

# **Интерфейс на естественном языке в интеллектуальных системах.**

***Лекция 9.***

***Специальность : 230105***

## **Основные требования к процессу понимания интеллектуальной системой запросов на естественном языке.**

- Язык общения – деловая и научная проза русского языка.**
- Форма запроса – простое распространенное вопросительное предложение русского языка.**
- Изменение базы знаний предметной области осуществляется без ведома подсистемы понимания естественного языка.**
- Тематика общения ограничена содержимым базу знаний.**

**В рассматриваемом здесь варианте подсистемы общения предполагается фиксированный формат базы знаний. “Прокачка” словаря новой лексикой на основе меняющегося содержимого базы знаний – тема отдельного исследования и здесь не рассматривается. Отдельного рассмотрения заслуживает вопрос пополнения самой базы знаний предметной области на основе входных запросов об объектах реального мира и их характеристиках, информация о которых отсутствует в базе знаний.**

**Здесь мы рассмотрим вариант предметной базы знаний на основе фреймовой сети, рассмотренный нами в предыдущей лекции.**

## **Модель языка как преобразователя “Смысл $\Leftrightarrow$ Текст”.**

**В рамках теоретического подхода “Смысл $\Leftrightarrow$ Текст” Естественный Язык (ЕЯ) рассматривается как особого рода преобразователь, который выполняет переработку заданных смыслов в соответствующие им тексты и заданных текстов в соответствующие им смыслы. Причем предполагается многозначное отображение множества текстов на множество смыслов.**

**Под текстом в рассматриваемом подходе понимается особый конструкт, являющийся дискретным представлением звучащей (устной) речи как последовательности непрерывных акустических сигналов.**

**Смысл, как и текст, представляет собой особый конструкт, но еще более сложный и далекий от наблюдения. В подходе “Смысл $\Leftrightarrow$ Текст” под смыслом понимается информация, подлежащая передаче и восприятию. Причем относительно понятия смысла вводится понятие равнозначности текстов как эквивалентности (с точки зрения носителя данного ЕЯ в пределах некоторой оговоренной точности) означаемых. Тогда смысл представляется как инвариант синонимических преобразований одних равнозначных текстов в другие.**

**Под самой моделью языка понимается некоторый гипотетический преобразователь, который существует в сознании носителя ЕЯ и устанавливает соответствие между смыслами и текстами данного ЕЯ.**

# Основные аспекты модели “Смысл $\Leftrightarrow$ Текст”.

Руководствуясь идеями Н.Хомского по формальным грамматикам как исчислениям, в модели “Смысл $\Leftrightarrow$ Текст” следует различать :

–“Лингвистическая часть”, в которой собраны сведения о конкретном ЕЯ (то есть соответствия “смыслы $\Leftrightarrow$ тексты”) и которая представляет собой исчисление, то есть набор предписаний (что можно/нельзя и что нужно делать).

–“Алгоритмическая часть”, в которой содержится описание механизма (или процедур) использования сведений о ЕЯ, представляемых “лингвистической частью”.

Модель “Смысл $\Leftrightarrow$ Текст” с формальной точки зрения есть не порождающее, а преобразующее устройство – не генератор текстов, а транслятор смыслов в тексты и обратно.

Следует отметить, что в формальной грамматике (по Н.Хомскому) *алгоритм использования грамматики для порождения или распознавания фраз должен задаваться отдельно от нее !*

## **Уровни представления высказываний и строение модели “Смысл $\leftrightarrow$ Текст”.**

**Понятие смысла, которое используется в подходе “Смысл $\leftrightarrow$ Текст”, предполагает наличие более чем одного текста, которые несут заданный смысл (явление синонимии). При этом возможна также и омонимия – наличие более чем одного смысла, которые соответствуют заданному тексту. В силу наличия в ЕЯ указанных явлений прямой одношаговый переход от смыслов непосредственно к текстам (и наоборот) оказывается практически крайне затруднительным и запутанным. С целью решения указанной проблемы в модели “Смысл $\leftrightarrow$ Текст” вводится ряд уровней промежуточных представлений ЕЯ-высказываний при переходе от смысла к тексту :**

- Семантический – запись смысла без расчленения на фразы;**
- Синтаксический и морфологический уровни – с подразделением на поверхностный и глубокий подуровень;**
- фонологический – здесь для фразы задается последовательность символов-фонем и просодем (минимальных значащих единиц, задающих интонацию фразы);**
- Фонетический – здесь для фразы задается последовательность символов, представляющих звуки и просодии (ударения, паузы и т.п.).**

# **Общая схема анализа высказывания.**

- **Морфологический Анализ (МА) слов входного предложения (см. [3,4]).**
- **Анализ результатов МА. При наличии в предложении слов-предикатов (слов таксономической категории этикеток (LABL), обозначающих ситуации (SIT), см. предыдущую лекцию), для которых определена Модель Управления (МУ) [2], вызвать модуль Семантического Анализа (СЕА.Pr), отвечающий за обработку предикатов. При отсутствии в предложении предикатов вызывается модуль Синтаксического Анализа (СИА.NP), отвечающий за обработку Именных Групп (ИГ).**
- **Для каждого слова-предиката из входного предложения Модуль СЕА.Pr определяет роли участников обозначаемой словом ситуации и указывает их возможные характеристики.**
- **СИА входного предложения с выделением предикативных, комплетивных и обстоятельственных Синтаксических Отношений (с помощью модуля СИА.КОМП).**
- **Вызов СИА.NP с выделением ИГ и проведение Синтаксических Отношений (СИО) в ИГ.**
- **Вызов модуля СЕА для проведения семантических отношений в ИГ.**

## **Представление знаний и словарь интеллектуальной системы.**

**При разработке интерфейса на естественном языке для интеллектуальной системы с фреймовой моделью в основе представления предметных знаний все используемые системой знания следует распределить между словарем (знания об используемом Естественном Языке) и фреймовой сетью.**

**Словарь содержит информацию об основах слов и словосочетаний. Словарная информация для отдельной основы состоит из заглавной, морфологической, синтаксической и семантической зон.**

**В заглавной зоне содержится запись основы; если основа может изменяться, то одна из ее форм принимается за “основную”, а остальные снабжаются отсылками к “основной”.**

**В морфологической зоне содержатся сведения, необходимые для работы процедур морфологического анализа, описанных в [4,5].**

**В словаре следует различать два вида синтактико-семантических зон : для слова-события и для слова-понятия. Зона для события содержит описание МУ и значений Лексических Функций (ЛФ) слова. Зона для понятия содержит поле предлогов и поле значений ЛФ.**

# Лексические функции.

Рассматриваемый подход к языку как преобразователю “Смысл $\leftrightarrow$ Текст” использует большое количество сведений о каждом слове рассматриваемого ЕЯ (в нашем случае русского). Помимо МУ, для каждого заглавного (т.е. ключевого) слова  $C_0$  указываются все слова (или словосочетания), определенным образом связанные с ним по смыслу, а именно :

- 1) “парадигматические варианты” слова, или его “замены” – средства, которые могут или должны заменять  $C_0$  в тех или иных контекстах и при тех или иных условиях;
- 2) “синтагматические партнеры” слова, или “параметры”, - средства, которыми идиоматично, т.е. несвободно выражаются при данном слове некоторые смыслы.

Определение. Под лексическим коррелятом понимается парадигматический вариант, либо синтагматический партнер слова. С помощью Лексической Функции (ЛФ) описывается зависимость, связывающая слово и его лексический коррелят.

ЛФ  $f$  описывает зависимость, определяющую для некоторого слова или словосочетания  $X$  такое множество слов или словосочетаний  $\{Y_i\}=f(X)$ , что для любых  $X_1$  и  $X_2$  верно следующее : если  $f(X_1)$  и  $f(X_2)$  существуют, то между  $f(X_1)$  и  $X_1$ , с одной стороны, и между  $f(X_2)$  и  $X_2$ , с другой стороны, всегда имеет место одно и то же смысловое отношение.  $X$  – это аргумент лексической функции  $f$ , а  $\{Y_i\}$  – ее значение, или выражение. ЛФ вводятся для описания лексической сочетаемости, а не семантики, поэтому для некоторых ЛФ выражающие их слова не находят отражения в семантическом графе высказывания.



