

О ПРЕДСТАВЛЕНИИ РЕЗУЛЬТАТОВ СЕМАНТИКО-СИНТАКСИЧЕСКОГО АНАЛИЗА ТЕКСТА КОНЦЕПТУАЛЬНЫМИ МОДЕЛЯМИ ДАННЫХ¹

Г.В.Лезин,* В.А. Тузов**

*Санкт-Петербургский экономико-математический институт РАН
lezin@emi.spb.su

**Санкт-Петербургский государственный университет
spbuoasis3@peterlink.ru

1. Введение

В работе рассматриваются предварительные результаты исследований методов взаимодействия двух информационных систем – семантико-синтаксического анализатора текста на русском языке (СС-анализатора) и базы знаний, содержащей описание той или иной конкретной предметной области и предназначенной для накопления извлекаемой из текста информации.

СС-анализатор осуществляет локальный анализ предложений текста и для каждого из предложений строит его семантико-синтаксическую модель. На этапе СС-анализа решается проблема семантической неоднозначности слов предложения (по крайней мере, для "правильно построенных" и однозначно трактуемых предложений). Каждому из слов предложения сопоставляется его значение в виде описания соответствующей лексемы-функции. Предложение в модели представлено суперпозицией лексем-функций. Аргументами лексемы-функции служат лексемы, связанные с данной подчинительными связями. Значением лексемы-функции в предложении является сущность (объект действительности), обозначенная соответствующим элементом предложения.

Ядром СС-анализатора является семантический словарь русского языка. Словарные статьи словаря содержат семантико-синтаксическое описание лексем. Каждая из статей имеет два раздела – синтаксический и семантический. В синтаксическом разделе размещена информация, необходимая для выполнения СС-анализа предложения. В семантическом разделе приведена семантика лексемы в виде формулы, задающей способ вычисления значения лексемы-функции по значениям ее аргументов. СС-анализатор разработан одним из авторов статьи [1].

Возможны, по-видимому, различные интерпретации формул, описывающих семантику лексем. Мы в этой работе рассмотрим интерпретацию семантических формул и семантико-синтаксических моделей предложений в терминах концептуальных моделей данных. Рассматриваемая в работе концептуальная модель данных реализована в системе управления базой знаний MAZE. Общее описание языка представления знаний в системе MAZE (М-языка) и примеры его использования для моделирования текстов можно найти в [2].

2. Язык представления знаний

Наша цель – использовать базу знаний MAZE как средство для обработки результатов СС-анализа текста и извлечения из текста содержащейся в нем информации, и одна из первых задач здесь – интерпретация семантико-синтаксической модели на языке базы знаний. В этом параграфе мы вкратце рассмотрим основные элементы М-языка.

¹ Работа выполнялась при поддержке РФФИ, грант № 00-01-00443

2.1. Объекты и отношения. Формальное сообщение, посылаемое в базу знаний, записывается последовательностью предложений М-языка. Предложение может содержать утверждение об отношениях между объектами концептуальной модели, иметь форму правила, устанавливающего зависимости между отдельными классами утверждений, либо представлять собой запрос к базе знаний.

Объектами в концептуальной модели представляются сущности моделируемой предметной области. Допускается определение следующих видов объектов: классы (понятия, множества), экземпляры понятий, бинарные отношения. Вид объекта явно указывается в его обозначении. Например, обозначения понятий записываются прописными буквами, а экземпляров – строчными.

Текущее состояние объектов базы знаний описывается наборами предикатов. Для обозначения переменных в предикатах мы зарезервируем буквы X, x, Y, y, Z, z, и в случае необходимости будем индексировать эти буквы цифрами 1, ..., 9. Переменные, обозначенные прописными буквами X, Y, Z, определены на множестве понятий (классов) концептуальной памяти, а переменные, обозначенные строчными буквами x, y, z - на множестве экземпляров.

Факт принадлежности понятия E1 или экземпляра v понятию E в М-языке устанавливается отношением E:E1 и, соответственно, E:v. Далее мы будем говорить в таких случаях, что объект E1 *конкретизирует* объект E (является его подклассом), объект v – конкретный пример, экземпляр объекта E; само отношение называется отношением *конкретизации*. Факт E:E1, будучи зарегистрированным в базе знаний, задает значение переменных X=E, Y=E1, для которого предикат X:Y истинен. Предикат X:Y имеет вполне прозрачную операционную семантику [3]:

$$\forall(X, Y) [X:Y \leftrightarrow \forall(x) (Y:y \rightarrow X:x ; \neg (X:x) \rightarrow \neg(Y:y))] \quad (1)$$

