

Лингвистические информационные ресурсы и их применение для задач компьютерной обработки Естественного Языка.

Лекция 11 (Часть 2).

О некоторых практических аспектах описания Лексического Значения предикатного слова в компьютерном словаре.

Специальность : 230105

Общая схема анализа высказывания.

- Морфологический Анализ (МА) слов входного предложения.
- 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), 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.

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.

Смысловые соотношения между цельными лексическими единицами описываются с помощью аппарата стандартных Лексических Функций (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.

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.**

Информация моделей управления.

Отношение между лексемой и ее глубинным синтаксическим актантом по Модели Управления :

$$\begin{aligned} & t_government_pattern(Lexema, T_surf_real, D_synt, & (1) \\ & t_semantic_contence(Role_name, ActantSemClassList)) \end{aligned}$$

T_surf_real – способ поверхностной реализации актанта лексемы $Lexema$;

D_synt - тип отношения подчинения между лексемой и ее глубинным синтаксическим актантом (в реальных МУ это 1,2,3, либо 4);

$t_semantic_contence$ – семантическая интерпретация глубинного синтаксического актанта лексемы. Включает : $Role_name$ – название роли обозначаемой актантом сущности и список $ActantSemClassList$ возможных семантических классов.

Характеризованный ролевой состав Лексического Значения (ЛЗ) для предикатного слова семантического класса $SemClass$:

$$t_characterized_role_contents(SemClass, CharacterizedRoleList), \quad (2)$$

Здесь $CharacterizedRoleList$ – список описаний семантических интерпретаций актантов, каждое описание задается составным объектом типа $t_semantic_contence$ (1).

Цель и задачи.

Цель. *исследование возможностей использования существующих в языке закономерностей для решения проблемы автоматизации построения Моделей Управления и контроля их корректности.*

Задачи.

- 1) Выделить составляющие (элементы) толкования слова, которые соответствуют описанию отдельного актанта и могут задаваться настройщиком.
- 2) Исследовать зависимости между составляющими толкования и элементами описания семантических валентностей слова.
- 3) Выявить существующие зависимости между отдельными составляющими описания глубинного актанта слова с целью выделения ключевых элементов в соответствии с предложенной И.А. Мельчуком концепцией МУ.
- 4) На основе выделенных ключевых элементов описания актанта слова разработать методику систематизации и контроля корректности информации справочника МУ.
- 5) В соответствии с выявленными зависимостями разработать алгоритмическую составляющую перехода от фрагмента толкования слова к описанию его глубинного синтаксического актанта.
- 6) Разработать структуру базы знаний программного комплекса ведения справочника МУ.

Описание способа поверхностной реализации актанта.

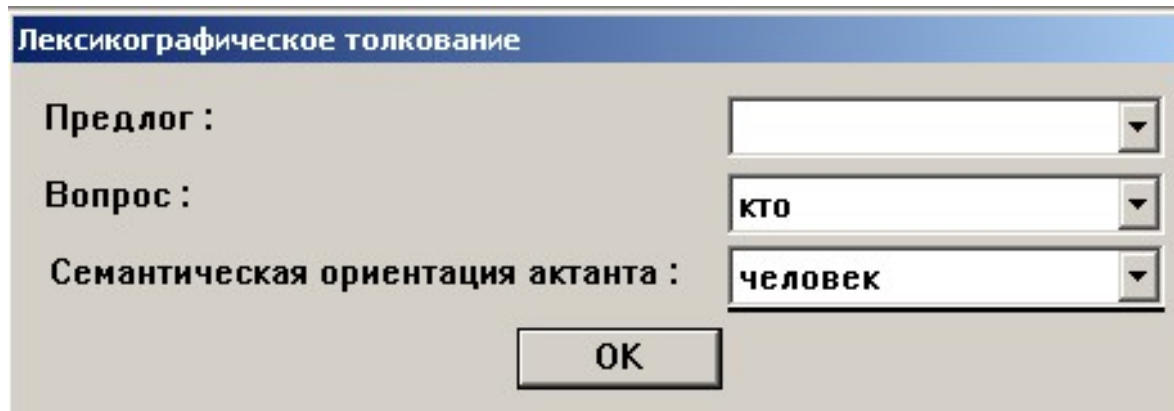
t_surf_real(Prep, Synt_class, Gram_info), где

Synt_class - синтаксический класс слова;

Prep – предлог, который должен присутствовать в поверхностной реализации актанта (поле может быть пустым);

Gram_info – набор грамматических признаков словоформы.