# Классификация и примеры Лексических Функций.

Среди всех возможных ЛФ с точки зрения лексической сочетаемости наибольший содержательный интерес представляют те ЛФ, у которых имеются фразеологически связанные выражения, то есть те ЛФ, у которых есть выражения, допустимые при одних аргументах, но не допустимые при других. ЛФ же, у которых все выражения возможны при любых аргументах, или ЛФ-константы, не представляют содержательного интереса с точки зрения описания лексической сочетаемости.

Среди ЛФ, имеющих фразеологически связанные выражения, различаются стандартные и нестандартные ЛФ. Стандартная в данном языке ЛФ (СЛФ) должна – в отличие от нестандартной ЛФ – удовлетворять двум следующим требованиям :

- 1) Вне зависимости от используемого ЕЯ СЛФ должна быть определена для достаточно большого числа аргументов.
- 2) СЛФ характеризуется достаточным богатством ее языковых выражений.

Примерами СЛФ могут послужить Syn – синоним (Syn(“эксперимент”)=“опыт”), Anti – антоним (Anti(“простой”)=“сложный”), S<sub>0</sub>, A<sub>0</sub>, Adv<sub>0</sub>, V<sub>0</sub> - синтаксические дериваты от C<sub>0</sub>, т.е., соответственно, существительное, прилагательное и глагол, имеющие тот же смысл, что и C<sub>0</sub> (V<sub>0</sub>(“эксперимент”)=“экспериментировать”, A<sub>0</sub>(“эксперимент”)=“экспериментальный”, Adv<sub>0</sub>(“эксперимент”)=“в порядке эксперимента”).

В дальнейшем мы будем рассматривать только СЛФ.

## Этап синтаксического анализа входного предложения.

**Определение.** Словоформой будем называть отрезок текста между двумя соседними пробелами. При этом знаки препинания считаются отдельными словоформами.

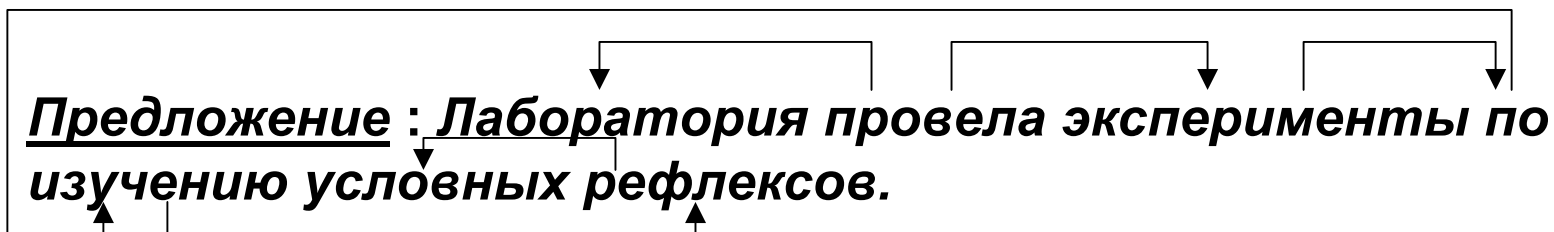
**Определение.** Основной уровень предложения образуют словоформы, не входящие в обороты и придаточные предложения.

На этапе СИА на основе полученной на этапе МА морфологической информации словоформ и информации из словаря строится синтаксическая структура входного предложения. Фактически СИА устанавливает синтаксические отношения не между словоформами (или словосочетаниями), а между характеризованными обобщенными лексемами на выходе МА. Мы будем рассматривать представление синтаксической структуры в виде дерева зависимостей (= дерева синтаксического подчинения). Подавляющее большинство предложений русского языка описываются проективными деревьями.

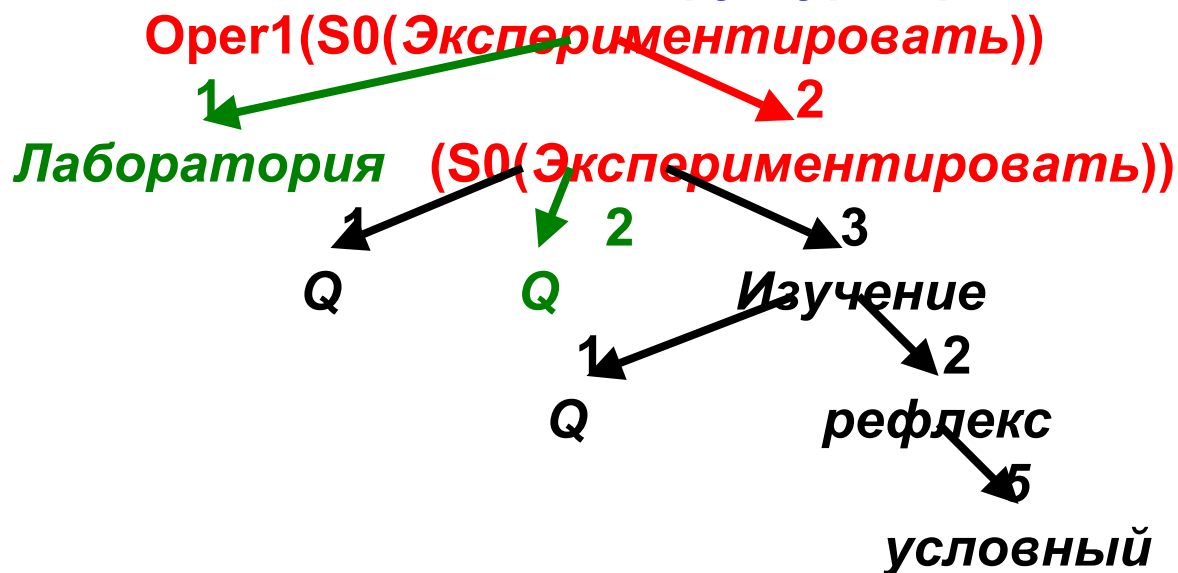
**Определение.** Пусть  $x$  – произвольная непустая цепочка и  $X$  – множество всех входящих в  $x$  точек,  $\rightarrow$  есть отношение зависимости как произвольное бинарное отношение на  $X$ , при котором граф  $\langle X, \rightarrow \rangle$  является деревом. Считается, что дерево зависимостей  $\langle X, \rightarrow \rangle$  для цепочки  $x$  (а также соответствующее отношение подчинения) называется проективным, если для любых трех точек  $\alpha, \beta$  и  $\gamma$  цепочки  $x$  из того, что  $\alpha \rightarrow \beta$  и  $\gamma$  лежит между  $\alpha$  и  $\beta$ , следует, что  $\gamma$  зависит от  $\alpha$ . Графически для цепочечного способа изображения дерева зависимостей проективность означает возможность проведения дуги (ветви) дерева таким образом, чтобы : 1) никакие две из них не пересекались и 2) корень дерева не лежал ни под одной из дуг. Если первое из условий соблюдается, а второе нет, то говорят о слабой проективности. Построение синтаксической структуры проективности.

В научной и деловой прозе подавляющее большинство предложений проективны. Построение синтаксической структуры проективных предложений упрощается по сравнению с непроективными предложениями за счет близости синтаксически связанных слов по положению в предложении. Так, в проективном предложении ИГ может быть образована только из смежных слов, а разрыв самой ИГ не допускается.

К понятию проективности : пример дерева глубинного синтаксиса для простого распространенного предложения русского языка.



Глубинная синтаксическая структура предложения :



Обозначения :

 Поддерево, заменяемое лексическим правилом\*

 Поддерево, заменяемое синтаксическим правилом\*

\*Примечание. Здесь имеются в виду лексические и синтаксические правила синонимических преобразований деревьев глубинного синтаксиса теории языка как преобразователя “Смысл $\leftrightarrow$ Текст”.