В М-языке класс представляет в базе знаний множество элементов, имеющих общие *атрибуты*. Записью E(E1) мы утверждаем факт наличия атрибута E1 у объекта E. Утверждение E(E1,...,Ek) задает для объекта E список атрибутов. В качестве атрибутов могут указываться лишь классы объектов. Отметим, что список атрибутов никогда не считается завершенным и может пополняться по мере поступления в базу знаний новых сообщений. Текущее состояние списка атрибутов понятия E определяется предикатом E(X). Предикат X(E) позволяет установить наличие в базе знаний классов, а предикат x(E) — экземпляров, для которых E определено в качестве атрибута. Для каждого экземпляра e из класса E(A) существует вполне конкретный экземпляр a атрибута A. Для задания значений атрибутов в языке предусмотрено трехместное отношение вида e(A.a). Распространяя это отношение на классы, мы получаем отношение вида X(Y.Z) с операционной семантикой

$$\forall(X, Y) [X(Y.Z) \leftrightarrow \forall(x) (X : x \rightarrow x(Y) ; \exists(z) (x(Y.z) ; Z : z ; Y : z))] \quad (2)$$

В базе знаний текущее состояние класса E в отношении X(Y.Z) определено значениями предикатов E(Y.Z), X(E.Z) и X(Y.E).

Семантика атрибута – двойственна:

- атрибут объекта задает класс допустимых значений определяемого им свойства объекта;
- атрибут является бинарным отношением, связывающим объект со значением свойства.

2.2. Правила. В М-языке логика зависимостей между объектами и отношениями конкретной предметной области задается правилами. Структура правила после преобразования его записи на исходном языке в представление для базы знаний имеет вид:

$$\langle \text{конъюнкция предикатов} \rangle \Leftarrow (\langle \text{конъюнкция предикатов} \rangle)$$

Элемент конъюнкции в правиле - предикат того или иного отношения. Как обычно, все переменные, упомянутые в следствии правила, должны быть использованы также и в каждом из дизъюнктов его посылки. Правила, наряду с фактами, регистрируются в базе знаний.

2.3. Концептуализации отношений. Классические концептуальные модели данных [5,6] статичны и описывают отношения, "свойственные" объектам. Отношения в этих моделях не зависят от времени, источника сообщения или каких-либо других обстоятельств, ограничивающих область действия того или иного отношения. В реальных же текстах, как правило, описывается череда событий или явлений, в рамках которых отношения между объектами могут возникать и исчезать, приводятся мнения различных источников информации об этих событиях, часто противоречивые, - в общем, тексты содержат информацию, в которой отношения между объектами локализованы, привязаны к тем или иным обстоятельствам. Для решения этих задач концептуальная модель данных в М-языке дополнена специальным видом отношений – отношениями концептуализации.

Терминальным отношением концептуализации называется отношение вида X{R(Z)}, где X – класс объектов, R – одно из основных отношений языка (например, рассмотренные выше конкретизация, объект-атрибут, объект-атрибут-значение), Z - набор переменных отношения. Смысл этого отношения мы определим следующим образом:

а) Предикаты концептуализаций $X\{R1(Z)\}$ и $X\{R2(Z)\}$ истинны тогда и только тогда, когда истинен предикат концептуализации $X\{R1(Z);R2(Z)\}$, где $(R1(Z); R2(Z))$ - конъюнкция предикатов терминальных отношений $R1(Z)$ и $R2(Z)$. Концептуализация, т.о. представляет собой конъюнкцию предикатов терминальных отношений, сопоставленную некоторому объекту и истинную только в отношении с этим объектом.

б) Операционная семантика концептуализации:

$$\forall(X,Z)[X\{R(Z)\} \leftrightarrow \forall(x) (X:x \rightarrow x\{R(Z)\})] \quad (3)$$

Т.о., концептуализация класса распространяется на все элементы этого класса.

в) Отношение вида $X\{Y\}$, где X и Y - классы объектов, позволяет описать иерархическую структуру концептуализаций. Его операционная семантика:

$$\forall(X,Y,Z)(X\{Y\} \leftrightarrow X\{R(Z)\} \rightarrow Y\{R(Z)\}), \quad (4)$$

где $R(Z)$ - конъюнкция терминальных отношений, заданных концептуализацией класса X .

Включив объект Y в концептуализацию объекта X мы распространяем область истинности предикатов $R(Z)$ концептуализации X на концептуализацию Y .

Примеры использования М-языка мы рассмотрим в следующем разделе, разбирая методы построения описаний лексем в семантическом словаре.

