

# ОБ АВТОМАТИЧЕСКОМ ВЫЯВЛЕНИИ РЕФЕРЕНЦИАЛЬНОЙ СВЯЗНОСТИ ПОВЕСТВОВАТЕЛЬНОГО ТЕКСТА

## ON AUTOMATIC DISCLOSURE OF REFERENCIAL COHERENCY IN NARRATIVE TEXT

*Лезин Г.В. (lezin@emi.nw.ru)*

*Санкт-Петербургский экономико-математический институт РАН*

В докладе задача выявления референциальной связности рассматривается как компонента общей задачи автоматического формирования дискурса текста. Предлагается представлять текст в дискурсе в виде концептуальной модели. Обсуждаются вопросы определения денотативного статуса именных групп и повторного именованя денотативных объектов.

### *1. Введение. Постановка задачи*

**1.1.** Проблема автоматического "понимания" повествовательных текстов является одной из актуальнейших в области современных информационных технологий. Говоря о "понимании" текста информационной системой, мы имеем в виду автоматический процесс соотнесения информации, содержащейся в тексте, с информацией, имеющейся в базе знаний информационной системы. Результатом процесса является представление информации, содержащейся в тексте, на формальном языке базы знаний. Полученное таким образом представление текста часто называют **дискурсом** текста, а сам процесс получения дискурса – дискурсивным анализом текста .

Ключевым моментом дискурсивного анализа является выбор структуры дискурса. Различают два уровня структуры [1,2]: уровень, на котором представлена система связей и отношений между крупными фрагментами текста, его макроструктура, и уровень **микроструктуры**, когда целью анализа становится выявление минимальных дискурсивных единиц текста и первичных связей и отношений между этими единицами. Можно выделить три типа таких первичных связей:

- референциальные, устанавливающие отношения референции для обозначенных в тексте объектов, а также между объектами текста и имеющимися в базе знаний сущностями;
- временные, отражающие совпадение во времени или временные последовательности процессов, ситуаций, изменений состояний, упоминаемых в тексте;
- ситуативные связи, отражающие сочинительно-подчинительный характер отношений между отдельными дискурсивными единицами.

Мы различаем термины "**референтное отношение**" и "**референциальная связь**". Референтным отношением фиксируется связь между словом в тексте (именем) и обозначенной этим словом внеязыковой сущностью (референтом, денотатом). Мы считаем, что два имени референциально связаны, если этим именам в тексте сопоставлен общий денотат (имена кореферентны), либо если денотаты, сопоставленные этим именам, связаны в тексте отношениями типа "класс-подкласс", "класс-индивид", "класс или индивид-свойство, или состояние, или роль". Выявление референтных отношений и референциальных связей – основополагающая задача дискурсивного анализа текста.

В тексте референтное отношение может быть задано:

- "объектно", когда толкование (лексический смысл) имени совпадает со смыслом имени в тексте;
- местоимением или местоменным словом (анафором);
- метономическим сдвигом, когда обозначаемый "участник" ситуации задан именем какого-либо другого участника этой ситуации [3];
- метафорой.

В докладе на примерах "объектно" заданных референтных отношений рассматриваются представление этих отношений в памяти дискурса и выявление референциальных связей в повествовательных текстах. Представляемая работа выполнена в рамках исследовательского проекта, нацеленного на разработку методов автоматического дискурсивного анализа текстов общественно-политической тематики.

**1.2.** В нашем проекте целью дискурсивного анализа текста на микроуровне является формирование дискурса в виде концептуализации текста, т.е. в виде сети концептуальных отношений между маркированными в тексте денотатами. В процессе дискурсивного анализа текст обрабатывается последовательно, предложение за предложением. Дискурс очередного предложения просто встраивается в ранее построенную часть дискурса текста, не модифицируя, а лишь дополняя ее. Модификация дискурса в процессе анализа текста – самостоятельная задача, решение которой на данном этапе исследований не планировалось.

Выделяются следующие этапы анализа предложений :

1) Семантико-синтаксический разбор предложения.

На этом этапе строится дерево зависимостей между словами предложения и для каждого из слов предложения устанавливается его лексическое значение. Как правило, удается установить единственное

лексическое значение слова в предложении, но возможны случаи, когда в результате разбора мы получаем лишь набор вариантов значения. В нашем проекте используется семантико-синтаксический анализатор предложений, разработанный В.А. Тузовым [7].

2) Выделение на дереве зависимостей именных и предикативных групп. Оформление толкований выделенных групп в виде дескрипций (пропозициональных форм) [8], записанных на языке концептуальных моделей. Преобразование дерева зависимостей в общий предикат предложения, представленный списком последовательно вычисляемых дескрипций [5]. На этом этапе мы получаем внеконтекстную семантическую модель предложения.

3) Вычисление полученного списка дескрипций в контекстах ранее построенного дискурса и базы знаний о предметной области. В процессе вычисления референтные переменные в построенных на предыдущем этапе дескрипциях получают константные значения, а сами дескрипции замещением переменных вычисленными значениями приобретают форму минимальных дискурсивных единиц анализируемого текста.

4) Регистрация полученных дискурсивных единиц в дискурсе анализируемого текста.

Этот подход к определению смысла текста (в виде конъюнкции пропозиций отдельных предложений) и его (текста) денотативного значения (в виде конъюнкции константных значений пропозиций) был предложен еще в 80-х годах в рамках теории референции (прежде всего трудами Н.Д. Арутюновой и Е.В. Падучевой [8]).