Толкование лексического значения слова по введенным пользователем данным.



Лексикографическое толкование

Предлог :

Вопрос :

Семантическая ориентация актанта :

OK

Рис.1. Исходные данные для описания актанта слова.

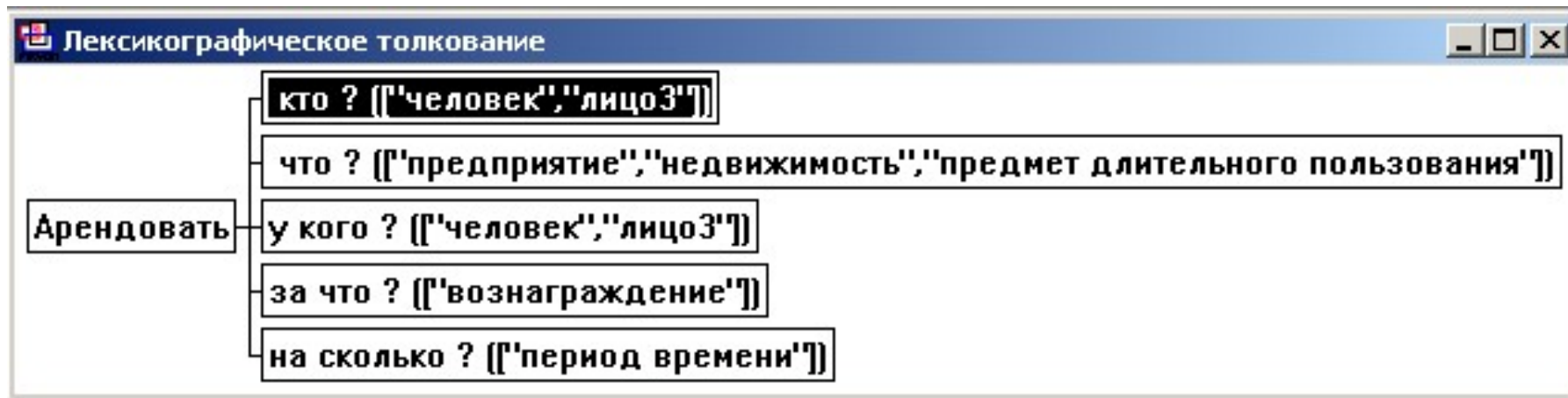


Рис.2. Толкование лексического значения глагола *арендовать*.

Типичная лексика вопросов и морфологические характеристики будущей валентности.

Таблица 1.

Предлог	Вопрос	Синтаксический класс	Грамматическая характеристика
	кто	Существительное	Именительный падеж
у	кого	Существительное	Родительный падеж
из	чего	Существительное	Родительный падеж
	чего	Существительное	Родительный падеж
до	чего	Существительное	Родительный падеж
	кому	Существительное	Дательный падеж
	чему	Существительное	Дательный падеж
на	что	Существительное	Винительный падеж
	что	Существительное	Винительный падеж
в	что	Существительное	Винительный падеж
на	кого	Существительное	Винительный падеж
на	сколько	Существительное	Винительный падеж
	сколько	Существительное	Винительный падеж
	чем	Существительное	Творительный падеж
за	чем	Существительное	Творительный падеж
на	чем	Существительное	Предложный падеж
в	чем	Существительное	Предложный падеж
	какой	Полное прилагательное	Именительный падеж

Установление типа отношения глубинного синтаксиса.

- для валентностей с порядковыми номерами от 1 до 4 номер валентности напрямую отображается в тип отношения глубинного синтаксиса (D_synt);
- валентностям с порядковыми номерами от 5 до 14 соответствует тип отношения 5¹;
- для валентностей с порядковыми номерами больше 14 тип отношения подчинения определяется как “неизвестный” (резерв).

¹ Число неоднородных адъюнктов как зависимых членов этого типа отношения в реальных текстах не превышает 10

Таблица 2. Типовые грамматические формы выражения глубинных актантов.

Тип отношения	Предлог	Синтаксический класс	Грамматическая характеристика
1		Существительное	Именительный падеж

Свойства семантических валентностей.

Свойство 1. Если зависимость с некоторой ролевой ориентацией является семантической валентностью при заданном слове, то она, как правило, выражается идиоматично. Это означает, что морфологическое выражение семантической валентности определяется как содержанием самой валентности (семантическим классом выражаемых актантом слов), так и семантическим классом предикатного слова.