3. Интерпретация описаний лексем

3.1. Классификация лексем. Основой описания формальной семантики слов в семантическом словаре является классификатор основных понятий русского языка. Полное множество лексем словаря разбито на два подмножества: подмножество базовых и подмножество производных лексем. Все лексем из подмножества базовых разнесены по семантическим классам, причем базовая лексема не может относиться более, чем к одному классу. Слово может иметь несколько различных значений (лексем). Уникальное обозначение базовой лексем в словаре образовано парой <слово><класс>. Производные лексем описываются на основе базовых и не вносятся в классификатор. Для производных лексем нет однозначного соответствия <класс><лексема>, и уникальное обозначение производной лексем приходится задавать парой <слово><номер>, указывая порядковый номер лексем в списке значений слова.

Классификатор имеет иерархическую структуру с отношением "один" ← "многие". Цепочка цифр в обозначении класса задает его место в общей иерархии классификатора.

Пример 1.

Лексеме ВРУЧЕНИЕ \$15210 соответствует следующее место в иерархии классов: "существительное" (класс \$1) ← "действие" (\$15) ← "занятие" (\$1521) ← "приобретение" (\$15210), а лексеме ОСМАТРИВАНИЕ\$15403 - "существительное" (класс \$1) ← "действие" (\$15) ← "движение" (\$1540) ← "осматривание"(\$15403). Обе лексем принадлежат общему классу \$15 "существительное" ← "действие".

По принципу классификации базовые лексем поделены на две категории: объекты и их свойства. Каждому из классов может быть сопоставлена собственная иерархия свойств.

Пример 2.

Лексем СТАРЫЙ\$12411/071 и ЮНЫЙ\$12411/071 принадлежат общему классу "существительное" (\$1) ← "физический объект" (\$12) ← "живой"(\$124) ← "человек" (\$1241) ← "индивид"(\$12411) и относятся к общему классу его (класса \$12411) свойств - "возраст" (/07) ← "молодой-старый" (/071).

В этой работе мы интерпретируем обозначение лексем как обозначение класса, элементами которого являются внеязыковые сущности, составляющие (наряду с отношениями между ними) содержание исходного текста. Выявление этих сущностей в тексте и установление связей между ними является основной задачей семантического анализа текста в целом (связного текста). Тем не менее, в словаре имеется достаточно большое количество слов, обозначающих вполне конкретные сущности (имена стран, организаций, конкретных людей).

3.2. Виды семантических описаний. В простейшем (и достаточно распространенном) случае семантическим описанием лексем является только указание класса, которому она соотнесена, т.е. задано ее обозначением. На М-языке такое описание представляется отношением конкретизации:

$$\langle \text{обозначение лексем} \rangle : \langle \text{переменная} \rangle, \quad (5)$$

где <переменная> определена на классе сущностей, обозначаемых лексемой.

В других случаях вид описания определяется характером взаимодействия описываемой лексемы с другими, попадающими в сферу ее действия в предложении. Описания лексем в этих случаях представлено в словаре в виде функций над значениями взаимодействующих лексем. В записи функций, описывающих семантику лексем, используются:

а) Базисные функции, априорно заданные в семантической модели русского языка, лежащей в основе семантического словаря [1]².

б) Ссылки на лексем, на основе которых толкуется определяемая.

Полный список используемых в словаре базисных функций образован 33 унарными и 21 бинарной функцией. Функции определены на множестве внеязыковых сущностей, так или иначе упоминаемых в тексте.

Примеры бинарных функций:

- 1) Caus(x,y) - x является причиной, поводом, инициализирующим фактором события y - x "каузирует" y.
- 2) Lab(x,y) - x подвергается воздействию y.
- 3) Hab(x,y) -x имеет свойство или состояние y, y "в распоряжении (временном)" x.
- 4) Usor(x,y) - x применяется (используется, предназначается) для y.
- 5) Loc(x,y) - x находится, располагается в y.
- 6) Temp(x,y) - время x-а находится в интервале времени y-а.
- 7) Content(x,y) - x содержится в (является частью) y.

В качестве примера унарных функций мы приведем те из них, которыми характеризуются состояния сущностей:

- 1) Perf(x) - отмечает завершенность x-а.
- 2) Inscr(x) - отмечает начало нового состояния (действия, получения или проявления свойства).
- 3) Oper(x) - процесс выполнения действия, проявления свойства.
- 4) Fin(x) - отмечает прекращение существования, завершение действия или состояния.
- 5) Em(x) - способность проявить свойство, выполнить действие.
- 6) Func(x) - отмечает факт существования.
- 7) Cont(x) - отмечает постоянство действия, проявления свойства.