Есть одно "маленькое НО": для реализации этого подхода необходимо иметь в распоряжении семантический словарь с формальными толкованиями лексических значений слов, а значит необходим семантический язык для записи таких толкований. Вот здесь-то и образовался затор. Мысль о возможной концептуализации текста как о способе и средстве представления содержащейся в нем информации, наверное, лишь в последние годы начинает осознаваться отечественными специалистами по лингвистической семантике и лексикографами. Естественно, что в ранге **получивших признание** таких словарей, причем достаточно полно охватывающих лексику языка, **в России нет**. (Прекрасная формализация сделана, судя по публикациям, в базе глаголов, разрабатываемой под руководством Е.В. Падучевой [3, 9], но это – только для глаголов и при отсутствии развитой онтологии. Очень интересные работы в этом направлении ведутся С. Ниренбургом [13], но для английского языка.). Соответственно нет и достаточно "хорошо известного научной общественности" семантико-синтаксического анализатора предложений, ориентированного на построение концептуализаций. Тем не менее, опыт определения такого рода семантического языка и опыт построения достаточно обширного семантического словаря русского языка **есть**. Это **экспериментальные** семантический словарь и семантико-синтаксический анализатор В.А. Тузова [4, 7]. Система В.А. Тузова динамично развивается и вполне пригодна для использования в вычислительных экспериментах, связанных с исследованием методов дискурсивного анализа, это – главное.

**1.3.** Семантический словарь и семантический язык, подходящие для разрабатываемого нами подхода, должны удовлетворять двум требованиям:

а) Семантический язык, используемый в словаре должен быть строго формализован.

б) Формальные описания лексики в словаре должны быть ориентированы на построение концептуализаций текста.

Чтобы содержательно пояснить эти требования, мы рассмотрим пример, в котором сопоставим два способа записи формального толкования глагола *разрезать*: разработанного Е.В. Падучевой и приведенного в ее сравнительно недавно опубликованной книге [3] и **ее же** толкование, но написанное на языке В.А. Тузова (имеющееся в текущей версии словаря толкование сделано несколько иначе). Отметим заранее, что оба этих способа удовлетворяют нашим требованиям.

У Е.В. Падучевой:

**разрезать** 1.1 (как в *разрезать арбуз*)

*Легенда* — исходная лексема парадигмы.

*Актуальная структура* —

Имя	Синтаксис	Ранг	Роль	Таксономия
X Y	<i>Сб</i>	<i>Центр</i>	Агенса	ЛИЦО
Z W	<i>Об</i>	<i>Центр</i>	Пациенса	ПРЕДМЕТ
	сущ. твор. на +	<i>Периф</i>	Инстр	ПРЕДМЕТ: имеет острый край
	сущ. вин.	<i>Периф</i>	Результат	ЧАСТИ ПРЕДМЕТА

*T-категория* — действие: обычное

*Схематическое толкование* —

*экспозиция* — до *t*, предшествующего МН (моменту наблюдения) У был целый

1) —

2) —

3) —

4) *деятельность* — в *t* до МН X действовал с целью

5) *способ* — воздействовал на У: с помощью Z-а

б) *каузация* — это вызывало

7) процесс в Объекте — шел процесс в У-е: синхронный деятельности: имеет предел: У утрачивал целостность

8) результат (совпадающий с целью; он же — предел процесса) — наступило и в МН имеет место состояние: существует W — отдельные части У-а

9) следствие — тем самым У не существует как единый предмет

10) -

Тема — деформация

Аспектуальная характеристика — СВ\ парный НСВ — *разрезать* 1.1, действие

Компоненты толкования заданы в виде значений признаков. Предполагается, что для каждого из признаков имеется оговоренный набор возможных значений – элементарных составляющих смысла [3].

Это же толкование, переписанное на языке В.А. Тузова:

Синтактика:

(X:ЧЕЛОВЕК\$1241~!Им, Y:ВЕЩЬ\$121~!Вин, Z: ИНСТРУМЕНТ\$121313~!Тв, W:!наВин)

Семантика:

а) РАЗРЕЗАТЬ\$152500 (АГЕНС:Х, ПАЦИЕНС:У, ИНСТР:Z, РЕЗУЛЬТАТ:W)

б) PerfCaus(X>FinOper(X, РЕЗАНИЕ\$152500(АГЕНС:Х, ПАЦИЕНС:У, ИНСТР:Z)),

And(FinFunc(Y), IncepFunc(W:MultFlanc(Y)))

Строка а) семантики представляет собой описание лексемы-класса с именем РАЗРЕЗАТЬ\$152500. Этому классу принадлежат все ситуации, обозначаемые в разных текстах глаголом *разрезать* в заданном лексическом значении. Часть б) содержит толкование лексемы-класса, заданной в а). Отметим, что в толковании на языке В.А. Тузова понятие "момент наблюдения" скрыто в интерпретации семантических примитивов (унарных и бинарных базисных функций по В.А. Тузову):

– PerfCaus – признак действия, нацеленного на результат, с каузатором, заданным в первом операнде функции;

– FinOper – действие, заданное вторым операндом закончилось в некоем интервале времени Т, явно не указанном;

– FinFunc – в некоем интервале времени Т закончилось существование операнда;

– IncepFunc – в некоем интервале времени Т начал существовать операнд.

Запись W:MultFlanc(Y) толкуется как "W представляет собой множество частей (кусков) целого Y, полученного в результате механического воздействия (разбивания, разбрызгивания) на Y".

Текстовая форма толкования:

"X, закончив *резание* Y-а инструментом Z, скаузировал состояние, в котором Y как целое перестал существовать, и начали существовать W в качестве частей Y."

Толкование задано в терминах классов классификатора (онтологии) лексем русского языка, являющегося неотъемлемой частью семантического словаря. Записью РЕЗАНИЕ\$152500 обозначено множество денотатов лексического значения слова *резание* в классе \$152500.

Толкования написаны на внешне совершенно разных семантических языках. Общее внешнее сходство этих языков выражается лишь в том, что оба они строго формальны.

Однако, сопоставляя, нельзя не заметить их содержательной общности: толкованиями на обоих языках устанавливается практически идентичный (с точностью до степени разработанности онтологии русского языка) набор **концептуальных** отношений между толкуемыми объектами (классами и их конкретными представителями). Поэтому все далее сказанное может быть, вообще-то, в равной степени отнесено к обоим формальным представлениям.