Свойство 2. Семантические валентности слова немногочисленны и могут быть описаны явным образом в словаре.

Свойство 3. Существуют ролевые зависимости, содержание которых всегда являются частью значения соответствующего предикатного слова. Эти ролевые зависимости являются семантическими валентностями при любых предикатных словах.

Свойство 4. Существуют ролевые зависимости, которые являются содержанием семантических валентностей только для слов определенных семантических классов, в остальных случаях – содержанием чисто грамматических зависимостей.

Свойство 5. Вне зависимости от принадлежности к категории “чисто семантических” ролевые зависимости могут как иметь, так и не иметь типовых морфологических форм выражения актантных значений.

Классификация семантических валентностей.

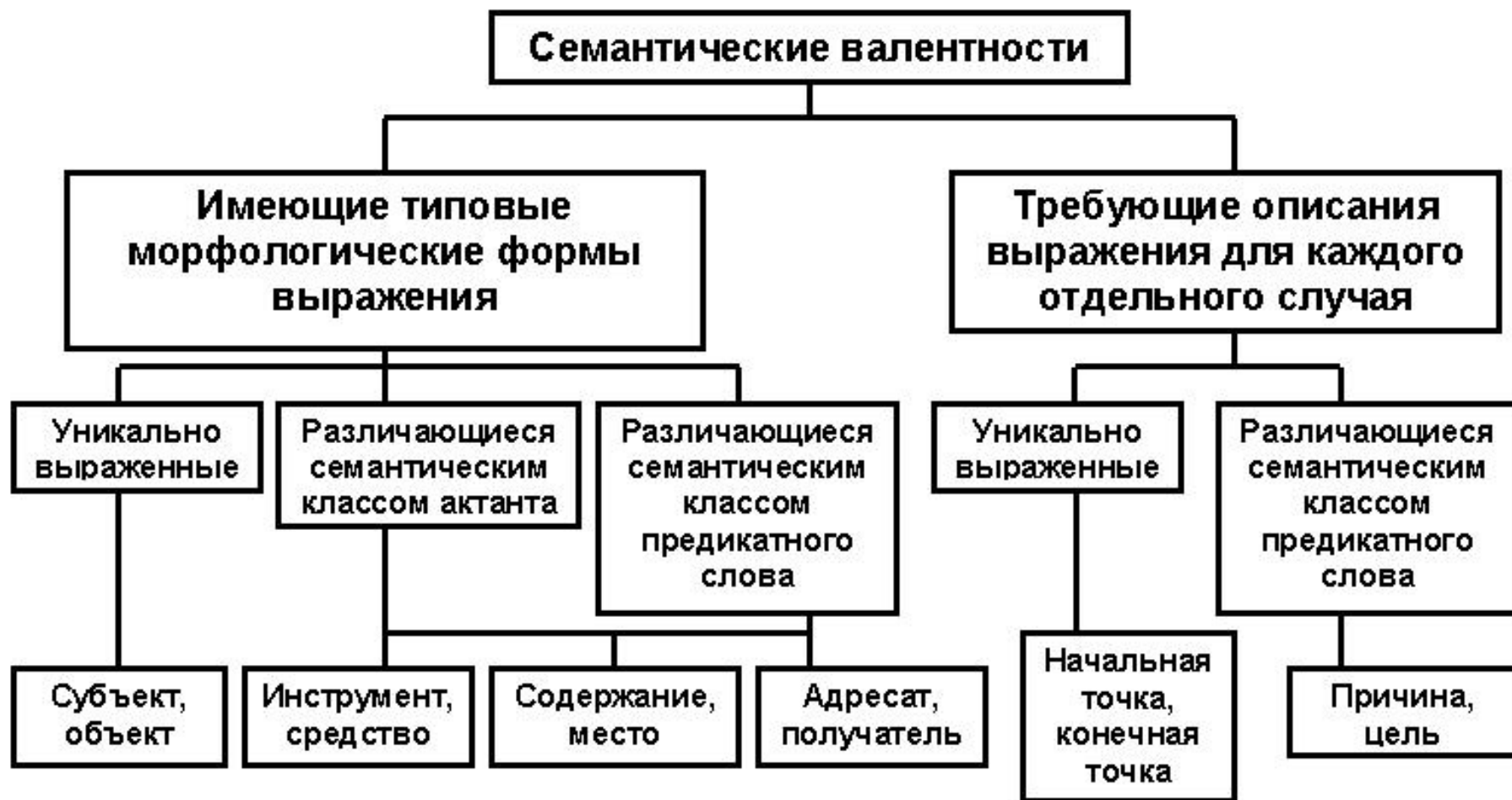


Рис.3.