# **Основные принципы построения алгоритма синтаксического анализа.**

**Метод фильтров – традиционный метод построения дерева зависимостей фразы русского языка. Следует сформулировать ряд общих рекомендаций по применению синтаксических фильтров для ускорения выбора правильного варианта анализа предложения :**

- 1. Использование глобальных фильтров для отсева слов, между которыми не могут быть установлены синтаксические отношения.**
- 2. Применение фильтров в процессе выбора цепочки характеризованных лексем в соответствии с рассматриваемым вариантом синтаксического разбора.**
- 3. Использование семантической информации из словаря в процессе построения синтаксической структуры запроса.**
- 4. Использование параллельного перехода от этапа СИА к этапу СЕА.**

## Основные типы синтаксических фильтров.

1). Отсеивающие фильтры – реализуются в виде требований к способам поверхностной реализации и семантической интерпретации глубинных синтаксических актантов предикатного слова, описываемых его МУ. При применении отсеивающих фильтров обязательной является проверка наличия во входном предложении словоформ, обуславливающих изменение падежей, требуемых способом поверхностной реализации того или иного актанта. К числу таких словоформ русского языка относятся количественные числительные и отрицательные частицы. Количественное числительное (не типа 1) изменяет падеж следующего за ним существительного с именительного на родительный (ср. “Завод арендует у колхоза земельные участки” → “Два завода арендуют у колхоза земельные участки”). Стоящая перед предикатом-глаголом или предикатом-кратким прилагательным отрицательная частица может изменять падеж существительного, связанного с данным предикатом отношением СУБЪЕКТ (Sub, тип 1 отношения глубинного синтаксиса [2,3]), с именительного на родительный (ср. “Доцент Иванов был на заседании кафедры” → “Доцента Иванова не было на заседании кафедры”).

2). Локальные фильтры. Пример – проведение обстоятельственного отношения (ему соответствует тип отношения глубинного синтаксиса 5, см. [3]). Этот тип отношения может быть проведен на существительное без предлога, если оно стоит в творительном падеже.

3). Фильтры для устранения непроективности при выделении ИГ. Пример – непроективность в вопросительных предложениях с вопросительными местоимениями “СКОЛЬКО” и “КАКОЙ”. Так, для вопроса со словом “СКОЛЬКО” синтаксический фильтр определяет наличие непосредственно за данным словом слова в родительном падеже. Если такого слова там нет, то подпрограмма устранения непроективности будет искать в предложении слово в родительном падеже, которое подходило бы по Семантическому Классу (СК) под отношение СУБЪЕКТ, либо первое комплетивное поверхностно-синтаксическое отношение, которому на уровне глубинного синтаксиса соответствует тип отношения 2, см.[3].

4). Фильтры для выделения границы однородной группы по запятым, союзам, единству морфологических характеристик.

## Общая структура алгоритма СИА (без рассмотрения оборотов).

- 1) Выделить из множества  $S$  словоформ основного уровня подмножество  $P$  слов-предикатов, для которых определены МУ.
- 2) Для каждого слова-предиката  $p \in P$  выбрать список слов-кандидатов, удовлетворяющих требованиям  $МУ(p)$  к способам поверхностной реализации и семантической интерпретации актантов с применением отсеивающих фильтров.
- 3) Провести обстоятельственные отношения (см. предыдущий слайд) с применением локальных фильтров.
- 4) Выделить ИГ с проведением определительных и других синтаксических отношений, отличных от связи предиката с его актантами по МУ посредством предикативного и комплетивных отношений (типы 1, 2, 3 и 4) и соответствующих типу Глубинного Синтаксического Отношения (ГСО) 5, см. [3]. Определительное отношение устанавливается между главным словом-существительным и зависимым от него словом, которое может быть прилагательным, причастием, определительным наречием и существительным. В случае неоднозначности, вызванной наличием у слова МУ (причастие, отглагольное существительное и т.п.) используются семантические фильтры в виде требований к семантическому классу слова, на которое проводится данный вид отношений.
- 5) С помощью синтаксических фильтров обработать имеющиеся в предложении случаи непроективности.

## Семантический анализ запроса.

На этапе Семантического Анализа (СЕА) синтаксическая структура запроса преобразуется во внутреннее представление, называемое Семантическим Графом (СГ). СГ состоит из множества вершин-понятий, связанных между собой через вершины-предикаты. Часть дуг СГ имеют тот же смысл и именуется так же, как и дуги фреймовой сети. Поскольку сам фрейм мы рассматриваем как сеть из нескольких вершин и отношений, то для каждого отдельного фрейма в результирующем СГ мы получим подграф с дугами, именуемыми обозначениями имен слотов рассматриваемого фрейма.

## **Этап семантического анализа предложения.**

**На вход алгоритма Семантического анализа предложения (СЕАП) поступает обобщенная синтаксическая структура (=обобщенное дерево разбора) предложения. Эта структура содержит :**

- 1) Для каждого слова, имеющего Модель Управления (МУ), из возможных МУ омонимичных слов должна быть выбрана МУ в соответствии с ситуацией употребления рассматриваемого слова в тексте.**
- 2) Альтернативные варианты установления обстоятельственных отношений (из рассматриваемых на уровне глубинного синтаксиса это тип отношения 5).**
- 3) Альтернативные границы Именных Групп (ИГ) и альтернативные варианты разбора каждой ИГ.**
- 4) Для каждого слова входного предложения – информацию о числе и роде (используется при семантической обработке местоимений).**

**Мы будем рассматривать двухэтапный СЕАП : на первом этапе строится неструктурированный Семантический Граф (СГ) предложения, на втором происходит структурирование полученного СГ [1].**



## **Смысловые валентности предикатного слова.**

**В словарной статье Русского общесемантического словаря (РОСС) семантические валентности предикатного слова описываются в поле ВАЛ (VAL). Имена смысловых валентностей (названия ролей обозначаемых актантами сущностей) могут быть напрямую отображены в названия дуг Семантического Графа.**

**Используемый в идеологии РОСС и основанных на ней системах ФРАП-ПОЛИТЕКСТ-ДИАЛИНГ набор смысловых валентностей в общих чертах сходен с описанными Ю.Д. Апресяном в [2] (см. Таблицу 1).**

**В системах ФРАП-ПОЛИТЕКСТ-ДИАЛИНГ, помимо семантически наполненных отношений, используются два вспомогательных отношения : F\_АСТ (П\_АКТ, первый актант) и S\_АСТ (В\_АКТ, второй актант), которые используются для описания валентностных структур слов-отношений (слов таксономической категории ОТН (RELAT)).**

# Ролевые ориентации семантических валентностей.

Таблица 1.

Апр есян	Пример	ФРАП-ПОЛИТЕКСТ-ДИАЛИНГ
Sub(субъект)	поезд движется	СУБ (Поезд, двигаться)
Contrag (контрагент)	покупать у старьевщика	К-АГЕНТ (старьевщик, покупать)
Cap (глава)	вина перед коллективом	
Obj (объект)	гладить руку, стрелять в мишень	ОБ(рука, гладить)КОН-Т (мишень, стрелять)
Content (содержание)	знать об отъезде	СОДЕРЖ(отъезд, знать)
Adr (адресат)	сообщать президенту	АДР (президент, сообщать)
Recip (получатель)	давать детям, дарить людям	АДР (дети, давать)АДР (люди, дарить)
Via (посредник)	передовать через секретаря	ПОСРЕД(секретарь, передавать)
Is (источник)	брать в кассе	ИСХ-Т (касса, брать)
Loc (место)	находиться в лесу	ЛОК(лес, находиться)
Ab (начальная точка)	вывести из леса	ИСХ-Т(лес, вывести)
Ad (конечная точка)	везти в город	КОН-Т(город, везти)
Itin (маршрут)	идти по дороге	

## Ролевые ориентации семантических валентностей (продолжение).

Таблица 1 (продолжение).

Адресант	Пример	ФРАП-ПОЛИТЕКСТ-ДИАЛИНГ
Med (средство)	прибивать гвоздями	СРЕДСТВО (гвозди, прибивать)
Instr (инструмент)	резать ножом	ИНСТР (нож, резать)
Mod(способ)	обращаться плохо	АСПЕКТ (плохо, обращаться)
Cond (условие)	если P, то Q	УСЛ (P,Q)
Motiv(мотивировка)	награждать за храбрость	ПРИЧ(храбрость, награждать)
Caus (причина)	радоваться подарку, проистекать из-за Q	ПРИЧ (подарок, радоваться) ПРИЧ(Q, проистекать)
Result (результат)	превращать в воду	РЕЗЛТ(вода, превращать)
Dest (цель)	стремиться к общему благу	ЦЕЛЬ (благо, стремиться)
Asp (аспект)	превосходить по качеству	ОГРН (качество, превосходить)
Quant (количество)	пять человек	КОЛИЧ(пять, человек)
Period (срок)	отпуск на два месяца	ВРЕМЯ (на два месяца, отпуск)
Temp (время)	начаться в полночь	ВРЕМЯ (полночь, начаться)

# Особенность фреймовой сети.

С помощью фреймовой сети описываются знания системы об окружающем мире. Во фреймовой сети мы выделим (см.[1], стр.333) Абстрактную и Конкретную Сети (АС и КС).