Мы интерпретируем базисные функции как бинарные отношения. Учитывая семантику (2) термина "атрибут" в М-языке, мы можем представить на этом языке бинарные базисные функции семантических формул как необязательные, "нейтральные" атрибуты одного из ее аргументов. Как правило, это – первый аргумент бинарной функции. "Нейтральность" бинарной функции как атрибута проявляется в отсутствии класса выделенных ее обозначением объектов: бинарные функции определены на всем множестве упоминаемых в текстах сущностей. Унарные функции при переводе на М-язык трактуются аналогично бинарным. Унарную функцию мы представляем как атрибут с неопределенным (размытым) константным значением, не требующим специального указания.

Выделяются следующие основные виды описаний семантики лексем:

а) Значение лексемы определяется ее актантами:

$$\langle \text{обозначение лексемы} \rangle (*\langle \text{аргументы} \rangle *) \quad (6)$$

Аргумент лексемы в (6) задается в общем случае тройкой <атрибут>.<класс>~<Z-переменная>. Указание атрибута, класса или обоих вместе могут опускаться. Порядок перечисления аргументов – произвольный и не влияет на ее интерпретацию. Место аргумента в интерпретации лексемы обозначено ее атрибутом. При ссылке на лексему можно указывать лишь некоторые из заданных для нее обязательных аргументов. Мы будем интерпретировать атрибут в описании лексемы в точном соответствии с его операционной семантикой в М-языке. Это означает, в частности, что Z-переменные, определенные на множествах значений соответствующих атрибутов связаны

² Автор модели языка на начальной стадии своих исследований находился под сильным впечатлением от работы Мельчука [4]. Многие базисные функции по обозначению совпадают с лексическими функциями Мельчука, но за время работы большинство обозначений приобрело существенно иное содержание.

квантором существования с множеством значений лексемы: получив из текста конкретное значение лексемы, мы автоматически получаем факт существования конкретных значений ее атрибутов (см. (2)).

Пример 3.

В словаре лексема ОСМАТРИВАНИЕ описывается как:

ОСМАТРИВАНИЕ \$15403 (Z1: !Тв, Z2: !Род)

s: (*СУБЪЕКТ.Z1, ОБЪЕКТ.Circ(Z2) *);

Словарная статья имеет 2 раздела: в первом описывается взаимодействие слова, обозначающего лексему с другими словами предложения, во втором (помеченном значком 's:') приводится ее семантическая формула. (Унарная функция Circ(x) трактуется как "окрестность, окружение" х.) Интерпретация формулы на М-языке представляется конъюнкцией отношений:

ОСМАТРИВАНИЕ\$15403:X; X(СУБЪЕКТ.Z1); X(ОБЪЕКТ.Z2); Z2(Circ);

б) Значением лексемы определяется свойство одного из ее актантов (обозначенного как Z0). Общий вид формулы:

<слово><класс-свойство>(* Z0, <аргументы> *) (7)

Мы интерпретируем эту лексему как атрибут Z0 (см.(2)), ее значение – как значение атрибута, а ее аргументы – как актанты значения атрибута.

Пример 4.

БЫСТРЫЙ\$153/01(Z0:a> НЕЧТО\$1)

s: (*НЕЧТО\$1~Z0*)

Интерпретация:

Z0(БЫСТРЫЙ \$153/01.X);

АКТИВНЫЙ\$1100/12(Z0:a> НЕЧТО\$1, Z1: !вПред)

s: (*НЕЧТО\$1~Z0, ТЕМА:Z1)

Интерпретация:

Z0(АКТИВНЫЙ \$1100/12.X); X(ТЕМА.Z1);

в) Множество значений лексемы определяется функцией ее толкования.

Пример 5.

ОКИНУТЬ N%ОКИДЫВАНИЕ\$15403(Z1: ЖИВОЙ\$124~!Им, Z2:!Род, Z3:ВЗГЛЯД\$124/00~!Тв)

s: PerfOper~ОКИДЫВАНИЕ (* СУБЪЕКТ.Usor(Z1, ВЗГЛЯД\$124/00~Z3), ОБЪЕКТ.Z2 *)

Множество значений лексемы ОКИНУТЬ описывается как совпадающее с множеством значений действия ОКИДЫВАНИЕ в состоянии завершенного выполнения, заданного комбинацией унарных базисных функций Perf и Oper. Интерпретация:

(ОКИДЫВАНИЕ\$15403:X; X(PerfOper); X(СУБЪЕКТ.Z1); Z1(Usor.Z3); ВЗГЛЯД\$124/00:Z3; X(ОБЪЕКТ.Z2); Z2(Circ));

г) Значения лексемы определяются как подмножество значений функции ее толкования. Общий вид формулы:

<слово><класс>(*<аргументы>*) (<функция толкования>) (8)

Описание лексемы в этом случае интерпретируется как правило (см. п.2.2), посылкой которого служит обозначение лексемы с заданными для нее аргументами, а следствием – функция толкования.

