2$  – через  $\bar{H}_2$ . Прообраз общей гомоморфной части обозначим через  $H_1$ , а его дополнение в  $G_1$  – через  $\bar{H}_1$ .

Операция сшивания графов образов заключается в следующем:

1. Объединение  $G_{1P}$  с частью  $H_2$ :  $G_{1P} \cup \tau_2(H'_1)$ .

2. Сшивание части  $H_1$  с частью  $\tau_2(\bar{H}'_2)$  путем построения части  $H^c = G(E^c)$  так, что

$$\exists e_2 = (\tau(g_1), \tau_2^{-1}(\tau_2(g_2))) \in G_{2P} \Rightarrow \tau_2(e_2) \in H^c.$$

Тогда, граф  $G_3 = G_1 \cup \tau(\bar{H}'_2) \cup H^c$  отображает новый образ, который включает в себя два известных подобраза.

## 7. Ассоциативные подстановки

Обычно при описании знаний строят *семантические сети*, которые устанавливают атрибутивные и ролевые связи между сущностями.

*Атрибутивные связи* (кто – какой, кто – чей) описывают свойства объектов, принадлежность, родство и другие их характеристики.

*Ролевые связи* (кто – кому – что – как – для чего – где – когда и т.д.) устанавливаются при выявлении отношений процессуальных, системных, коммуникации, положения, следования, в результате чего строится *концептуальная модель* процесса.

Помимо вышеперечисленных, в ходе обучения и накопления опыта возникают *ассоциативные связи* – по сходству сущностей или их атрибутов или их отношений. Сходство может оцениваться по самым разнообразным критериям, которые устанавливаются жизненным опытом индивидуума, зависят от его психофизических характеристик. Психологи различают 3 вида ассоциаций: по сходству; по смежности; по контрасту.

- Что ли, ножик – вилкин муж? (К.И.Чуковский. От двух до пяти. Собр. соч. в 6 томах. Том 1. – М.: Худ. лит., 1965) – по-ви-

димому, из часто встречающейся ситуации, в которой нож и вилка сопутствуют друг другу, при том, что нож мужского рода, а вилка – женского, у девочки возникает модель семейной пары и теперь она спрашивает подтверждения у взрослых.

Ассоциативная связь фиксируется как факт наличия ненулевого значения при проверке отношения сопоставления. Вес этой связи учитывается при распознавании и оценке сходства образов ситуаций.

В процессе формирования семантической модели текста, описывающего ситуацию, ранее мы пользовались системными правилами – правилами грамматики языка текста и правилами подстановки, определенными плекс-грамматикой семантической модели, а также предметными (экспертными) правилами подстановки, исходящими из взаимозаменяемости реальных элементов (деталей) ситуации. Подобрав с помощью оценки сходства аналоги решаемой ситуации, можно путем сшивания графов подставить за один шаг сразу всё, что подходит, что полезно и не очень.

Правило подстановки на основе предметной взаимозаменяемости также можно усилить путем ассоциативной подстановки. Зачастую заменяющие элементы бывают связаны друг с другом. Например, от материала изделия зависит оборудование, инструмент и способ изготовления. Операция сшивания образов подставит сразу всю подборку.

Таким образом, *логика процессуальных отношений* – это вычисление истинности (качества) процессуальных отношений. Она представляет каждый процесс как **систему отношений**

*кто – что – из чего – чем – как – кому – для чего – где – при каких обстоятельствах – когда – сколько.*

Вычисление оценок (например, "хорошо", "удовлетворительно", "плохо") каждой составляющей (см., например, [Ивин, 1997], [Тарасов и др., 2010]), а также процессов, в результате которых появились "плохие" составляющие, позволит за несколько шагов по ассоциативной цепочке придти к первопричине возникновения неудовлетворительной ситуации. Тогда, чтобы получить ответ на извечный вопрос "*что делать?*", надо выполнить некоторые преобразования и подстановки в **графе процессуальных отношений** задачи.

## Пример 2. Поездка. "Я выхожу на следующей остановке".

1. Для каждого субъекта (слова, словосочетания), выделенного из текста в ходе морфологического и синтаксического анализа, находим в онтологии область существования:

*Я* – человек (предмет одушевленный, местоимение).

*Выходить* [из] – процесс:

- а) движение – перемещение с места на место (откуда/куда: комната, дом, транспорт, лес);
- б) комплектование – исключение/включение в состав участников процесса (откуда/куда: игра, бизнес);
- в) событие – смена состояния (откуда: кризис, ...).