АС (см.[1], стр. 333) хранит информацию о наиболее общих понятиях, событиях и характеристиках моделируемой предметной области (соответствует уровню фреймов классов). Среди вершин АС мы будем различать ПОНЯТИЯ, СОБЫТИЯ и ХАРАКТЕРИСТИКИ.

КС описывает конкретные объекты как реальную интерпретацию представленных в АС понятий реального мира (соответствует уровню фреймов-шаблонов).

Фреймовая сеть есть разновидность семантической сети. Отличительные особенности :

- Блочная структура информации в вершинах;
- В качестве семантических отношений допускаются “род-вид” (IS\_A) и “часть-целое” (PART\_OF)

## Фреймовое представление ситуации и Модель Управления.

В ряде работ по Искусственному Интеллекту, представлению и инженерии знаний [8-10] под фреймом понимается структура данных (образ), связанных с концептуальными объектами в памяти и необходимых для представления некоторой типичной/стереотипной ситуации. В частности, фрейм-ситуация [6] формируется представлениями о прототипической ситуации и ее элементах, имеющих фиксированные роли и положения, что соответствует описанию семантических актантов обозначающего ситуацию слова.

Учитывая тот факт, что описание предметной области, как правило, содержит более конкретную информацию о кандидатах на исполнение ролей обозначаемой словом ситуации, чем МУ, во фреймовых единицах, описывающих ситуации, будут описываться понятия, заполняющие актантные места в соответствии с ролью того или иного элемента ситуации. Соответствие семантического класса элемента ситуации, описываемой во фреймовой единице, и семантического класса актанта по МУ обозначающего ситуацию слова устанавливается путем проверки наличия родовидового отношения (с учетом транзитивности) между понятиями. Пример : *“предмет длительного пользования” IS A “предмет как результат труда человека”, “предмет как результат труда человека” IS A “материальный объект”, следовательно, “предмет длительного пользования” IS A “материальный объект”*.

# Основные группы предикативных слов деловой прозы.

- Предикаты действия – характеризуют состояние объектов моделируемого мира. Применительно к фреймовым сетям предикат действия должен иметь (как минимум !) ссылку на имя фрейма. Примеры : *заем, аренда, продажа* и т.п.
- Предикаты функционального типа – указывают на операции, которые необходимо выполнить над предикатами действия при обработке входного предложения. Пример : *сравнить, найти, вычислить*. Функциональные предикаты не имеют представления в семантической сети. Они вырабатывают обращение к процедуре, которая производит обработку предикатов действия (на этапе обработки).
- Пустые предикаты – к ним относятся глаголы-связки (*быть, есть*), модальные глаголы (*мочь, должен*) и другие глаголы, либо не несущие реальной семантической нагрузки (вообще или в данной предметной области), либо несущие информацию не о самом действии, а о характере его протекания. Пустым предикатам обычно не соответствуют никакие вершины в сети.

## **Обработка пустых и функциональных предикатов.**

**Состоит в устранении указанных предикатов путем перестройки синтаксической структуры высказывания с переносом исходящих из этих предикатов синтаксических отношений на предикат действия. Перестройка ведется с применением лексических и синтаксических правил перифразирования, см.[3], стр.141-176, а также [2], стр. 316-345.**

**Смысловые соотношения между цельными лексическими единицами описываются с помощью аппарата стандартных ЛФ. Поэтому деревья синтаксического подчинения (более точно – деревья глубинного синтаксиса), которым отвечают одни и те же семантические представления, а синонимия описывается с помощью аппарата стандартных ЛФ, принято называть ЛФ-синонимичными. Специальное исчисление в виде системы правил, которое любому заданному дереву ставит в соответствие все другие деревья, ЛФ-синонимичные с ним, получило название системы перифразирования [1,2].**

**Известные в теории “Смысл $\leftrightarrow$ Текст” синонимические преобразования деревьев Глубинных Синтаксических Структур (ГСС) сводятся к синонимическим заменам либо лексики (узлов), либо синтаксических связей (ветвей), либо – и это наиболее общий случай – одновременно того и другого.**

**Ситуации употребления в тексте пустых предикатов описываются посредством следующих стандартных ЛФ :**

- Глаголы-связки – с помощью ЛФ Sorul;**
- Пустые глаголы – с помощью ЛФ Orei, Func и Labor;**
- Фазовые глаголы – с помощью ЛФ Incer, Cont и Fin.**

## ЛФ для описания ситуаций употребления пустых предикатов.

*Copul* – связка : *Copul*(“получатель”) = “быть”. “являться”.

*Pred* – глагол, обозначающий “быть...”, то есть соответствующий “склеенной” комбинации *Copul*(*C0*) с *C0*. Пример : *Pred*(“рядом”) = “соседствовать”.

*Oper<sub>i</sub>* – 1-м актантом глагола (и его подлежащим) является *i*-й участник ситуации, а 2-м глубинным актантом (и первым дополнением) – имя самой ситуации (прочие актанты, если они есть – обозначения прочих участников ситуации) :

*Oper1*(“аренда”) = “сдавать” [что-л. в аренду],

*Oper2*(“аренда”) = “брать” [что-л. в аренду].

*Func<sub>i</sub>* – 1-м актантом глагола (и его подлежащим) является имя ситуации, 2-м глубинным актантом (и первым дополнением) – ее *i*-й участник.

*Labor<sub>ij</sub>* – 1-м актантом глагола (и его подлежащим) является *i*-й участник ситуации, 2-м глубинным актантом – *j*-й участник ситуации, а 3-м глубинным актантом (и вторым дополнением) – имя самой ситуации : *Labor<sub>32</sub>*(“аренда”) = “сдавать” [что-л. в аренду].

*Oper*, *Func* и *Labor* соотносятся друг с другом как конверсивы.

ЛФ для передачи значений фазовых глаголов :

*Inser* – ‘начинаться’, *Cont* – ‘продолжаться’, *Fin* – ‘переставать’.

*Gener* – такое родовое понятие, что ‘*Gener* + *C0*’ = ‘*C0*’, где *C0* – заглавная лексема. Пример : “передача в распоряжение” ← “ситуация передачи в распоряжение”.



# Лексические синонимические преобразования с целью устранения пустых предикатов.

Для выполнения ЛФ-синонимических замен с применением указанных ЛФ с учетом типов возможных вопросительных ситуаций могут быть использованы следующие лексические правила из описанных в [3] :

Расщепления со связкой :

$$C_0 \Leftrightarrow Pred(C_0) \quad (\text{лексическое правило № 13, [3], стр. 153})$$

$$Pred(C_0) \Leftrightarrow Copul(C_0) \xrightarrow{2} C_0 \quad (\text{лексическое правило № 14, [3], стр. 153})$$

Расщепление с  $Oper_1$  :

$$C_0 \Leftrightarrow S_0(C_0) \xleftarrow{2} Oper_1(S_0(C_0)) \quad (\text{лексическое правило № 17, [3], стр. 154})$$

Расщепление с  $Gener$  :

$$C_0 \Leftrightarrow Gener(C_0) \xrightarrow{5} Der(C_0) \quad (\text{лексическое правило № 28, [3], стр. 155})$$

$$S_0(Pred(C_0)) \Leftrightarrow Gener(C_0) \xrightarrow{2} C_0 \quad (\text{лексическое правило № 29, [3], стр. 155})$$

И аналогичное ему правило (3), описанное в [2], стр.320 :

$$X \Leftrightarrow Gener(X) \xrightarrow{2} A_0(X)$$

$$X \Leftrightarrow Gener(X) \xrightarrow{2} S_0(X)$$

## **Фреймовое представление предметных знаний : отличия от семантической сети.**

Множество понятий (семантических классов слов [2,3]), заполняющих актантные места предиката в заданном контексте, могут указываться во фреймовой сети только в том случае, если фреймом описывается сама МУ (ср. также употребляемое в отечественной литературе понятие “фрейм модели управления”). В этом случае каждому понятию, заполняющему то или иное актантное место, будет соответствовать значение слота. Имя слота в этом случае будет соответствовать ролевой ориентации актанта (Sub, Object, Contrag и т.п.).

В отдельных случаях, как при описании Лексических Значений (см. *Лекцию 8*), состав семантических валентностей предиката описывается посредством одного слота типа TABLE.