Пример 6.

ВРУЧЕНИЕ \$15210(Z1: !Тв, Z2: !Род, Z3: !Дат)

s: (*СУБЪЕКТ.Z1, ОБЪЕКТ.Z2, АДРЕСАТ.Z3*) (Caus_o(Z1, IncepHab(Z3, Z2)))

Интерпретация:

ВРУЧЕНИЕ\$15210:X; X(СУБЪЕКТ.Z1); X(ОБЪЕКТ.Z2); X(АДРЕСАТ.Z3)) ⇐
X(ConvCaus.Z1); X{Z3(IncepHab.Z2)}

Z1 каузирует новую ситуацию, в которой Z3 имеет Z2. Смена ситуаций представлена концептуализацией (см. п.2.3).
 $\text{ConvCaus}(X, Z1) = \text{Caus}(Z1, X)$ – конверсия отношения.

ВРУЧАТЬ N%-ВРУЧЕНИЕ\$15210(Z1: !Им, Z2: !Вин, Z3: !Дат)

s: Oper~ВРУЧЕНИЕ\$15210 (СУБЪЕКТ.Z1, ОБЪЕКТ.Z2, АДРЕСАТ.Z3)

Лексема ВРУЧАТЬ описана как производная от ВРУЧЕНИЕ.

Интерпретация:

ВРУЧЕНИЕ\$15210:X; X(Oper); X(СУБЪЕКТ.Z1); X(ОБЪЕКТ.Z2); X(АДРЕСАТ.Z3)

ОЦЕНИВАТЬ_ВЗГЛЯДОМ \$11016(N%~\$0(Z0: a> ВЗГЛЯД\$124/00, Z1: !Им, Z2: !Вин))

s: (*СУБЪЕКТ.Z1, АДРЕСАТ.Z2*) Oper~ОЦЕНКА\$11016(СУБЪЕКТ.Z1, АДРЕСАТ.Z2)

Интерпретация:

(ОЦЕНКА\$11016:Z0; Z0(Oper); Z0(СУБЪЕКТ.Z1); Z0(АДРЕСАТ.Z2)) \Leftarrow

(ВЗГЛЯД\$124/00:Z0; Z0(СУБЪЕКТ.Z1); Z0(АДРЕСАТ.Z2))

Лексемы СТАРИК и МАЛЬЧИК не имеют аргументов, но их описание относится к этому же виду:

СТАРИК \$12411~Z0: s> ЧЕЛОВЕК\$1241

s: СТАРЫЙ\$12411/071(Z0)

Интерпретация:

(Z0(СТАРЫЙ\$12411/071.X)) \Leftarrow \$12411СТАРИК:Z0;

МАЛЬЧИК \$12411~Z0: s> ЧЕЛОВЕК\$1241

s: Magn~ЮНЫЙ\$12411/071(Z0)

Интерпретация:

(Z0(ЮНЫЙ\$12411/071.X); X(Magn)) \Leftarrow \$12411МАЛЬЧИК:Z0;

4. Модели предложения

В предыдущих разделах мы рассмотрели интерпретацию семантических формул словаря в терминах концептуальных моделей данных. В результате мы получаем возможность использования базы знаний в процессе анализа связного текста и автоматического извлечения из него той или иной информации.

4.1. Семантико-синтаксическая модель предложения. Результат СС-анализа предложения представлен суперпозицией лексем-функций и описанием слов, обозначающих лексемы в предложении. Суперпозиция определяет порядок вычисления лексем-функций, заданный предложением – лексема вычисляется после того, как вычислены все ее аргументы. Способ же вычисления лексем задается описаниями обозначающих слов.

Пример 7.

Рассмотрим предложение

Вручая книгу, старик окинул мальчика быстрым оценивающим взглядом.

а) Суперпозиция лексем-функций этого предложения

окинул<X004.001>

(@Деепр Вручая<X000.001>

(@Вин книгу<X001.001>),

@Им старик<X003.001>,

@Вин мальчика<X005.001>,

@Тв оценивающим_взглядом<X007.001>

(@Тв быстрым<X006.001>)).