Мы исходим из следующей концептуальной интерпретации семантики лексемы [5]:

Для любой **конкретной ситуации** V1, упомянутой в тексте дискурса, справедливо:

а) если V1 обозначена глаголом "разрезать" в заданном лексическом значении, то дискурс текста содержит также некоторую конкретную ситуацию V2 из класса РЕЗАНИЕ\$152500 и набор отношений над конкретными значениями V2 и актантами глагола "разрезать";

б) обратно, если дискурс содержит конкретную ситуацию V2 и заданный толкованием набор отношений, то в дискурсе имеется (может быть неявно обозначенная) ситуация V1 из класса РАЗРЕЗАТЬ\$152500 с соответствующими значениями актантов.

Ясно, что, получив толкование, мы имеем дело с утверждением в пропозициональной форме, правда, завуалированной, т.к. в ней опущены некоторые параметры и элементы полной записи, восстанавливаемые по строго определенным правилам. Задача выявления опущенных параметров (неявно заданных переменных, значков квантификации, формальная интерпретация семантических примитивов) – чисто техническая. Ее решение определяется тем формальным языком, который мы выбираем для представления информации в базе знаний и дискурсе. Имея такой транслятор и применив его к словарю, мы получим новый словарь, толкования лексем в котором написаны уже на языке дискурса и базы знаний. Транслятор на выбранный таким образом язык не сложен. Общие правила трансляции нами разработаны и опубликованы на Диалоге [5, 6]. Реализовано несколько версий транслятора на язык используемой нами системы

управления базой знаний. По результатам этой работы проводились уточнения исходной версии семантического языка, и менялись требования к языку представления знаний.

Результат трансляции для рассмотренного нами примера имеет вид конъюнкции атомарных (терминальных) отношений на языке дискурса и базы знаний, по своей форме представления уже подготовленных для регистрации в памяти дискурса.<sup>1</sup>

а) лексема-класс:

\$152500 : РАЗРЕЗАТЬ; РАЗРЕЗАТЬ : V1; V1(АГЕНС:Х); V1(ПАЦИЕНС:У); V1(ИНСТР:Z);  
V1(РЕЗУЛЬТАТ:W);

б) толкование:

\$152500 : РЕЗАНИЕ; РЕЗАНИЕ : V2; V2(FinOper. T1); V2(АГЕНС:Х); V2(ПАЦИЕНС:У); V2(ИНСТР:Z);  
Y(FinFunc.T2); W(MultFlanc.Y); W(IncepFunc.T2); T1(Antaut.T2);

В приведенной записи использованы всего два вида отношений: отношение конкретизации <класс> : <подкласс или экземпляр класса> и отношение <класс или экземпляр>(<свойство определяемого объекта>.<значение свойства>). Эти отношения являются основополагающими для всех языков определения онтологий. T1 и T2 – переменные, определенные на множестве денотативных значений интервалов времени. Отношением T1(Antaut.T2) задано, что интервал T1 предшествует T2.

Есть один очень **важный момент**. Части а) и б) **этого вида** формальных описаний лексем в системе дискурсивного анализа работают по-разному.

Часть а) представляет собой **экстенционал**, шаблон, который после замены переменных константными денотативными значениями превращается в константное выражение, регистрируемое в памяти дискурса. Каждое использование в тексте глагола *разрезать* с денотативным статусом "новая информация" порождает в дискурсе текста **собственный** комплект константных терминальных отношений относительно каждого **конкретного** значения V1.

Часть б) – **интенционал**, правило, смысл денотативных значений V1 как для данного текста, так и для всех прочих, подлежащих дискурсивному анализу. Примером использования правила из б) может быть поиск в дискурсе текста или в базе знаний неявно обозначенных ситуаций, тождественных ситуациям, обозначаемым глаголом *разрезать*. Правило не нужно размножать в дискурсе. Достаточно иметь его в единственном экземпляре и применять по мере необходимости.

Следствием сказанного является тот факт, что для построения пропозициональной формы предложений текста исходными "кирпичиками" служат лишь части а) семантики лексем, полученные из словаря.

Возможны и **другие виды** описаний. Подробнее об этом см. в [5].

**1.4.** Итак, знания о языке представлены в нашей системе семантическим словарем В.А. Тузова, в котором каждому из слов сопоставлены его морфологические характеристики и описания его возможных лексических значений [4]. Прагматические знания о предметной области представляются в базе знаний. Результаты дискурсивного анализа текста накапливаются в специальной памяти – памяти дискурса. Информация в базе знаний и в памяти дискурса представлена одинаково, в виде концептуальных моделей. Мы можем ввести сквозную нумерацию классов лексем, лексем и конкретных объектов в дискурсах текстов, словаре и базе знаний. Полученный таким образом номер представляет денотат сущности в информационной системе. Дискурс текста представляет собой сеть метаотношений между денотатами, упомянутыми в тексте и имеющимися в системе.