Типовые формы морфологической реализации семантических валентностей.

Таблица 3.

Ролевая ориентация	Синтаксический класс	Грамматическая характеристика
Субъект	Существительное	Именительный падеж
Объект	Существительное	Винительный падеж
Содержание	Существительное	Предложный падеж
Адресат	Существительное	Дательный падеж
Получатель	Существительное	Дательный падеж
Инструмент	Существительное	Творительный падеж
Средство	Существительное	Творительный падеж
Место	Существительное	Предложный падеж

Таблица 4. Морфологическая реализация семантических валентностей при отсутствии типовых форм.

Рольевая ориентация	Синтаксический класс	Грамматическая характеристика	Предлог	Вопрос	Семантический класс актанта
Цель	Существительное	Винительный падеж	На	Что	Недвижимость
Цель	Существительное	Винительный падеж	На	Что	Предмет длительного пользования
Цель	Существительное	Винительный падеж	На	Что	Культурное мероприятие
Цель	Существительное	Творительный падеж	За	Чем	Физический объект
Причина	Существительное	Родительный падеж		Чего	Событие
Причина	Существительное	Родительный падеж		Чего	Импликатор события
Причина	Существительное	Родительный падеж		Кого	Импликатор события
Причина	Существительное	Дательный падеж		Чему	Событие
Причина	Существительное	Винительный падеж	На	Что	Событие
Цель	Существительное	Винительный падеж	На	Что	Незаконное либо недозволенное действие
Цель	Существительное	Родительный падеж		Чего	Действие
Срок	Существительное	Винительный падеж	На	Сколько	Период времени
Срок	Существительное	Родительный падеж	До	Чего	Событие
Срок	Существительное	Винительный падеж		Сколько	Период времени
Срок	Качественное прилагательное	Именительный падеж		Какой	Период времени
Начальная точка	Существительное	Родительный падеж	Из	Чего	Место
Конечная точка	Существительное	Винительный падеж	В	Что	Место

Использование информации семантических классов.

Таблица 5. Соответствия семантических классов выражаемых актантами слов ролевым ориентациям валентностей.

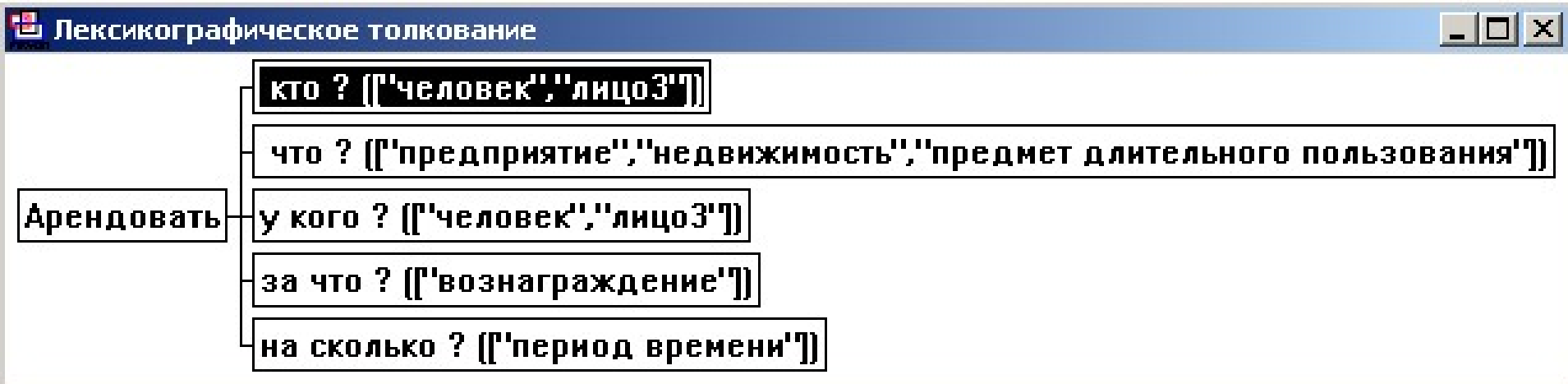
Ролевая ориентация актанта	Семантический класс актантного значения
Средство	Может расходоваться
Инструмент	Инструмент
Начальная точка	Место
Конечная точка	Место
Место	Место

Таблица 6. Соответствия реализаций ролевых зависимостей в качестве семантических валентностей семантическим классам предикатных слов.