б) Массив описаний:

Вручая <X000.001>

ВРУЧАТЬ {Глаг Деепр} (см. Пример 6)

l: Z1 => <X003.001>

Z2 => <X001.001>

книгу <X001.001>

КНИГА {Сущ. Жен Неодуш@ОНА @Вин}

s: КНИГА \$14110()

старик <X003.001>

СТАРИК {Сущ. Муж Одуш@ОНЪ\$17@Им} (см. Пример 6)

окинул <X004.001>

ОКИНУТЬ {Глаг Муж} (См. Пример 5)

l: Z1 => <X003.001>

Z3 => <X005.001>

Z2 => <X007.001>

*@DeeL1 => <X000.001>

мальчика <X005.001>

МАЛЬЧИК (см. Пример 6)

быстрым <X006.001>

БЫСТРЫЙ {Прил. МжСр @ОНЪ @Тв} (см. Пример 4)

оценивающим_взглядом <X007.001>

ОЦЕНИВАТЬ_ВЗГЛЯДОМ {Глаг Прич МжСр @ОНЪ @Тв } (См. Пример 6)

l: *@\$Какой3 => <X006.001>

В суперпозиции указывается слово, именуемое лексему в предложении и ссылка на описание слова в массиве описаний. Ссылка содержит индекс лексемы, сопоставленной слову, в массиве описаний, и порядковый номер лексемы в списке лексем, сопоставленных слову в семантическом словаре. Для каждого из аргументов лексемы в заголовке указывается также вид подчинительной связи (обозначение с префиксом '@').

Описание лексемы содержит следующую информацию:

- имя лексемы в семантическом словаре;
- морфологические характеристики слова, обозначающего лексему в предложении;
- описание лексемы, взятое из семантического словаря (мы для краткости заменили его ссылками на соответствующие примеры в предыдущем тексте);
- помеченный знаком 'l:' список переменных семантической формулы, значения которых заданы в предложении.

4.2. Шаблон семантической модели. Шаблоном семантической модели предложения мы назовем последовательность утверждений о свойствах сущностей, упоминаемых в предложении и записанных на М-языке. Каждое из утверждений представляет собой результат преобразования семантической формулы лексемы, полученной из словаря

Порядок расположения утверждений в шаблоне определяется суперпозицией предложения: сначала располагаются утверждения о лексемах - аргументах, затем утверждение о лексеме-функции.

Важный шаг получения шаблона – унификация переменных в утверждениях шаблона в соответствии с подчинительными связями между лексемой функцией и ее аргументами. Необходимая для унификации информация содержится в семантико-синтаксической модели. Как правило, не все обязательные аргументы формулы лексемы явно указаны при ее использовании в предложении. Чтобы отметить это обстоятельство, мы будем знаком 'X' с номером обозначать переменные, значения которых явно указаны в предложении и участвуют в

унифицирующих подстановках, и сохранять обозначения Z- переменных для обязательных, но явно не обозначенных аргументов.

Пример 8.

Вернемся к предложению из примера 7. Шаблон семантической модели этого предложения:

КНИГА\$14110:X1;

СТАРИК\$12411:X2;

ВРУЧЕНИЕ\$15210:X3(Oper,СУБЪЕКТ.X2,ОБЪЕКТ.X1,АДРЕСАТ.Z3);

МАЛЬЧИК\$12411:X4;

ВЗГЛЯД\$124/00:X5(СУБЪЕКТ.X2,АДРЕСАТ.X4);

X5(БЫСТРЫЙ\$153/01);

\$15403ОКИДЫВАНИЕ:X6;

X3{X6(PerfOper,СУБЪЕКТ.X2, ОБЪЕКТ.X4); X2(Usor.X5)};

Шаблону соответствует комплект правил, полученный из описаний лексем:

X(ConvCaus.Z1){Z3(IncepHab.Z2)} \Leftarrow ВРУЧЕНИЕ\$15210:X(СУБЪЕКТ.Z1, ОБЪЕКТ.Z2, АДРЕСАТ.Z3)

ОЦЕНКА\$11016:X(Oper, СУБЪЕКТ.Z1,ОБЪЕКТ.Z2); Z1(Usor.Z0)) \Leftarrow

ВЗГЛЯД\$124/00:X(СУБЪЕКТ.Z1,АДРЕСАТ.Z2)}

ОСМАТРИВАНИЕ\$15403:X(СУБЪЕКТ.Z1, ОБЪЕКТ.Circ(Z2)) \Leftarrow

ОКИДЫВАНИЕ\$15403:X(СУБЪЕКТ.Z1(Usor. ВЗГЛЯД\$124/00~Z3(СУБЪЕКТ.Z1, АДРЕСАТ.Z2)), ОБЪЕКТ.Z2);