Таким образом, применительно к фреймовым сетям слово с МУ следует идентифицировать как предикат действия тогда и только тогда :

- В его словарной статье есть ссылка на вершину фреймовой сети (имя фрейма);
- Имена слотов фрейма соответствуют названиям ролей актанта слова по его МУ, либо задан особый слот типа TABLE, в котором описывается актантная структура с конкретизацией семантического наполнения актантов в соответствии с ролевой ориентацией по МУ;

## Типы вопросительных ситуаций для запросов к фреймовой системе.

- Спрашивается о наличии характеристики, интерпретирующей понятие, например : *“Характерна ли каузация для Лексического Значения “Передача в распоряжение” ?”*.
- Спрашивается значение некоторой характеристики, определяющей множество объектов, например : *“Какова Модель Управления у глаголов Лексического Значения “купля-продажи” ?”*.
- Применительно к одноместным характеристикам спрашивается значение параметра некоторой характеристики, например : *“Какое количество семантических актантов имеют предикатные слова Лексического Значения “приобретение права на временное пользование” ?”*. Подобные параметры можно выделить у характеристик, имеющих тип LISP, TEXT, TABLE, либо EXPRESSION. Причем значение запрашиваемого параметра связывается с типом значения рассматриваемой характеристики (например, содержанием ячейки таблицы для типа TABLE). Для получения значения искомого параметра используются стандартные функции по работе со строками, столбцами и т.п.

## **Задача построения неструктурированного семантического графа предложения.**

**Задача данного подэтапа состоит в выборе семантически осмысливаемого (то есть не противоречащего информации фреймовой сети) варианта синтаксического разбора входного предложения, построении для него неструктурированного СГ и выделения кванторных слов-предикатов.**

**Функционально в данном этапе мы будем выделять обработку слов, имеющих МУ, и обработку ИГ (ИГ включают слова, не имеющие МУ). Такое деление этапа обусловлено тем, что слова с МУ и Именные Группы несут в предложении разные семантические нагрузки и для их обработки необходимы различные методы.**

**Лексическое значение слова, имеющего МУ, определяется самой МУ. Семантическая информация слова с МУ задается в самой МУ, то есть в словаре. Для таких слов информация из фреймовой сети используется только для подтверждения правильности сделанного алгоритмом СЕАП выбора.**

**В отличие от слов, имеющих МУ, семантические отношения, связывающие слова в ИГ, определяются в основном контекстом, то есть с помощью фреймовой сети.**

# Структурирование семантического графа.

Целью этапа структурирования (квантификации) СГ является замена имен индивидуальных сущностей в виде языковых выражений, кванторными словами, которые позволяют выделить среди множества сущностей те, которые задаются рассматриваемым выражением.

С точки зрения синтаксиса выделяются следующие группы кванторных слов :

- 1) Кванторные прилагательные – всякий, все, каждый, какой-нибудь, некоторый, никакой, какой бы то ни было, какой-то, кто-то, некто, некий и т.п.
- 2) Кванторные существительные – все, кто-нибудь, кто-либо.
- 3) Кванторные наречия – всегда, иногда, когда-нибудь, когда-либо.
- 4) Кванторные глаголы – существует, имеется, есть, нет.
- 5) Фразеологические сочетания – каков бы ни был, что бы ни, где бы ни, когда бы ни и т.п.

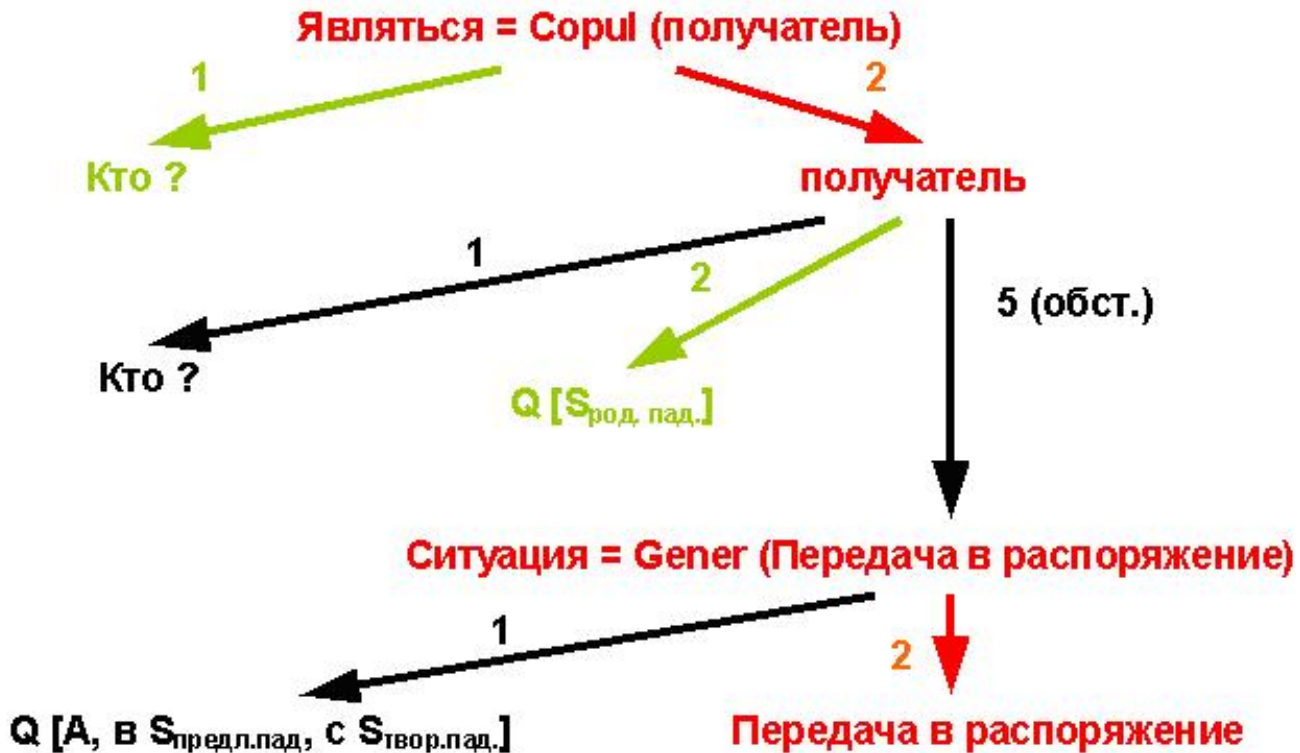
С учетом показанных особенностей запросов к фреймовой системе мы опускаем этап структурирования СГ, для интерпретации запроса достаточно иметь неструктурированный СГ, в котором выделяется вопросительное слово.

## Пример построения СГ запроса. Исходная ГСС.

Проиллюстрируем работу предлагаемого нами подхода к анализу ЕЯ-запросов на примере запроса : “Кто является получателем в ситуации передачи в распоряжение ?”.

В рассматриваемом примере имеем следующие слова с МУ : “являться”, “получатель”, “ситуация”

Глубинная синтаксическая структура запроса :