Далее показан пример результата выполнения этапов 1 и 2 дискурсивного анализа предложения<sup>2</sup>

(1) *Киев перешел в контраступление после безуспешных переговоров глав правительств России и Украины по поводу цен на газ.*

Внеконтекстная семантическая модель этого предложения имеет вид последовательности дескрипций:

КИЕВ [!X<sub>0</sub>]

\$12314:ГОРОД; ГОРОД:!X<sub>0</sub>; Киев~!X<sub>1</sub>; !X<sub>0</sub>(НАЗВАНИЕ.!X<sub>1</sub>);

РОССИЯ [!X<sub>2</sub>]

\$1231:СТРАНА; СТРАНА:!X<sub>2</sub>; Россия~!X<sub>3</sub>; !X<sub>2</sub>(НАЗВАНИЕ.!X<sub>3</sub>);

УКРАИНА [!X<sub>4</sub>]

\$1231:СТРАНА; СТРАНА:!X<sub>4</sub>; Украина~!X<sub>5</sub>; !X<sub>4</sub>(НАЗВАНИЕ.!X<sub>5</sub>);

ПРАВИТЕЛЬСТВО [!X<sub>6</sub>]

\$123407:ПРАВИТЕЛЬСТВО; !X<sub>2</sub>(ПРАВИТЕЛЬСТВО.!X<sub>6</sub>)

ПРАВИТЕЛЬСТВО [!X<sub>7</sub>]

\$123407:ПРАВИТЕЛЬСТВО; !X<sub>4</sub>(ПРАВИТЕЛЬСТВО.!X<sub>7</sub>);

<sup>1</sup> Чтобы получить подобное представление для толкования Е.В. Падучевой, необходимо предварительно разработать интерпретацию значений признаков 1) - 10) в виде системы атомарных отношений базы знаний.

<sup>2</sup> Этот и большинство следующих примеров подбирались в Национальном корпусе русского языка [//ruscorp.org.ru].

ГЛАВА [!X<sub>8</sub>]  
 \$12413205:ГЛАВА; !X<sub>6</sub>(ГЛАВА !X<sub>8</sub>);  
 ГЛАВА [!X<sub>9</sub>]  
 \$12413205:ГЛАВА; !X<sub>7</sub>(ГЛАВА !X<sub>9</sub>);  
 ГЛАВА [!X<sub>10</sub>]  
 ГЛАВА:!X<sub>10</sub>; !X<sub>10</sub>(Mult); !X<sub>10</sub>!X<sub>8</sub>; !X<sub>10</sub>!X<sub>9</sub>;  
 ГАЗ [!X<sub>11</sub>]  
 \$12115:ГАЗ; ГАЗ:!X<sub>11</sub>;  
 ЦЕНА [!X<sub>12</sub>]  
 ЦЕНА:!X<sub>12</sub>; !X<sub>12</sub>(ДАТ.!X<sub>11</sub>);  
 ПЕРЕГОВОРЫ [!X<sub>13</sub>]  
 ПЕРЕГОВОРЫ:!X<sub>13</sub>; !X<sub>13</sub>(МЕЖДУ.!X<sub>10</sub>); !X<sub>13</sub>(ТЕМА.!X<sub>12</sub>); !X<sub>13</sub>(НЕ\_УСПЕХ);  
 КОНТРНАСТУПЛЕНИЕ [!X<sub>14</sub>]  
 КОНТРНАСТУПЛЕНИЕ:!X<sub>14</sub>; !X<sub>13</sub>(Antaut.!X<sub>14</sub>); !X<sub>14</sub>(Perf\_Incep\_Oper); !X<sub>14</sub>(АГЕНС.!X<sub>0</sub>);

Обозначения:

Прописными буквами обозначены имена лексем, строчные буквы используются в автонимных именах. Значком '!X' с индексом обозначены референтные переменные, определенные на классах денотативных значений. Записью 'СТРАНА:!X<sub>2</sub>' обозначено отношение принадлежности значений !X<sub>2</sub> классу, заданному лексемой СТРАНА, записью 'Киев~!X<sub>1</sub>' указано, что все значения переменной !X<sub>1</sub> имеют автонимное имя 'Киев', а записью '!X<sub>2</sub>(НАЗВАНИЕ.!X<sub>3</sub>)' обозначено наличие у всех значений !X<sub>2</sub> свойства НАЗВАНИЕ, значение которого определено переменной !X<sub>3</sub>. Mult (признак замкнутого множества), Antaut (предшествование во времени), Perf\_Incep\_Oper (процесс начался) – семантические примитивы [4]. В заголовке дескрипции указано имя лексемы, выбранной для опорного слова именной группы. Дескрипцией задано условие, которому должно удовлетворять денотативное значение главной (операторной) переменной дескрипции, указанной в заголовке в квадратных скобках. Из примера видно, что в списке дескрипций отсутствуют в явном виде толкования наречия *после* и полузнаменательного глагола *перейти*. Лексические значения этих слов были проинтерпретированы на втором этапе анализа, и в результате толкование в дескрипции пополнилось указаниями, что КОНТРНАСТУПЛЕНИЕ трактуется в тексте как начавшийся процесс (см. Perf\_Incep\_Oper), и началу этого процесса предшествовали (см. Antaut) ПЕРЕГОВОРЫ. Интерпретация примитивов Perf\_Incep\_Oper и Antaut должна быть выполнена на следующем этапе, при выявлении временной связности текста.

В нашем примере среди возможных значений переменных !X<sub>0</sub> – !X<sub>14</sub> могут быть как сущности, уже обозначенные в дискурсе текста, так и новые, привносимые данным предложением. Так возможными значениями переменных !X<sub>2</sub> – !X<sub>9</sub> (обозначенными как *Россия*, *Украина*, *главы правительств*) являются денотаты, скорее всего описанные в базе знаний, а значением переменной !X<sub>14</sub> (*контрнаступление*) является новая информация и для ее регистрации в дискурсе необходимо сформировать новый денотат. Заменив переменные !X<sub>0</sub> – !X<sub>14</sub> выявленными или вновь сформированными константными значениями, мы получим **концептуализацию** предложения референциально связанную с дискурсом. Эти действия выполняются на третьем этапе дискурсивного анализа.

Реализация третьего этапа связана с необходимостью ответа по крайней мере на следующие вопросы:

- Каковы виды денотативных значений в памяти дискурса, их перечень и свойства?
- Какие механизмы для маркирования разных видов значений имеются в языке?
- Как работают эти механизмы в связанном тексте?

В лингвистике, в ее разделах, посвященных семантике языка, для ответов на эти вопросы многое сделано и делается (см., например, [3, 8–11]). Цель нашего проекта – попытаться объединить полученные результаты, формализовать их, сделать доступными для реализации в системах автоматического анализа текстов. В докладе рассматривается возможный подход к решению этой задачи на примере анализа объектных референтных отношений.