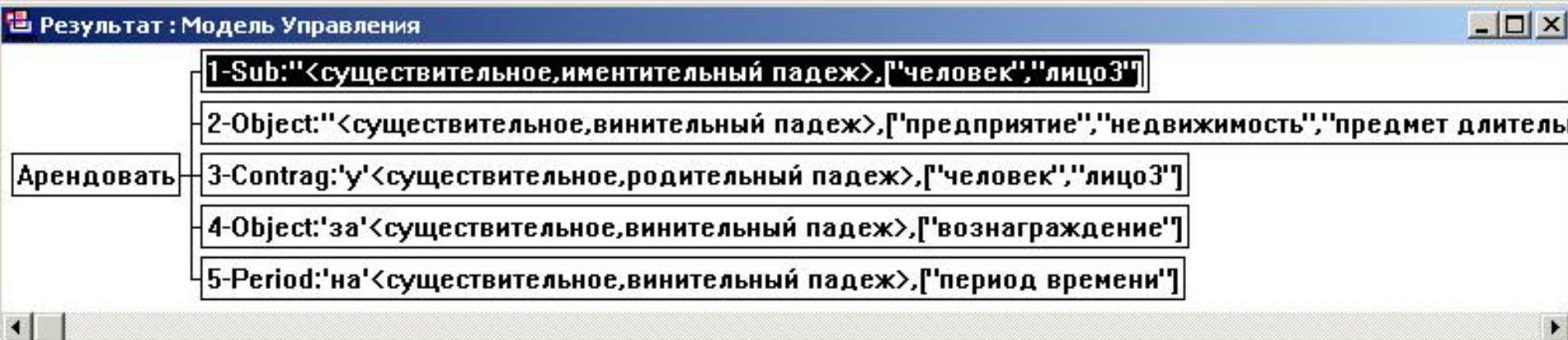
Ролевая ориентация актанта	Семантический класс предикатного слова
Начальная точка	Переставать находиться
Конечная точка	Начинать находиться
Адресат	Информационное воздействие
Причина	Внутреннее эмоциональное состояние человека
Контрагент	Ситуации с противоположностью действий участников
Цель	Очередь как множество ожидающих субъектов
Цель	Очередь как последовательность ситуаций
Срок	Перерыв
Срок	Приобретение права на временное пользование
Срок	Продолжение
Получатель	Передача

Модель Управления для глагола *арендовать* как результат обработки толкования.

Исходные данные : толкование Лексического Значения слова.



Результат : Модель Управления.



Формальный контекст для совокупности лексических значений предикатных слов.

$$K = (G, M, W, I) \quad (3)$$

Здесь :

Множество объектов G есть множество Лексических Значений. значений предикатных слов;

Множество формальных атрибутов M есть множество возможных ролевых ориентаций актантов для обозначаемых предикатными словами ситуаций;

Множество атрибутивных значений W есть множество всех подмножеств множества возможных семантических классов слов, каждое подмножество соответствует множеству семантических классов слов, способных замещать некоторую валентность предикатного слова заданного семантического класса.

Тернарное отношение I между G , M и W , которое задает частичное отображение G на W : $m(g) = w$, ставит в соответствие для семантической валентности ролевой ориентации $m \in M$ предикатного слова семантического класса $g \in G$ множество $w \in W$ семантических классов слов, способных замещать данную валентность.

Отношение “субконцепт-суперконцепт” между ЛЗ предикатных слов.

Определение 1. ЛЗ семантического класса *SemClass1* :

$t_characterized_role_contents(SemClass1, CharacterizedRoleList1)$

следует считать суперконцептом для лексического значения семантического класса *SemClass2* :

$t_characterized_role_contents(SemClass2, CharacterizedRoleList2)$

при условии, что для $\forall RoleName$:

$t_semantic_contents(RoleName, ActantSemClassList2) \in CharacterizedRoleList2$

$\exists t_semantic_contents(RoleName, ActantSemClassList1) \in CharacterizedRoleList1$:

каждому семантическому классу $ActantSemClass1 \in ActantSemClassList1$ можно поставить в соответствие семантический класс

$ActantSemClass2 \in ActantSemClassList2$:

либо $ActantSemClass2 = ActantSemClass1$, либо $ActantSemClass2$ и $ActantSemClass1$ связаны отношением IS_A.