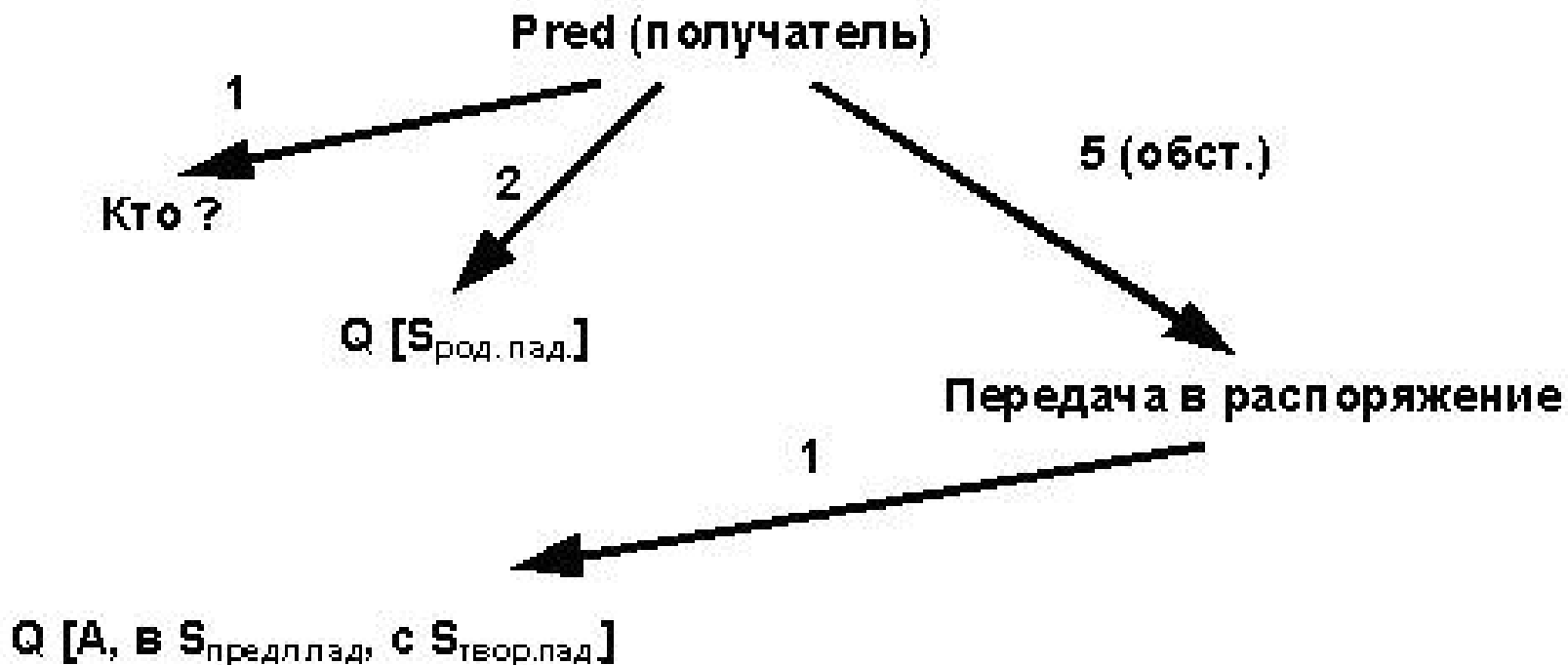


Условные обозначения :

- - Поддереву, заменяемое лексическим правилом;
- - Поддереву, заменяемое синтаксическим правилом.

# Преобразованная ГСС.

Применение лексического правила №14 с обслуживающим его синтаксическим правилом №6 из описанных в [3] позволяет устранить глагол-связку “является”, а применение модифицированного лексического правила №29 дает возможность склеить в один узел “ситуацию” и “передачу в распоряжение”, которые связаны родовидовыми отношениями. Преобразованная ГСС будет иметь вид :



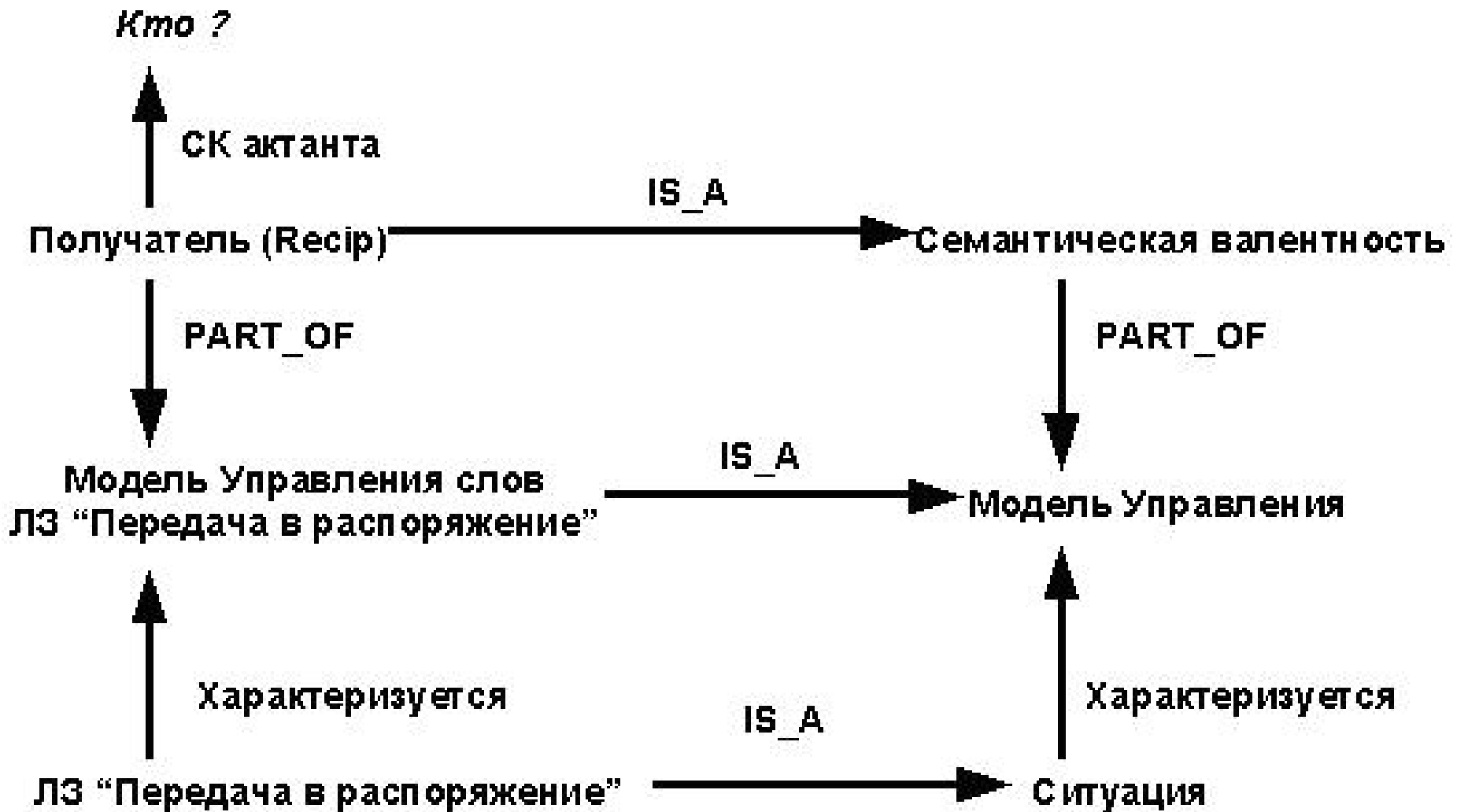
## **Замена обстоятельственного отношения.**

**Отличительной особенностью замены обстоятельственных отношений семантическими при обработке слов с МУ для запросов к фреймовой сети является то, что обстоятельственное отношение может развернуться в подграф в соответствии с типом данных слота.**

**В рассматриваемом примере запроса спрашивается значение параметра характеристики “Модель Управления”. Данная характеристика описывается слотом типа TABLE, названия строк таблицы соответствуют ролевым ориентациям семантических валентностей. Получатель – это ролевая ориентация семантической валентности. Помимо ролевой ориентации, семантическая валентность характеризуется семантическим классом слова, которое может ее замещать. Следовательно, в нашем случае вопрос был задан к параметру “валентность получателя” характеристики “Модель Управления” : спрашивалось о значении этого параметра, то есть значение Семантического Класса актантного значения. С учетом сказанного СГ для нашего запроса будет иметь представленный на следующем слайде вид.**