## 2. Виды денотативных значений и референциальный статус имен в тексте

**2.1.** Референтное отношение в сети дискурса представлено парой <имя>~<денотат>. В отношении указывается либо имя лексемы, полученной из словаря для слова, которым обозначен в тексте заданный денотат, либо автонимное имя [11], не сопровождаемое толкованием. Имени сопоставляется место вхождения соответствующего слова в текст.

Каждый из денотатов сети характеризуется признаком, определяющим, какого вида сущность им представлена:

- Конкретный объект внеязыковой действительности (индивид), обозначаемый далее как 'e<sub>i</sub>', где 'e' – признак денотата, а i – число, представляющее его в системе. Например, в референтном отношении ЧЕЛОВЕК~e<sub>i</sub>, 'ЧЕЛОВЕК' – имя лексемы, полученной из словаря для слова *человек*, обозначающего в тексте конкретного человека.

- б) Замкнутое множество индивидов ( $s_i$ ). Количество элементов во множестве может быть задано в тексте, а может быть неопределенным. (ТОЛПА~ $s_i$ )
- в) Класс объектов, общий для всех текстов и заданный семантическим словарем языка. Этот денотат – компонента языка. Его имя уникально, поэтому записанное прописными буквами имя лексемы можно считать и обозначением денотата.
- г) Класс объектов, действующий в дискурсе текста ( $d_i$ ). Мы различаем "понятия" в языке, заданные лексемами семантического словаря, инвариантные по отношению к разным текстам, и понятия в дискурсе текста. Различие выражается прежде всего в том, что обозначенное одним и тем же именем понятие в языке и дискурсе имеют разные денотативные значения. Более того, они могут иметь разные денотативные значения и в разных дискурсах. Тем самым обеспечивается возможность текстовых трансформаций и уточнений общего для всех текстов понятия, заданного описанием языка. В примере

(2) *Женщине в нашем обществе по-прежнему закрыт доступ на вершины власти.* (пример Ю.Д. Апрессяна, НОС, вып.3, С. XXIV)

референтное отношение ЖЕНЩИНА~ $d_i$  сопоставляет лексеме ЖЕНЩИНА, полученной из словаря, денотат  $d_i$  со смыслом *женщина в нашем обществе*.

д) Денотат-переменная, определенная на некотором множестве конкретно-референтных значений ( $x_i$ ) [7].

(3) *Каждый отец думает, что его ребенок самый умный.*

В (3) лексеме ОТЕЦ сопоставляется дискурсивный класс  $d_1$  и на этом классе определяется переменная  $x_2$ , способная принимать любое конкретно-референтное значение, относящееся к классу  $d_1$ .

Словосочетанием *каждый отец*, следовательно, определяются отношения ОТЕЦ~ $d_1$ ;  $d_1$ : $x_2$ ; где значком ':' обозначена принадлежность  $x_2$  классу  $d_1$ . Отметим, что в концептуализации предложения присутствует и вторая переменная, представляющая ребенка и референциально зависимая от  $x_2$ . Концептуальное отношение  $x_2$ (РЕБЕНОК. $x_3$ ) отражает факт существования ребенка  $x_3$  у отца  $x_2$ , и именно  $x_3$  характеризуется как *самый умный*.

**2.2.** Формированию референтного отношения предшествует процедура выявления референциального статуса слов текста – набора признаков, характеризующих способ установления заданного словом референциального отношения [8]. (Не все слова обладают свойством референции. Например, предлог, связывая в предложении пару имен (*последний из мозикан, лежит на столе*), не имеет собственного референта в дискурсе; его функции в тексте ограничиваются определением семантико-синтаксических отношений между словами. Предлог не имеет референциального статуса).

Зная морфологические характеристики слова, его лексическое значение и определив тип предложения и место слова в предложении при его актуальном членении, мы можем попытаться определить:

- зависимость/автономность выбора денотата;
- вид денотата, определяемого этим словом;
- наличие/отсутствие в дискурсе кореферента для данного слова (признак известное/новое).

Далее мы рассмотрим примеры правил определения перечисленных признаков. Приведенные нами наборы заведомо не полны и предназначены скорее для иллюстрации общего подхода.

**2.3.** Начнем с определения признака "зависимость/автономность" выбора денотата.

Каждое из слов текста, имеющее референциальный статус, является компонентой именной (ИГ) или предикативной (ПГ) группы и может быть либо опорным словом группы, либо аргументом опорного слова. Например словосочетание *красная куртка Маши* образует на уровне синтаксического подчинения ИГ *куртка(Маши, красная)*. Внеконтекстное толкование этой ИГ: "Любая ЖЕНЩИНА !X с именем *Маша* имеет некоторую (для каждой конкретной женщины свою) *куртку* !Y с некоторым (для каждой куртки своим) оттенком *красного* цвета !Z". Толкование распространяется на все значения референтной переменной !X, сопоставленной имени *Маша* в разных контекстах. В конкретном контексте имени *Маша* будет сопоставлен конкретный денотат  $e_i$ , и именно этим денотатом будут фактически определяться денотативные значения переменных !Y и !Z. Имена *красная* и *куртка* находятся в **референциальной зависимости** от имени *Маша*, которое уже не зависит ни от какого другого имени [3, С. 69] и имеет **автономный** статус. В логике зависимость между переменной X, связанной квантором всеобщности, и находящимися в сфере действия этого квантора переменными Y и Z, связанными квантором существования, называют сколемовой зависимостью. В большинстве ИГ [12], представляющих генитивные отношения аргумент опорного имени имеет зависимый статус, а имени аргумента – автономный.

**2.4.1.** Конкретное референциальное значение может маркироваться:

а) Именем собственным (*Газпром, Украина, Виктор Ющенко*).