Описание семантического класса совокупностью используемых в РОСС дескрипторов.

$sf(\text{SemClass}, \text{SFLst}, \text{CAT}, \text{CATSubCl})$ (4)

Здесь семантическому классу `SemClass` ставится в соответствие список `SFLst` Семантических Характеристик (СХ) слова в последовательности “более общая СХ – более специфическая СХ”, таксономическая категория `CAT` и ее подкласс `CATSubCl`.

Примеры (для верхней окрестности ЛЗ глагола *арендовать* :

`sf` (приобретение права на временное пользование,
["FINANCIAL", "CAUS", "BELNG", "LEGISL", "S0(эксплуатировать)", "PERIO", "TIME"],
LABL, SIT)

`sf` (займ, ["FINANCIAL", "PERIO", "TIME"], LABL, SIT)

`sf` (купля-продажа, ["FINANCIAL", "CAUS", "BELNG"], LABL, SIT)

`sf` (передача в распоряжение, ["CAUS", "BELNG"], LABL, SIT)

“Субконцепт-суперконцепт” : случай производных валентностей.

Определение 2. ЛЗ семантического класса SemClass1 :

sf(SemClass1,SFLst1,CAT,CATSubCl) считается суперконцептом для ЛЗ

семантического класса SemClass2 : sf(SemClass2,SFLst2,CAT,CATSubCl),

если в дополнение к требованиям Определения 1 при отсутствии у суперконцепта для актанта t_semantic_contents(RoleName1,ActantSemClassList1) актанта субконцепта с идентичной ролевой ориентации и задаваемым Определением 1 соответствием набора возможных семантических классов существует актант

t_semantic_contents(RoleName2,ActantSemClassList2) субконцепта :

\forall ActantSemClass1 \in ActantSemClassList1 : sf(ActantSemClass1,ASFLst1,CAT,CATSubCl)

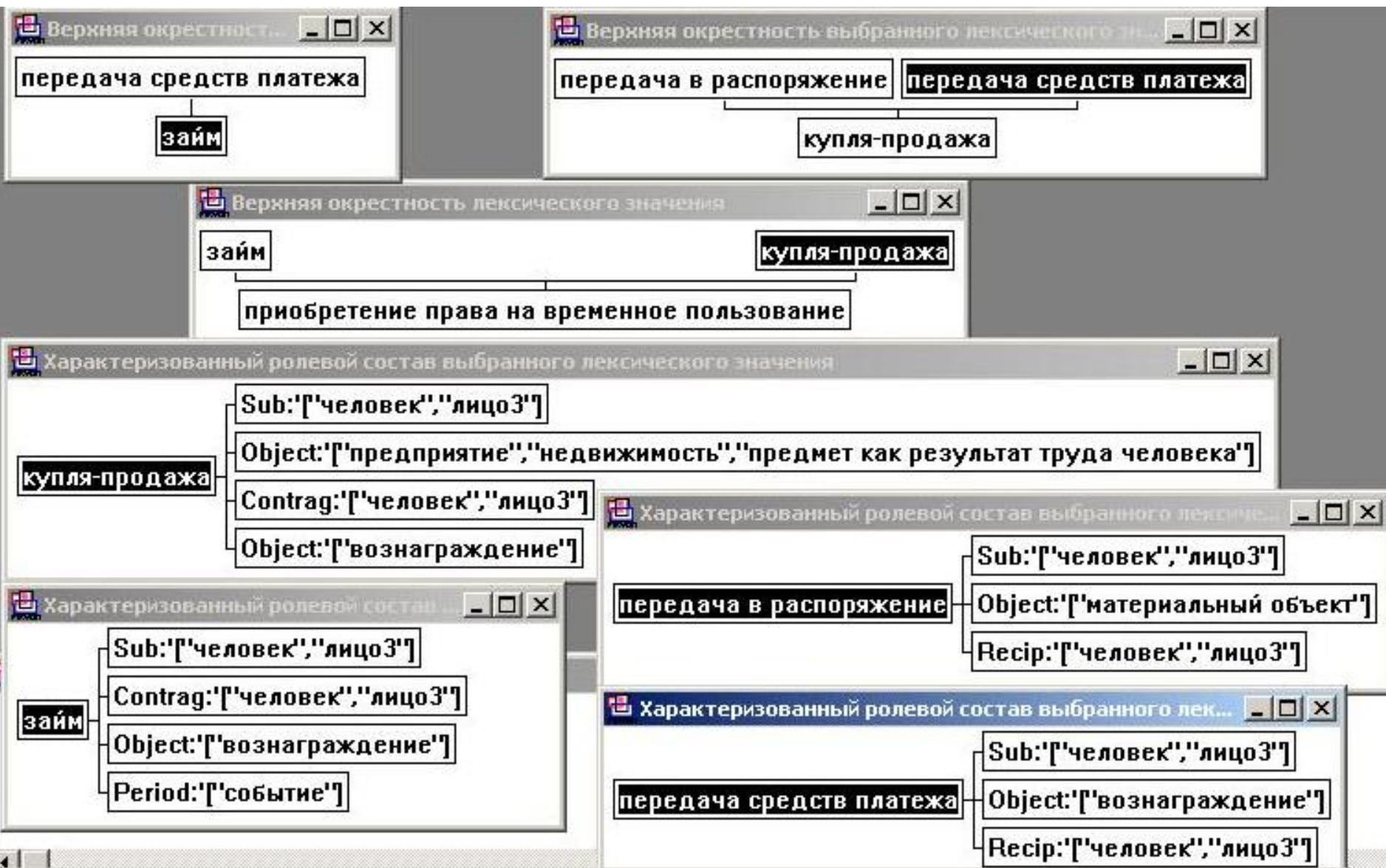
\exists ActantSemClass2 \in ActSClLst2 : sf(ActantSemClass2,ASFLst2,CAT,CATSubCl),