# СГ запроса.



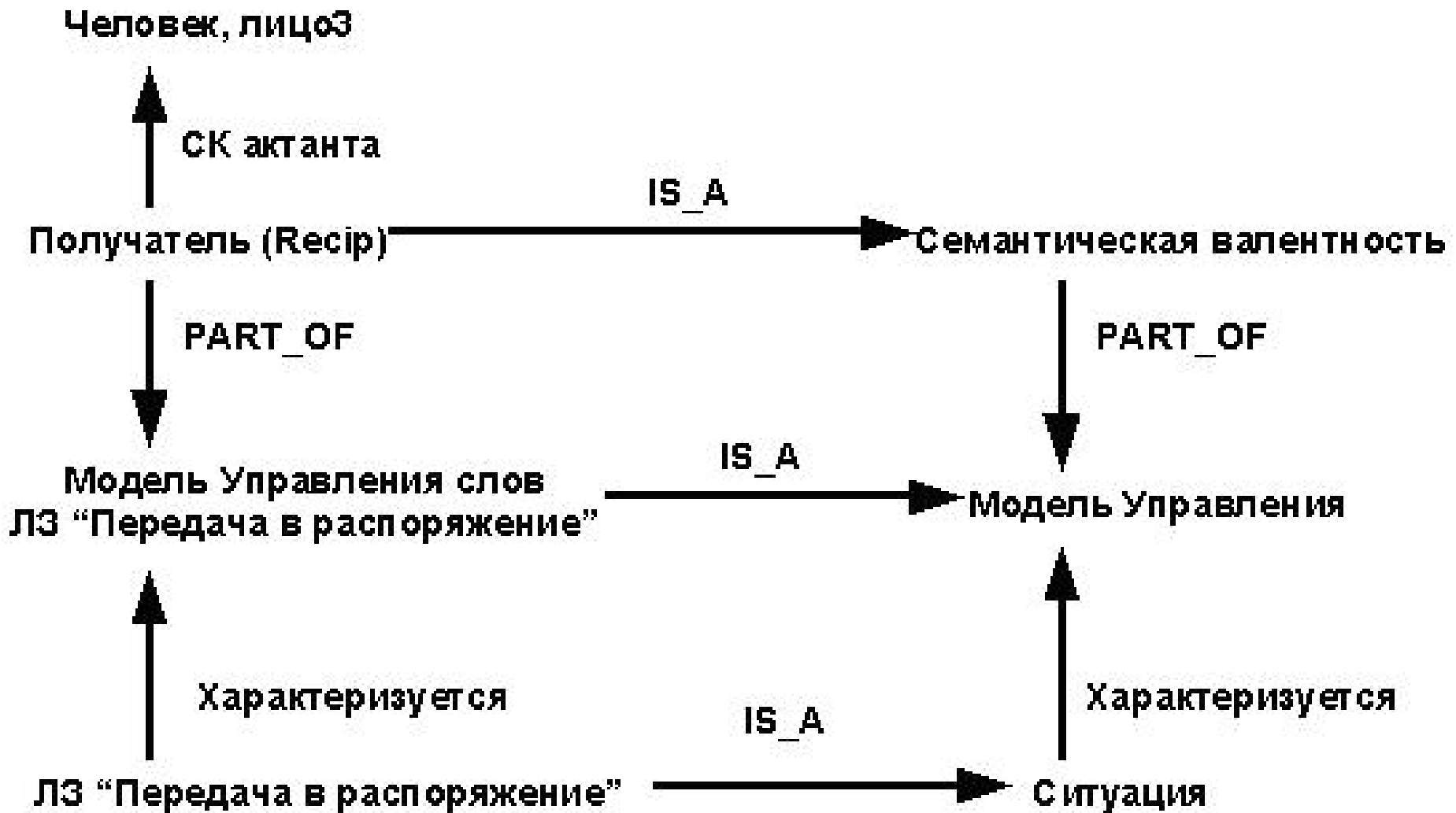
## **Интерпретация входного предложения.**

**На этапе интерпретации происходит сопоставление семантического графа запроса с фреймовой сетью. Сам процесс сопоставления является процессом свертывания семантического графа начиная от вершин, наиболее удаленных от вопросительной (вершин, не имеющих дочерних и имеющих ссылки на вершины фреймовой сети в своих словарных статьях) и кончая вопросительной вершиной.**

**СГ ответа формируется с использованием СГ запроса внесением в граф смысловой информации, полученной на этапах синтаксического (глубинно-синтаксического) анализа и преобразования ГСС в СГ. В простейшем случае задача формирования СГ ответа состоит в следующем :**

- Определить с помощью МУ семантические падежи запрашиваемого события, которые должны быть указаны в синтезируемой фразе;**
- Определить, какую информацию о каждом участнике (=заполнитель семантического падежа в виде СК актанта по МУ в соотношении с информацией фреймовой единицы) события;**
- Определить, допустимо ли стилистически одновременное одновременное выражение всех участников (и их характеристик) данного события в одном предложении русского языка.**

# СГ ответа для рассматриваемого примера.



## **Синтез синтаксической структуры ответа.**

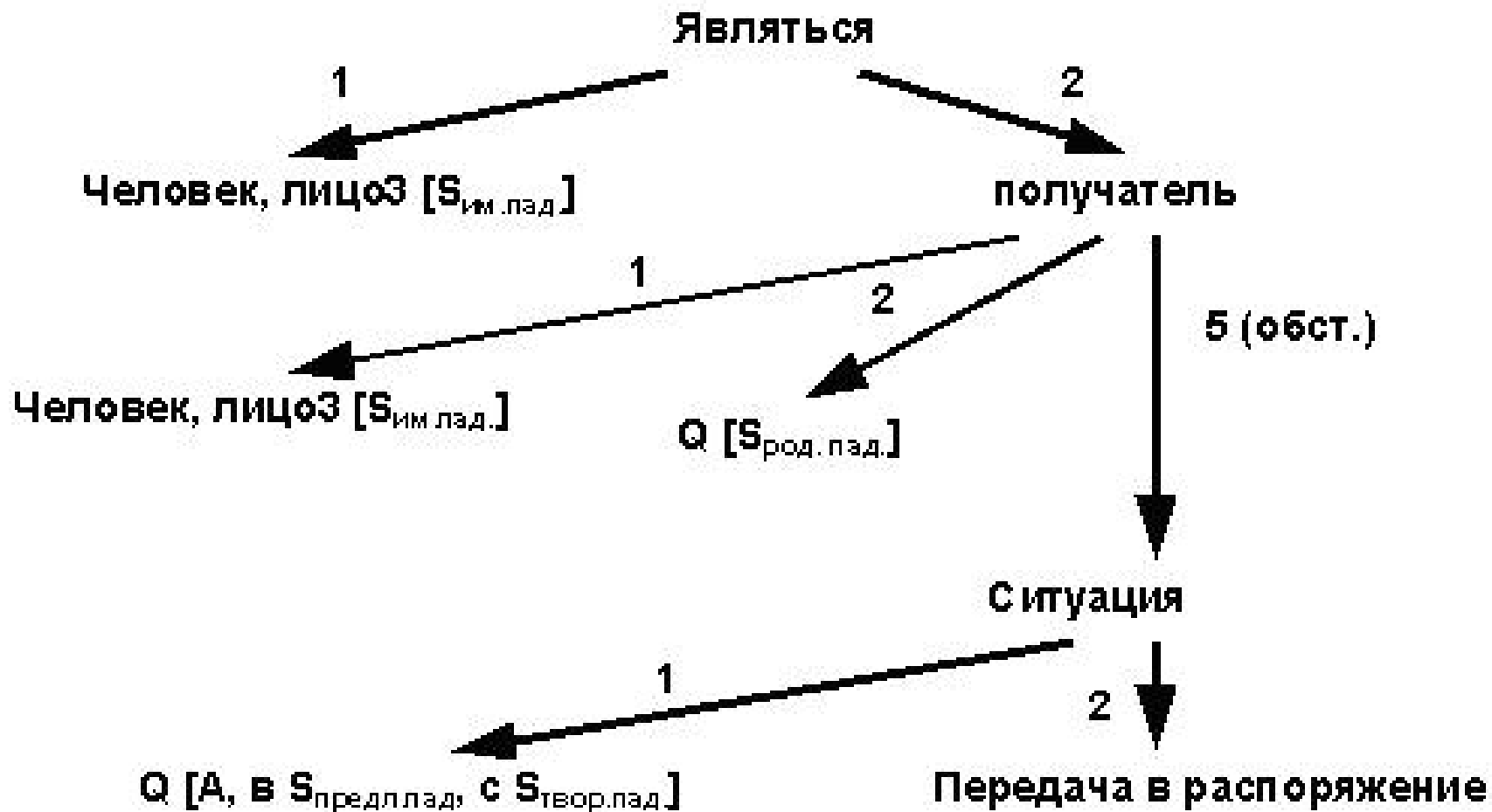
**Дерево синтаксической структуры ответа строится путем лексикализации информации вершин СГ (поиска по словарю слов со ссылками на представленные в вершинах события, понятия и их характеристики) с сопутствующей заменой представленных в СГ семантических отношений на синтаксические.**

**На практике процесс выбора вариантов преобразования фрагмента СГ в поддереву синтаксической структуры представляет собой поиск с возвратом. Для уменьшения возможного перебора при наличии между словами исходного запроса обстоятельственных отношений, которые в СГ отображаются в подграф (как в рассмотренном примере), каждое подобное развертывание обстоятельственного отношения “запоминается” для более быстрого отыскания “пути назад” при построении синтаксической структуры ответа.**

**В случае невозможности лексикализации информации тех или иных вершин СГ в синтезируемом дереве синтаксического подчинения используют правила синонимических преобразований деревьев, используемые при обработке пустых и функциональных предикатов.**

**Полученная на основе СГ ГСС ответа (в соответствии с информацией МУ слов “являться”, “получатель” и “ситуация”) представлена на следующем слайде.**

# ГСС ответа для рассматриваемого примера.



# Определение порядка слов в синтезируемом предложении.

Цель : установить порядок во отношении “правее”, “левее” между стоящими в вершинах дерева словоформами синтезируемого предложения. Указанная задача может быть выполнена в три этапа (см. [1], [3]) :

- 1) Формирование Исходных Групп (ИсГ) – последовательностей слов, связанных “локальными” (=смежными) синтаксическими отношениями.
- 2) Объединение исходных групп в Приведенные Группы (ПГ) – последовательности ИсГ, связанных “квазилокальными” синтаксическими отношениями. ПГ соответствует группам основных членов предложения : группе подлежащего, группе сказуемого, группе дополнений и т.п.
- 3) Расстановка ПГ, связанных “глобальными” синтаксическими отношениями.