*Остановка:*

- а) сооружение – место, специально оборудованное для кратковременной приостановки движения транспорта с целью входа/выхода пассажиров и (или) груза;
- б) движение – фаза процесса.

2. Каждый из вариантов толкования субъекта текста добавляет в рабочую модель свою информацию – объект или атрибут или отношение:

2.1. Прилагательное '*следующая*' – означает, что надо рассматривать по меньшей мере две остановки. В списке предметов появляется остановка следующая.

2.2. '*Остановка*' принадлежит двум областям (предмет и фаза движения).

Остановке как предмету *сопутствует* предмет '*транспорт*', а транспорту – '*груз*'. В таблицу предметов добавляется транспорт (груз уже есть).

В то же время, остановка является частью процесса '*Движение*', поэтому в список процессов добавляется пятый – движение и шестой элемент – '*останавливаться*', связанный со следующей остановкой.

3. Формирование списка отношений между процессами и участниками.

4. Уточнение параметров времени. Поскольку некто '*Я*' выходит на второй остановке, то длительность действия '*выходить*' можно приравнять к длительности остановки. Соответственно, совпадает и их время. Остаются неизвестными время и длительность движения, но это можно уточнить только если в

базе знаний есть информация о маршрутах транспорта. Кроме того, если считать, что анализируемое высказывание (как правило) произносится где-то в начале движения, то, очевидно, время глагола *'выхожу'* (настоящее) должно быть скорректировано на будущее.

В результате, с помощью ассоциативных связей базы знаний уточняется семантическая модель текста. В нашем примере в модели сюжета:

1) появились "подразумеваемые" предметы *остановка* (предыдущая) и *транспорт*, а также процессы *остановка – движение – остановка*;

2) установилась система отношений, связывающая между собой все субъекты, кроме второго и третьего вариантов толкования процесса *"выходить"*.

**Пример 3. Задача про Ивана-царевича** [Лихтарников, 1996]. *Собрался Иван-царевич на бой со Змеем Горынычем, трехглавым и треххвостым. "Вот тебе меч-кладенец, – говорит ему баба Яга. — Одним ударом ты можешь срубить либо одну голову, либо две головы, либо один хвост, либо два хвоста. Но запомни: срубишь один хвост — два вырастут, срубишь два хвоста — голова вырастет, срубишь голову — голова вырастет, срубишь две головы — ничего не вырастет". За сколько ударов Иван-царевич может срубить Змею все головы и хвосты?*

После сокращения (реферирования) семантической модели путем отбрасывания элементов, не влияющих на решение, условные задачи не трудно представить в табличной форме (табл. 1-3), удобной для передачи решателю [Кучуганов и др., 2010]. Предметные (задачные) отношения – формулы, ограничения, правила формируются с помощью толкового словаря и грамматики языка текста.

*Таблица 1. Предметы*

Ситуация	Наименование	Головы, шт.	Хвосты, шт.
Начальная	Змей Г.	3	3
Целевая	Змей Г.	0	0

Таблица 2. Действия

Наименование	Кто	Чем	Как	Что	Сколько	Кому
Срубить	Иван-царевич	Меч	Вы- честь	?	?	Змей Г.
Вырастать	-	-	При- бавить	?	?	Змей Г.

Таблица 3. Отношения

№ п/п	Формула
1	<i>Срубить.Что IN [Голова, Хвост]</i>
2	<i>Срубить.Сколько IN [1, 2]</i>
3	<i>(Срубить.Что = 'Хвост') ∧ (Срубить.Ск = 1) ⇒ (Вырастать.Что = 'Хвост'), (Вырастать.Ск = 2)</i>
4	<i>(Срубить.Что = 'Хвост') ∧ (Срубить.Ск = 2) ⇒ (Вырастать.Что = 'Голова'), (Вырастать.Ск = 1)</i>
5	<i>(Срубить.Что = 'Голова') ∧ (Срубить.Ск = 1) ⇒ (Вырастать.Что = 'Голова'), (Вырастать.Ск = 1)</i>
6	<i>(Срубить.Что = 'Голова') ∧ (Срубить.Ск = 2) ⇒ (Вырастать.Что = nil)</i>

## 8. Заключение

Таким образом, процесс синтеза семантической модели текста, описывающего ситуацию или сюжет, мы интерпретируем как процесс вывода корректной означивающей конструкции, адекватной человеческому восприятию текста.