причем один из ASFLst1 и ASFLst2 содержит другой в качестве подписка, а SFLst2 характеризуется тем, что если SFLst1 последовательно разбивается

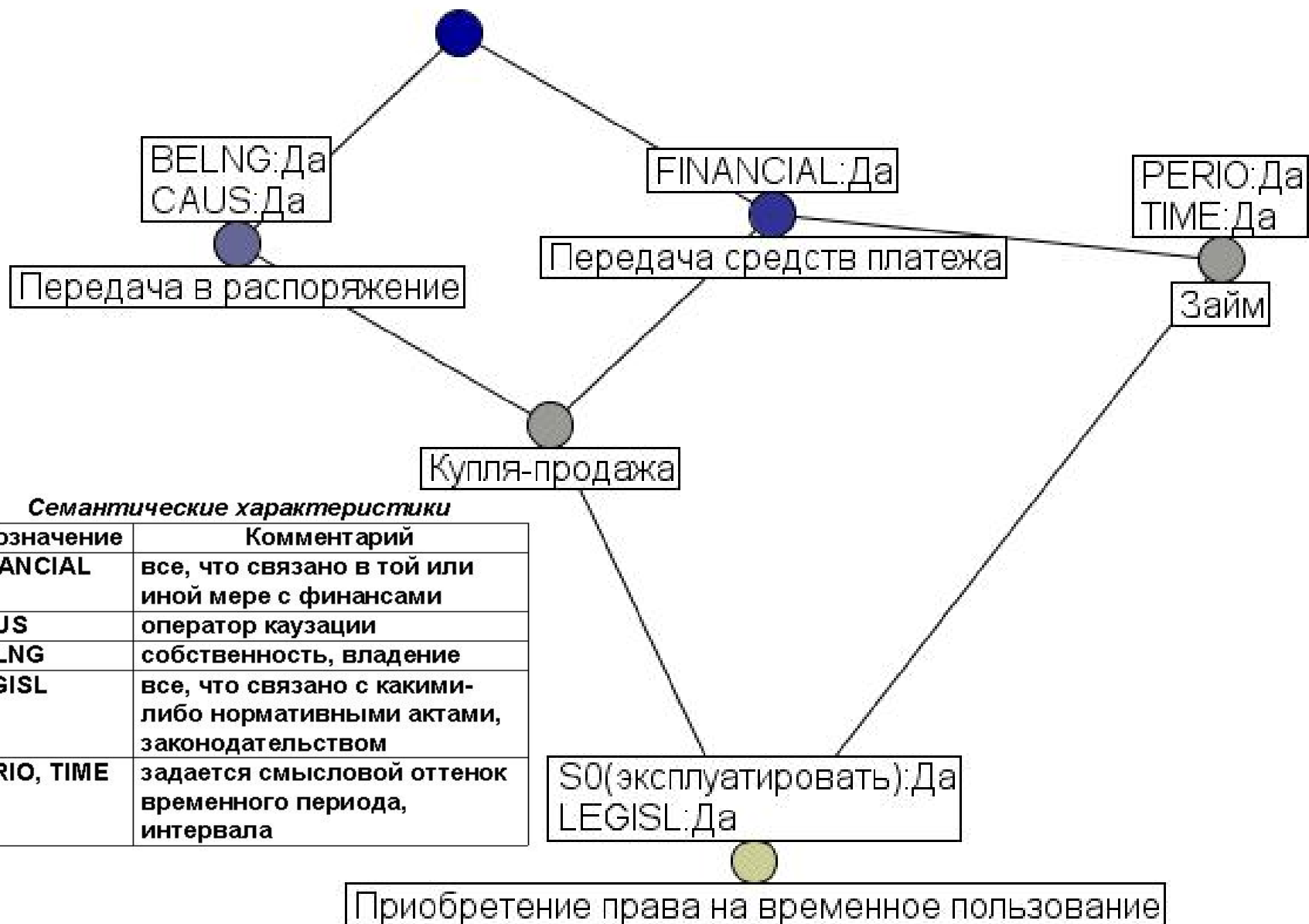
на подписки : $SFLst1_1 \cup SFLst1_2 \cup \dots \cup SFLst1_N$, $\forall i \neq j$ $SFLst1_i \cap SFLst1_j = \emptyset$