Каждый из указанных этапов выполняется с помощью специальных правил. Первая задача может быть решена с помощью автоматной грамматики. Будем задавать состав и строение ИсГ так называемыми “обоймами”, каждая из которых отвечает одному из типов ИсГ : именные (ИсГ<sub>S</sub>), адъективные (ИсГ<sub>A</sub>), глагольные (ИсГ<sub>V</sub>) и наречные (ИсГ<sub>Adv</sub>). Предложные ИсГ рассматриваются как разновидность именных. ИсГ состоит из слова (ядра группы) и всех слов, связанных с ним “локальными” синтаксическими отношениями. В большинстве случаев компоненты ИсГ являются подчиненными (=слугами) ядра группы (исключая предлоги, союзы и т.п.). Обойма задает максимальный состав возможных ИсГ (данного типа) русского языка. На следующем слайде в таблице 1 приведен порядок членов в различных типах ИсГ.

## Порядок членов в различных типах ИсХ русского языка.

Таблица 2

Номер позиции в ИсГ	Тип ИсГ			
	ИсГ <sub>S</sub>	ИсГ <sub>A</sub>	ИсГ <sub>V</sub>	ИсГ <sub>Adv</sub>
1	Сочинительный союз			
	(и)	(и)	(и)	(и)
2	Отрицание / отрицательная частица			
	(не только)	(едва ли не)	(только)	(отнюдь не)
3	Предлог (для)	Наречие определительное		
		(очень)	(крайне)	(очень)
4	Кванторное слово (всех)	Прилагательное (уставший)	Глагол (волнуется)	Наречие (далеко)
5	Указат. прилаг. (этих)	<p><b>Примечание.</b> Необходимо иметь в виду, что обойма задает “грубый порядок слов”. Для учета нюансов русского языка используются уточняющие правила. Пример – неоднородные прилагательные ядра, занимающие в обойме позицию 9 : “Интересные французские математические журналы”. Здесь полезным является задействование семантических аспектов.</p>		
6	Количествен. числ. (шести)			
7	Притяжательн. прил. (наших)			
8	Порядков. числ. (третьих)			
9	Прилагат. (важных)			
10	Существит. (платьев)			
11	Имя или атрибут – неизм. слово (беж)			

## Приведенные группы.

Все ИсГ, не зависящие непосредственно от вершины дерева зависимостей фразы, соединяются с исходными группами своих управляющих слов (=“хозяев”) и образуют Приведенные Группы (ПГ). Простейшее правило определения порядка ИсГ в ПГ состоит в том, что подчиненные ИсГ ставятся справа от управляющей ИсГ (=“нейтральный” порядок). Исключения возникают в первую очередь в тех случаях, когда управляющая ИсГ (т.е. ее ядро) имеет МУ. В данном случае сначала определяется “нейтральный порядок”, который в дальнейшем может быть изменен по специальным правилам.

Взаимное расположение ПГ определяется в русском языке большим количеством различных по силе факторов, сложно взаимодействующих между собой. Для русского языка можно выделить 4 группы правил расстановки ПГ :

- 1) Расстановка глагола, его актантов и частей сложного сказуемого;
- 2) Расстановка обстоятельств;
- 3) Оформление логического выделения (учет темы и ремы синтезируемого предложения);
- 4) Правила, корректирующие окончательный порядок слов в предложении путем поиска неприемлемых ситуаций.



## **Порядок следования приведенных групп.**

После применения правил первого и второго типа будет получен “нейтральный” порядок слов, т.е. такой порядок, которого придерживался бы человек при построении предложения вне контекста. В первом приближении можно определить следующий порядок приведенных групп :

**ПГ обст. – ПГ по предикативному СИО (по ГСО 1) – ПГ глагола – ПГ по СИО 1К (1-му комплетивному СИО, ему соответствует ГСО 2) – ПГ по СИО 2К (по ГСО 3) – ПГ по СИО 3К (по ГСО 4) – ПГ по обст.**

Правила третьего типа из представленных на предыдущем слайде учитывают влияние контекста на порядок слов. ПГ, входящие в ремю (т.е. то новое, что сообщается впервые), ставятся правее ПГ, образующих тему.

Пример правила четвертого типа : Если с одной стороны от глагола не находится ни одной ПГ, а с другой стороны их несколько, то :

- 1) В левую сторону переносится любая ПГ, кроме той, к которой был задан вопрос;**
- 2) В правую сторону переносится любая ПГ.**

# **Морфологический синтез словоформ ответа.**

**Подобно морфологическому анализу, Морфологический Синтез (МС) может быть реализован декларативным и процедурным способами [4,5,6].**

**При декларативном способе в словаре системы каждой основе сопоставлены различные возможные для нее наборы Морфологической Информации (МИ) и соответствующие им словоформы. Работа декларативного МС состоит в считывании из словаря по основе и набору МИ (полученному на этапе синтаксического синтеза) соответствующей словоформы этой основы.**

**Процедурный МС основан на классификации слов русского языка по типу формирования словоформы из основы (пример - для существительного – 7 типов, см. [5]). При процедурном МС формирование словоформ осуществляется с помощью таблиц окончаний и суффиксов, расклассифицированных по частям речи. Конкретный тип основы определяет правило, по которому осуществляется приписывание окончания основе. В целях учета случаев изменения основ (*приобрести - приобретение*), все возможные формы основы слова заносятся в словарь.**

## Литература (основная).

1. Попов Э.В. Общение с ЭВМ на естественном языке. – М.: Наука, 1982. С. 232-347.
2. Апресян Ю.Д. Избранные труды. В двух томах. Т.1. Лексическая семантика. Синонимические средства языка. - М.: Школа "Языки русской культуры", 1995. С. 79-94, 119-163
3. И.А.Мельчук Опыт теории лингвистических моделей "смысл $\leftrightarrow$ текст" : Семантика, синтаксис / И.А.Мельчук.-[Переизд.]. // Школа "Языки русской культуры". Москва, 1999.
4. Белоногов Г.Г., Новоселов А.П. Автоматизация процессов накопления, поиска и обобщения информации. - М.: Наука, 1979.
5. Белоногов Г.Г. и Богатырев В.И. Автоматизированные информационные системы. Под ред. К.В. Тараканова. - М.: Сов. радио, 1973.
6. Фреймовое представление семантики текста // Лингвистический вестник. Вып. 2. Ижевск : УМО "Sancta Lingua", 2000. С. 114-124, см. также <http://libelli.narod.ru/sonnet66/Russian/maslennikova/frame.html>

## Литература (дополнительная).

7. Минский М. Фреймы для представления знаний. - М.: Энергия, 1979.
8. Чарняк Ю. Умозаключения и знания // Новое в зарубежной лингвистике. Вып. XII. М., 1983. Ч. I. С. 171-207. Ч. II. С. 272-317
9. Шингарева Е.В. О двух направлениях представления семантики текста (тезаурус и фрейм) // НТИ. Сер. 2. Информ. процессы и системы. 1982. № 8. С. 1-8
10. Гончаренко В.В., Шингарева Е.А. Фреймы для распознавания смысла текста/ Отв. Ред. Р.Г. Пиотровский. – Кишинев : Штиинца, 1984
11. Романов А.А. Системный анализ регулятивных средств диалогического общения. – М.: Б.и., 1988.