Оценка адекватности семантической модели осуществляется путем проверки грамматической корректности модели, установления соответствия между элементами новой Сем и имеющимися в базе системы, а также выявления отличий в свойствах, отношениях и связях новых элементов. В то же время, поскольку плекс-грамматика, как и другие графовые грамматики, весьма удобна для визуального отображения, в особо важных случаях

лицо, принимающее решение, лично проверяет автоматически построенные модели.

Оценка адекватности семантической модели позволяет:

- оценить степень новизны ситуации;
- проверить корректность формулировки задачи;
- пополнить проблемную ситуацию информацией, подразумеваемой, но не заданной в явном виде или достроить (доопределить) ситуацию по аналогии с известными.

Корректные семантические модели ситуаций можно рассматривать как способ извлечения знаний из экспертов и из базы опыта, накопленного в предметной области. Образно говоря, происходит ассоциативная укладка деталей в гештальт – целостный образ, вектор так же, как из мозаичного набора можно сложить множество картин – гештальтов, в зависимости от ментальности автора. И так же восприятие картины как целостного образа зависит от ментальности зрителя.

В целом, на наш взгляд, инструментарий ассоциативной семантики повысит качество коммуникаций "человек-машина".

### **Литература**

1. FEDER J. *Plex languages*. // *Information Sci.*, 1971. - №3, P. 255-241.
2. FU K. S. *Syntactic Methods in pattern recognition*. // Academic Press. New York and London, 1974. Русский перевод: Фу К. Структурные методы в распознавании образов. – М.: Мир, 1977.
3. ZADEN L. A. *Toward a Theory of Fuzzy Information Granulation and its Centrality in Human Reasoning and Fuzzy Logic* // *Fuzzy Sets and Systems*. – 1997. Vol. 90. – P. 111-127.
4. ЗАДЕ Л. А. *Роль мягких вычислений и нечеткой логики в понимании, конструировании и развитии информационных / интеллектуальных систем*. – *Новости Искусственного Интеллекта*, №2-3, 2001. С. 7-11.
5. ИВИН А. А. *Основы теории аргументации*. – М.: ВЛАДОС, 1997.



6. КУЧУГАНОВ В. Н. *Автоматический анализ машиностроительных чертежей*. – Иркутск: Изд-во Иркутского ун-та, 1985. – 112 с., ил.
7. КУЧУГАНОВ В. Н. *Вербализация реальности и виртуальности. Ассоциативная семантика*. – Искусственный интеллект и принятие решений, 2011. – №1. – С. 95-106.
8. КУЧУГАНОВ В. Н., СЕНТЯКОВА А. В. *Интеллектуальный тренажер по программированию*. – Труды Конгресса по интеллектуальным системам и информационным технологиям "AIS'10". Научн. издание в 4-х томах. – М.: Физматлит, 2010. – Т.1. – С. 534-541.
9. ЛИХТАРНИКОВ Л. М. *Занимательные логические задачи*. (Для учащихся начальной школы) / Оформление С. Григорьева — Спб.: Лань, МИК, 1996. — 125 с.
10. ТАРАСОВ В. Б., КАРЛУЦКАЯ А. П. *Нечеткие лингвистические модели предпочтений когнитивных агентов*. – Труды Конгресса по интеллектуальным системам и информационным технологиям "AIS'10". Научн. издание в 4-х томах. – М.: Физматлит, 2010. – Т.2. – С. 277-284.

## ELEMENTS OF ASSOCIATIVE SEMANTIC THEORY

**Valery Kuchuganov**, Izhevsk State Technical University, Izhevsk, Doct. Sc., Head of Automated Information Processing and Control Systems Department, Professor (kuchuganov@istu.ru)

*Abstract: In the paper the tools are proposed for reducing the ambiguousness of semantic models of statements, thus increasing the tolerance of the man-machine dialog and the effectiveness of knowledge acquisition systems. The semantic model of the situation, the story is described by a plex grammar containing 4 sorts of symbols with various number of attaching points. The process of semantic text model synthesis is considered as a process of the inference of a correct denotative structure adequate to the human text perception. Operations of matching and joining of pattern graphs are proposed. The adequacy of the semantic model is assessed by checking its grammatical validity, matching the elements of the new model to the*

*ones stored in the base of the system, as well as detecting the differences between the properties, relations and links of the new elements.*

Keywords: semantic model, plex grammar, pattern graph, associative semantics, virtuality vector, adequacy assessment