б) Опорным существительным ИГ, когда в составе его аргументов присутствуют имена с конкретным денотативным значением и автономным референциальным статусом (*резиденция Газпрома*).

в) Опорным существительным ИГ, когда в составе его аргументов присутствуют неопределенные местоимения *какой-то, некий, чей-то (какой-то зверь, чье-то пальто)*.

г) Неопределенными местоимениями *кто-то, что-то, кто-нибудь, что-нибудь, кто-либо, что-либо, кое-кто, кое-что, некто, нечто (В дверь кто-то постучал)*.

д) Опорным существительным ИГ в условиях его использования в качестве актанта глагола-сказуемого в актуально-длительном (конкретно-процессном) значении [9].

(4) *Поезд мчался в разноцветном тоннеле осенних лесов.* ["Наш современник", 2003]

е) Опорным существительным ИГ в условиях его использования в качестве актанта глагола-сказуемого совершенного вида, если глагол относится к одной из вендлеровских категорий: предельные процессы (accomplishment) или скачки (achievement). (*Поезд пришел вовремя. Дом рухнул.*)

**2.4.2.** Конкретные множества конкретных единичных объектов могут задаваться:

а) Непосредственно лексическим значением слова (*толпа, автоколонна, группа*);

б) Множественным числом - опорным именем ИГ, удовлетворяющим условиям п. **2.2.2.**

(5) *Управление делами вскрыло конверты.* ["Известия", 2001.08.10]

в) Конкретным перечислением или указанием (не обязательно конкретного) количества.

В примере (1) генитивный ряд *глав правительств России и Украины* в процессе построения предиката предложения преобразуется в последовательность конкретно референтных дескрипций, из которых лишь одна !X<sub>10</sub> определяет в качестве возможного значения замкнутое множество денотатов.

Еще примеры:

(6) *Несколько человек, работавшие в цеху, вскрыли дверь, когда подсобка уже горела открытым огнем.* ["Коммерсантъ-Daily", 1996.01.25]

(7) *В состав группы "прямого действия" вошли исполнитель акции, трое партийцев из группы поддержки и фотограф.* ["Завтра", 2003.03.16]

**2.4.3.** Абстрактные референциальные значения маркируются:

а) Существительными, лексически ориентированными на именование абстрактных значений (*математика, общество*) при отсутствии конкретизирующих признаков (*общество моих друзей*)

б) существительным – именной частью именного составного сказуемого с глаголами-связками с отвлеченным или ползунаменательным значениями (*Настоящий охотник – всегда оптимист. Арестанты оказывались обыкновенными изможденными людьми.*)

в) Инфинитивом в условиях использования оценочных наречий в качестве предиката, выражающего ту или иную оценку действия, названного инфинитивом (*Детям вредно пить вино.*) [11]. Референты актантов такого инфинитива также абстрактны, если не удастся установить конкретность их референции (*Ивану вредно пить вино.*)

г) Субъектом X в предикации вида X – P, где P – предикат описывающий постоянное, вневременное свойство P, если для X не удалось установить признака конкретной референции. (*Дети любознательны.* Имя *Дети* маркирует класс. Но в: *Эти дети любознательны* вид денотата для имени *дети* определен в предыдущей части текста) [11].

**2.4.4.** Переменные в тексте маркируются:

а) В составе именной группы, если опорное слово этой группы имеет референтом абстрактное или конкретное множество денотативных значений, и в составе именной группы присутствуют местоимения *любой, каждый, всякий*.

(8) *... в обозримом будущем каждый участковый получит автомобиль в служебное пользование, чтобы своевременно выезжать в жилой сектор.* ["Известия", 2001.12.04]

В этом примере концептуализация *каждый участковый* представляется отношениями УЧАСТКОВЫЙ~d<sub>1</sub>; d<sub>1</sub>:x<sub>2</sub>. Денотат-переменной x<sub>2</sub> представлены значения актантных ролей АДРЕСАТ в ПОЛУЧИТЬ и АГЕНС в ВЫЕЗЖАТЬ.

Отметим, что и в примере (2) и в (8) денотат-переменные играют роль **параметров**, связывающих отношения в концептуализациях, и именно этой функцией обусловлена необходимость их использования.

б) Местоимениями в косвенно-вопросительных предложениях в условиях, когда субъекту главного предложения неизвестен ответ на вопрос.

(9) *Они долго решали, куда поехать летом* (пример из [11]).

В этом примере *Они, долго* и *решали* имеют конкретно-референтные значения, определяемые предшествующим контекстом. Предположим, что в конкретном дискурсе этими значениями являются: e<sub>1</sub> для *Они*, e<sub>2</sub> для *решали* и e<sub>3</sub> для *долго*. В рассматриваемом предложении место поездки – неизвестная величина, представляемая переменной-денотатом (обозначим ее как x<sub>1</sub>). Но тогда столь же неизвестной является и контекстная конкретизация действия ПОЕХАТЬ. Референтом этой лексемы в предложении также является переменная (обозначим ее как x<sub>2</sub>). Концептуализация предложения (9):

РЕШАТЬ~e<sub>2</sub>; ДОЛГО~e<sub>3</sub>; e<sub>2</sub>(АГЕНС.e<sub>1</sub>, ЗАДАЧА.x<sub>1</sub>, ПРОДОЛЖИТЕЛЬНОСТЬ.e<sub>3</sub>); ПОЕХАТЬ:x<sub>2</sub>;  
x<sub>2</sub>(АГЕНС.e<sub>1</sub>, МЕСТО.x<sub>1</sub>)

2.5. Признаки "зависимость/автономность выбора денотата" и "вид денотата" (см. п. 2.2) определяются на этапе построения предиката предложения и сопоставляются **каждой** из его референтных переменных. Как можно видеть из примеров, смысл референциального статуса слова в тексте не исчерпывается наличием этих признаков. В ряде случаев в процессе определения денотативного статуса слова выявляется необходимость модификации лексического толкования слова, ввода в предикат предложения новых дескрипций с отсутствующими в лексических толкованиях слов отношениями. В примере (1) это выразилось в построении дополнительных дескрипций для словосочетания *глав правительств России и Украины*. Аналогичные преобразования необходимы и при построении предикатов предложений в примерах (7)–(9).