Z0(СТАРЫЙ\$12411/071.X) \Leftarrow \$12411СТАРИК:Z0;

Z0(ЮНЫЙ\$12411/071.X); X(Magn); \Leftarrow \$12411МАЛЬЧИК:Z0;

Порядок следования (и последующей обработки) утверждений шаблона задан суперпозицией предложения. Первыми указываются утверждения о значениях лексем КНИГА - значение X1 и СТАРИК - значение X2. Эти переменные, а также переменная X4 (лексема МАЛЬЧИК) являются исходными для последующей процедуры унификации. Значение атрибута АДРЕСАТ лексемы ВРУЧЕНИЕ явно не обозначено в предложении и для него сохранено обозначение Z-переменной, полученное из формулы. Унифицированные значения атрибутов переменной лексемы ОЦЕНИВАТЬ_ВЗГЛЯДОМ получены после унификации этой переменной с переменной лексемы ВЗГЛЯД в составе лексемы ОКИДЫВАНИЕ (унифицирующая переменная X5).

Последнее из утверждений шаблона оформляет значение лексемы ОКИДЫВАНИЕ как концептуализацию значения лексемы ВРУЧЕНИЕ. Это решение принято на основе следующих данных семантико-синтаксической модели предложения:

1) Обе лексемы принадлежат классу \$15 - действие. Значение одной из них (X3) находится в состоянии Oper - "выполнение действия", другой (X6) - в состоянии PerfOper - "завершенное действие".

2) Оба действия выполняются совместно (одновременно) одним и тем же субъектом, на что указывает связывающее их синтаксическое отношение @Деепр - деепричастие.

В этих условиях завершенное действие явно является составной частью выполняемого, что и отражено отношением концептуализации.

Суперпозиция предложения в совокупности с описаниями подчинительных связей лексем предложения, приведенные в семантико-синтаксической модели, полностью определяют вид и порядок унификации переменных шаблона.

Шаблон предложения, в соответствии со сказанным в предыдущих разделах работы, можно рассматривать как конъюнкцию терминальных утверждений, принимающих значение "истина" на множестве внеязыковых сущностей:

- заданных описанием предметной области анализируемых текстов;
- полученных в результате обработки ранее введенных текстов и составляющих текущее состояние базы знаний;
- полученных в результате обработки фрагмента текста, предшествующего моделируемому предложению;

- дополнительно образованных в результате обработки моделируемого предложения.

Определенное таким образом множество внеязыковых сущностей образует контекст, в терминах которого строится семантическая модель предложения.

Семантической моделью предложения мы называем шаблон предложения, в котором значения переменных замещены системными (уникальными) обозначениями внеязыковых сущностей из заданного контекста, при которых конъюнкция терминальных утверждений шаблона принимает значение "истина". Построив семантическую модель предложения, мы определим его роль, дейктическую функцию в общем контексте базы знаний.

4.3. Проблема построения модели. Эта проблема выводит нас далеко за рамки данной статьи, и, если бы нам удалось ее решить, мы фактически решили бы проблему семантического анализа связного текста. Мы далеки от ее решения и здесь на конкретном примере мы лишь наметим некоторые простые и частные подходы к ней. Основная цель примера - иллюстрация возможностей рассмотренных в работе методов описания семантики русского языка и интерпретации этих методов в терминах концептуальных моделей данных.

Пример 9.

Вернемся к примеру 8. Будем рассматривать наше предложение как часть некоего текста, предыдущий фрагмент которого уже разобран и зарегистрирован в специальном разделе базы знаний.

Рассматриваемое предложение может содержать указания о сущностях, ранее в тексте не упомянутых и впервые обозначенных именно здесь. Поэтому, несмотря на то, что шаблон предложения - конъюнкция терминальных отношений с общими переменными, мы не можем рассматривать его как один общий запрос к базе знаний и вынуждены вычислять его последовательно, утверждение за утверждением.

Вычисление начнем с попытки пополнения базы знаний комплектом правил, сопоставленных шаблону. Отсутствие какого-либо из правил в базе знаний уже свидетельствует о возможном отсутствии в ней сущности, обозначенной соответствующей лексемой.

Формируем первый запрос к базе знаний, точнее к ее разделу с фрагментом текста.

? КНИГА\$14110:X1 - не упоминалась ли ранее в тексте какая-либо книга?