то $\forall SFLst1_i, i = 1, \dots, N$ есть общий подписок для SFLst2 и SFLst1 .

Верхняя окрестность ЛЗ глагола “арендовать”.



Описания ЛЗ предикатных слов наборами СХ.



Выводы.

- Методика построения МУ предикатного слова на основе его лексикографического толкования позволяет развивать предметно-ориентированные словари МУ рядом исследовательских коллективов параллельно, при этом сама концепция МУ изначально заложена в систему.
- Описанная методика выявления близких по смыслу лексических значений дает возможность определить место нового слова в существующей лексической системе. Это позволяет избежать противоречий и дублирования информации в базе данных МУ.
- Использование реализующего математические методы Формального Концептуального Анализа специализированного ПО ToscanaJ (<http://toscanaj.sourceforge.net/>) позволяет оценить адекватность набора характеристик семантических валентностей, лежащего в основе распознающих правил, и при необходимости корректировать сами правила.

Перспективы дальнейших исследований.

- Применение предложенного А.С. Нариньяни семантически-ориентированного подхода для анализа фрагментов толкования на Естественном Языке и выделения значимой для построения МУ информации. Здесь необходимо задействовать знания о лексикографическом толковании как о языковом жанре.**
- Использование системы лексических значений предикатных слов для подбора элементов лексикографического толкования, в наибольшей степени соответствующих описанию актантов нового слова. Это позволяет сократить объем перебора при анализе представленных на Естественном Языке фрагментов толкования.**

Рекомендуемая литература.

- 1) Емельянов Г.М., Михайлов Д.В. Построение Модели Управления предикатного слова на основе его лексикографического толкования. // Международное научно-теоретическое издание "Таврический Вестник Информатики и Математики", №1, 2005, с. 35-48.
- 2) D.V.Mikhailov, G.M.Emelyanov Application Of The Predicate Word's Lexical Meanings's System For Automation Of Updating Of The Dictionary Of Government Patterns. // Interactive Systems And Technologies : The Problems of Human-Computer Interaction. - Collections of scientific papers. - Ulyanovsk : ULSTU, 2005. - pp. 164-168.
- 3) Михайлов Д.В., Емельянов Г.М. Применение семантических полей словаря РОСС в задаче построения Модели Управления предикатного слова. // Математические методы распознавания образов : 12-я Всероссийская конференция : Сборник докладов. - М.: МАКС Пресс, 2005, с. 382-385.
- 4) Нариньяни А.С. Автоматическое понимание текста - новая перспектива // Труды международного семинара Диалог-97 по компьютерной лингвистике и ее приложениям, М.: Наука, 1997, с. 203-208.

Рекомендуемая литература (продолжение).

- 5) И.В. Азарова, О.А. Митрофанова, А.А. Синопальникова Компьютерный тезаурус русского языка типа WORDNET // Компьютерная лингвистика и интеллектуальные технологии : Тр. Междунар. конференции Диалог'2003.- М.: Наука, 2003. С. 43-50.**
- 6) Uta E. Priss The Formalization of WordNet by Method of Relational Concept Analysis // In: Fellbaum, Christiane (ed.), WordNet: An Electronic Lexical Database and Some of its Applications, MIT press, 1988, p. 179-196.**