### 3. Референциальная связность текста

3.1. Два имени в тексте кореферентны (референциально тождественны), если обозначают один и тот же объект, и, соответственно, в дискурсе кореферентность имен представляется референтным отношением <список имен>~<денотат>. Обычно, автор текста использует новое имя для обозначения ранее упомянутого денотата лишь в одном из двух случаев: либо, если автор уверен, что обозначаемый объект активирован в сознании читателя, и, следовательно, последний легко увяжет новое обозначение с ранее использованным антецедентом; либо эти два обозначения имеют "привычную" для читателя тождественность, сложившуюся вне текста и ставшую нормой или общего языка, или языка профессионального, свойственного предметной области [2]. В случае автоматического анализа текста эта особенность создания и восприятия текста моделируется использованием специальной кратковременной памяти, в которой задерживаются результаты анализа нескольких последних предложений текста. Поиск антецедента для анализируемого имени начинается с поиска в кратковременной памяти и, лишь при неудаче, продолжается в памяти дискурса.

3.2. При обозначении конкретных денотатов повторное обозначение может осуществляться:

а) Именем, денотативные значения которого принадлежат классу, в котором находится класс антецедента.

(10) *Да что же могло привлечь **соболя** на этом пустыре? Здесь не только зайцу, рябчику, но даже захудалому мышонку затаиться негде. И все-таки, к моему удивлению, пронырливый **зверек**, как говорится, приступил тут же к генеральной **ревизии**. Мохнатого **ревизора** интересовали буквально каждый пенек и коряга.* [Юный натуралист, 1975] – (СОБОЛЬ, ЗВЕРЬ, РЕВИЗОР) ~ e<sub>1</sub>.

(11) *Почему на почте не **приняли** паспорт к отправлению ценной бандеролью, сказать трудно. Запретов на **это действие** нет.* ["Известия", 2002.03.19] – (ПРИНЯТЬ, ДЕЙСТВИЕ) ~ e<sub>1</sub>.

(12) *Вот у входа в первую среднюю школу собралась большая **группа учащихся**. **Мальчики и девочки** терпеливо ожидают у двери, у которой стоит техническая служащая.* ["Северный колхозник", 1957.10.01] – УЧАЩИЙСЯ~s<sub>1</sub>; ГРУППА~s<sub>2</sub>; МАЛЬЧИКИ~s<sub>3</sub>; ДЕВОЧКИ~s<sub>4</sub>; s<sub>2</sub>:s<sub>3</sub>; s<sub>2</sub>:s<sub>4</sub>;

В примере (12) именем *учащиеся* обозначено множество s<sub>1</sub> конкретных значений. Генитивным отношением *группа учащихся* из s<sub>1</sub> вычленяется подмножество s<sub>2</sub>. Именами *мальчики*, *девочки* в тексте обозначены замкнутые множества s<sub>3</sub>, s<sub>4</sub>. Лексемы УЧАЩИЙСЯ, МАЛЬЧИК, ДЕВОЧКА принадлежат общему семантическому классу, поэтому множество, образованное в результате концептуальной интерпретации сочинительной связи в выражении *мальчики и девочки*, референциально тождественно s<sub>2</sub>. Отметим, что референциальная связность в концептуализации этого фрагмента текста выражена не отношением референтности, а принадлежностью s<sub>3</sub> и s<sub>4</sub> общему множеству s<sub>2</sub>.

б) Именем, обозначающим значения актантной роли в текущем исполняемом действии.

В примере (11) лексема РЕВИЗОР – имеющееся в языке имя агентивного актанта действия РЕВИЗИЯ, а в примере (13) ОРАТОР – имя выступающего (с речью).

(13) *Пыл лидера "Яблока" остудил и **выступавший** следом **г-н Гозман**. Несмотря на то, что **СПС** не участвует в конгрессе напрямую, заявил он, **правые** считают всех собравшихся в зале союзниками. Кроме того, **СПС** возрождается, утверждал **оратор**, перечисляя последние электоральные успехи **партии**.* ["Эксперт", 2004] – (Гозман, ОРАТОР) ~ e<sub>1</sub>; (спс, ПАРТИЯ, ПРАВЫЙ) ~ e<sub>2</sub>.

в) Указанием замкнутого множества, объединяющего ранее маркированные объекты. Как правило, в этих случаях референция к объединяющему множеству отмечается использованием дейктических указателей (например, *все эти ...*)

(14) *В зале порхали **девушки** в стразах и в ботфортах с перьями диковинных птиц, метались **операторы** и **фотографы**, устало позировали известные **персонажи**. Все эти люди пришли посмотреть на ювелирную коллекцию ...* ["Эксперт: Вещь", 2003] – ДЕВУШКА~s<sub>1</sub>; ОПЕРАТОР~s<sub>2</sub>; ФОТОГРАФ~s<sub>3</sub>; ПЕРСОНАЖ~s<sub>4</sub>; ЧЕЛОВЕК:s<sub>5</sub>; s<sub>5</sub>:s<sub>1</sub>; s<sub>5</sub>:s<sub>2</sub>; s<sub>5</sub>:s<sub>3</sub>; s<sub>5</sub>:s<sub>4</sub>.

Здесь, как и в (12), референциальная связность выражена отношениями принадлежности общему множеству.

в) Указанием свойства, состояния или действия (завершенного или продолжающегося), если для ранее упомянутого объекта эти его атрибуты были либо ранее указаны в тексте, либо заданы в базе знаний о предметной области.