Мы можем получить:

– отрицательный ответ;

– единственный положительный ответ - X1= <обозначение сущности>;

– несколько ответов - X= <список обозначений>.

При отрицательном ответе мы можем сформировать общий запрос:

? \$14110:X1 - не упоминался ли какой-либо объект из класса, которому принадлежит лексема?

Отрицательный ответ на этот запрос позволяет решить, что сущность впервые упомянута в тексте и требует нового уникального обозначения. В любом случае в результате мы имеем одно или несколько значений X1.

Следующий запрос:

? СТАРИК\$12411:X2

Обработка результатов этого запроса аналогична предыдущей. Отличие заключается в том, что лексема СТАРИК имеет толкование и при отрицательном ответе в качестве общего используется запрос, полученный из посылки правила:

? X2(СТАРЫЙ\$12411/071.X)

Анализ утверждения

ВРУЧЕНИЕ\$15210:X3(Oper,СУБЪЕКТ.X2,ОБЪЕКТ.X1,АДРЕСАТ.Z3);

значительно сложнее предыдущих. Мы имеем дело с действием (класс \$15) в состоянии выполнения. Поэтому наряду с маловероятным простым упоминанием этого действия в предыдущем фрагменте текста мы можем ожидать наличие там действий, связанных с данным. Это, в частности могут быть действия дарения, продажи, награждения, наконец, простой передачи объекта адресату. В любом из этих случаев данное действие является составной частью упомянутых и должно оформляться в составе их концептуализации. Формируя общий запрос

?X3(ConvCaus.X2){Z3(IncepHab.X1)}

мы должны быть уверены, что в процессе его выполнения будут инициализированы проверки наличия в ранее введенном тексте упоминаний о всех действиях, так или иначе связанных с переходом объекта от субъекта к адресату. Формулы семантического словаря не всегда удовлетворяют такому требованию и применение рассматриваемого в работе подхода к анализу текста связано со значительной правкой статей словаря.

Сравнительно прост анализ утверждений

ВЗГЛЯД\$124/00:X5(СУБЪЕКТ.X2,АДРЕСАТ.X4);

ОКИДЫВАНИЕ\$15403:X6;

X3 {X6(PerfOper,СУБЪЕКТ.X2, ОБЪЕКТ.X4); X2(Usor.X5)};

Значение лексемы ОКИДЫВАНИЕ определено как составная часть действия ВРУЧЕНИЕ и область определения всех его отношений ограничена этим действием. Поэтому вполне правомочно решение о том, что значения X5 и X6 - новые сущности, обозначенные текстом предложения.

5. Заключение

Анализ, результаты которого приведены в работе, показывает, что модель семантики, реализованная в семантическом словаре русского языка, может быть достаточно адекватно представлена в терминах концептуальных моделей данных. Существенным, на наш взгляд, результатом анализа является определение шаблона семантической модели предложения. Будучи эквивалентом семантико-синтаксической модели, шаблон определяет эту модель в контексте описания предметной области и множества внеязыковых сущностей, упоминаемых в тексте. Шаблон позволяет продолжить семантический анализ предложения, рассматривая его как элемент общей семантической модели связного текста.

Литература

1. Тузов В.А. Система семантико-синтаксического анализа предложений на русском языке // Информационные технологии в гуманитарных и общественных науках. СПб.:СПб ЭМИ РАН, 2002. Вып.12. С. 1–19.
2. Боярский . К.К., Каневский Е.А., Лезин Г.В., Попова А.И.. Формализация знаний в гуманитарных исследованиях // Экономико-математические исследования: математические модели и информационные технологии. СПб.:Наука, 2000. Вып.1. С. 248–264
3. Лезин Г.В., Мамедниязова Н.С. О представлении семантики концептуальных моделей в базах знаний. // В сб.: "Труды Международного семинара Диалог'2000 по компьютерной лингвистике и ее приложениям". Т.2, Протвино 2000, С. 235–242.
4. Мельчук И.А. Опыт теории лингвистических моделей "Смысл-текст" //М.:Наука. 1974.
5. Плесневич Г.С. Логика моделей "классы-бинарные отношения. I. // Изв. РАН. Теория и системы управления. 1997. №5. С. 17-26.
6. Плесневич Г.С. Логика моделей "классы-бинарные отношения. II. // Изв. РАН. Теория и системы управления. 1998. №5. С. 69-80.
7. Каневский Е.А. Атрибуты базовых и производных лексем в семантическом словаре. // Информационные технологии в гуманитарных и общественных науках. СПб.:СПб ЭМИ РАН, 2002. Вып.12. С. 29–35.