(15) Он был выведен из задумчивости приходом одного **приятеля** — молодого человека. — Здравствуй, Иванов, — сказал **пришедший** ... [М.В. Авдеев. Тамарин (1851)] – ПРИХОД ~ e<sub>1</sub>; ПРИЯТЕЛЬ ~ e<sub>2</sub>; ЧЕЛОВЕК:e<sub>2</sub>; e<sub>2</sub>(ВОЗРАСТ.МОЛОДОЙ); e<sub>1</sub>(АГЕНС.e<sub>2</sub>); СКАЗАТЬ~e<sub>3</sub>; e<sub>3</sub>(АГЕНС.e<sub>2</sub>).

В этом примере референциальная связность выражена в концептуализации текста референциальной тождественностью агентов в действиях ПРИХОД и СКАЗАТЬ.

Другим примером рассматриваемого вида референции может быть референция прилагательным *правые* в (13). Отметим, что и именем *пришедший* в (15), и именем *правые* в (13) обозначена заведомо известная информация. Но есть и различия: имя *пришедший* ситуативно, обозначает состояние обозначаемого денотата, и информацию об изменении состояния следует искать в уже построенной концептуализации текста, а имя *правые* маркирует денотат его постоянным свойством. Наличие этого свойства может быть задано вне текста, в базе знаний.

г) Список кореферентных имен в референтном отношении может пополняться в результате интерпретации глаголов называния. Эти глаголы приходится интерпретировать непосредственно в алгоритме выявления референциальных связей.

(16) "**Профессор**", как назвал себя **продавец**, обещал своим **снадобьем** спасти любого ... [Аргументы и факты, 2001.02.14] – (профессор, ПРОДАВЕЦ) ~ e<sub>1</sub>;

Использование автонимных имен в текстах, а также толкования глаголов называния подробно рассмотрены в [11].

**3.3.** Мы ограничились рассмотрением лишь нескольких достаточно часто используемых в текстах приемов повторного обращения к ранее упомянутой информации. Мы видим, что выявление референциальной связи между фрагментами текста не сводится просто к пополнению списка кореферентных имен в референтных отношениях. В процессе выявления связности концептуализация текста может пополняться новыми, не заданными в толкованиях отношениями (см. примеры (12) и (14)). Существенным обстоятельством является необходимость интерпретации толкований глаголов называния непосредственно в процессе выявления связности. Интерпретация сводится к пополнению списка имен в референтном отношении, открывая новую возможность обращения к объекту в последующем тексте.

### **Заключение**

Материалом этого доклада мы попытались со своей "прикладной" стороны еще раз обратить внимание (прежде всего лексикографов) на эффективность концептуальных моделей в задаче реализации дискурсивного анализа текста. Как пишет Е.В. Падучева [3, с. 155], "До последнего времени в лингвистике всерьез принималось только соотношение между формой и смыслом. Отношение высказывания к действительности (составляющее содержание лингвистической теории референции) изучалось в известной степени независимо от лексической семантики или семантики синтаксиса."

Сложилась ситуация, когда при общей ориентации на "отражение «наивной», или языковой картины мира" лексикограф, разрабатывая формализованное толкование слова (и даже используя при этом неявно заданный семантический метаязык), практически не связывает эту разработку с возможностью ее дальнейшего использования в системах, осуществляющих автоматические манипуляции с толкованиями. В результате мы, даже имея формализованное толкование, не получаем возможности автоматического конструирования семантики предложения на основе толкований составляющих его слов.

### **Список литературы**

1. Кибрик А., Паршин П. Дискурс // [www.krugosvet.ru/articles/82/1008254/1008254a1.htm](http://www.krugosvet.ru/articles/82/1008254/1008254a1.htm)
2. Кибрик А.А., Плуныя В.А.. Функционализм // *Фундаментальные направления современной американской лингвистики*. М.: 1997. С.276-339.
3. Падучева Е.В. Динамические модели в семантике лексики. // М.: "Языки славянской культуры", 2004.
4. Тузов В.А. Компьютерная семантика русского языка. // СПб.:Изд-во С.-Петербур. ун-та, 2004.
5. Лезин Г.В., Тузов В.А. О представлении результатов семантико-синтаксического анализа текстов концептуальными моделями данных // *Труды Международного семинара Диалог'2003 по компьютерной лингвистике и ее приложениям*. Москва: Наука, 2003. С. 383-393.
6. Лезин Г.В. Герасимов С.М. Каневский Е.А., Опыт построения предикатных форм предложений. // [www.dialog-21.ru/dialog2006/materials/html/Lezin.htm](http://www.dialog-21.ru/dialog2006/materials/html/Lezin.htm)
7. Тузов В.А. Компьютерная грамматика русского языка // *Вестник С.—Петербур. ун-та*. Сер. 10. Прикладная математика, информатика, процессы управления. СПб.: Изд-во С.—Петербур. ун-та, 2004. Вып. 1-2. С. 94 – 100.
8. Падучева Е.Д. Высказывание и его соотносительность с реальной действительностью. // М.: "Наука", 1985.
9. Падучева Е. В. Семантические исследования: семантика времени и вида в русском языке. Семантика нарратива. // М.: "Языки русской культуры", 1996.
10. Булыгина Е.В., Шмелев А.Д. Языковая концептуализация мира. // М.: "Языки рус. культуры", 1997.

11. Изворска Р. Формальная семантика. // Фундаментальные направления современной американской лингвистики. М.: Изд. МГУ. 1997. С. 207-230
12. Леонтьев А.П., Леонтьева А.Л. Еще раз к вопросу о семантике генитивных отношений. // Труды Международного семинара Диалог 2006 по компьютерной лингвистике и ее приложениям. М.: Изд. РГГУ, 2006, С. 335-341.
13. Raskin V. The "How's" and "Why's" on Ontological Semantic. // Труды Международного семинара Диалог 2005 по компьютерной лингвистике и ее приложениям. С. 570-577.