

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

Лекция 10 (Часть 1).

**Моделирование семантических
отношений в лексических базах данных.**

Специальность : 230105

Постановка проблемы.

Используемые для решения задач компьютерной обработки текстов лексические БД (RussNet, Толково-комбинаторный словарь современного русского языка, Русский общесемантический словарь), как правило, заполняются вручную и их сопровождение требует определенных лингвистических навыков. Наиболее перспективный путь решения проблемы систематизации информации лексических БД основан на использовании существующих в них иерархий. Подобные иерархии задаются на основе представленной в БД семантической информации о каждом слове (семантических отношений между словами).

С другой стороны, иерархии, которые могут быть построены на основе семантических отношений, но в разных лексических БД, имеют сходный характер и ставится задача формально описать сходство/различие самих семантических отношений.

Мы рассмотрим использование математических методов Формального Концептуального Анализа (ФКА) и реализующего эти методы программного обеспечения для моделирования семантических отношений в русскоязычных лексических БД.

Формальный Концептуальный Анализ.

ФКА это метод анализа данных, основанный на теории решеток и формализации контекста [1]. Формальный контекст это тройка (G, M, I) , где G - множество формальных объектов, M - множество формальных атрибутов и I - бинарное отношение, определенное между G и M . Деривационный оператор применяется к множеству объектов $X \subseteq G$,

$$X' = \{m \in M \mid \forall g \in X : glm\} \quad (1)$$

или к множеству атрибутов $Y \subseteq M$,

$$Y' = \{g \in G \mid \forall m \in Y : glm\} \quad (2)$$

Формальный концепт, определенный для формального контекста $K=(G, M, I)$, это пара (X, Y) , при условии, что $X \subseteq G$, $Y \subseteq M$, $X = Y'$ и $Y = X'$. Для формального концепта (X, Y) вводится понятие экстенционала X и интенционала Y . Множество всех концептов контекста (G, M, I) определяется как $\beta(G, M, I)$. На $\beta(G, M, I)$ определяется отношение субконцепт-суперконцепт: концепт $(X1, Y1)$ является субконцептом концепта $(X2, Y2)$, если

$$X1 \subseteq X2 \text{ (} Y2 \subseteq Y1 \text{)} \quad (3)$$

Отношение “субконцепт-суперконцепт”.

Пусть пара (X, Y) есть формальный концепт для контекста K .

Множество Y включает в себя в точности те атрибуты из множества M , которые относятся ко всем объектам из множества X .

И наоборот, множество X состоит в точности из тех содержащихся в множестве G объектов, которые имеют все атрибуты из Y .

Определение. Если упорядоченная совокупность (X, Y) является формальным концептом, то X называется экстенсом (экстенсионалом), Y – интенсом (интенсионалом) формального концепта.

Определим бинарное отношение \leq “меньше либо равно” или “является суперконцептом” на множестве $\beta(G, M, I)$ всех концептов контекста K следующим образом.

Если (X_1, Y_1) и (X_2, Y_2) являются концептами контекста K , то концепт (X_1, Y_1) является субконцептом (обозначается : $(X_1, Y_1) \leq (X_2, Y_2)$) для концепта (X_2, Y_2) тогда и только тогда, когда $X_1 \subseteq X_2$ и $Y_2 \subseteq Y_1$. (X_2, Y_2) называется суперконцептом для концепта (X_1, Y_1) .

Данное определение соответствует философскому соглашению о том, что концепт всегда имеет меньший экстенсионал, но больший интенционал, чем любой его суперконцепт.

Концептуальная решетка.

Определение. Множество $\beta(G, M, I)$ всех концептов формального контекста с определенным отношением порядка (\leq) называется концептуальной решеткой.

Концептуальные решетки отображаются графически с помощью линейных диаграмм, на которых формальные концепты обозначены узлами.

При чтении диаграммы экстенционал формального концепта состоит из всех формальных объектов, полученных начиная с формального концепта и далее из всех объектов, которые обозначены как субконцепты этого концепта. Аналогично, интенционал концепта состоит из всех формальных атрибутов, которые обозначены как суперконцепт данного формального концепта. Зависимости между атрибутами могут быть описаны с помощью импликации. Для $X, Y \subseteq M$, импликация $X \rightarrow Y$ содержится в контексте, если каждый объект, обладающий всеми атрибутами из X , также обладает всеми атрибутами из Y .

Модель системы ролевых ориентаций семантических валентностей.

Для решения задачи формального описания и визуализации сходств и различий семантических валентностей применением реализующего методы ФКА специализированного специализированного ПО ToscanaJ (<http://toscanaj.sourceforge.net/>) система ролевых ориентаций семантических валентностей может быть смоделирована в виде многозначного контекста :

$$K = (G, M, W, I) \quad (4)$$

в котором множеству объектов G соответствует множество названий ролей обозначаемых актантами сущностей, множеству M формальных атрибутов - множество названий характеристик, используемых при идентификации валентностей, а именно : семантический класс предикатного слова, семантический класс и морфологические характеристики актанта. Множество W есть множество значений указанных характеристик. Отношение I между G , M и W задает частичное отображение G на W : $m(g)=w$ и ставит в соответствие названию характеристики валентности заданной ролевой ориентации значение этой характеристики.

Свойства семантических валентностей.

Сформулируем свойства семантических валентностей слова, которые могут составить основу их классификации и распознавания.

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 актантов 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.

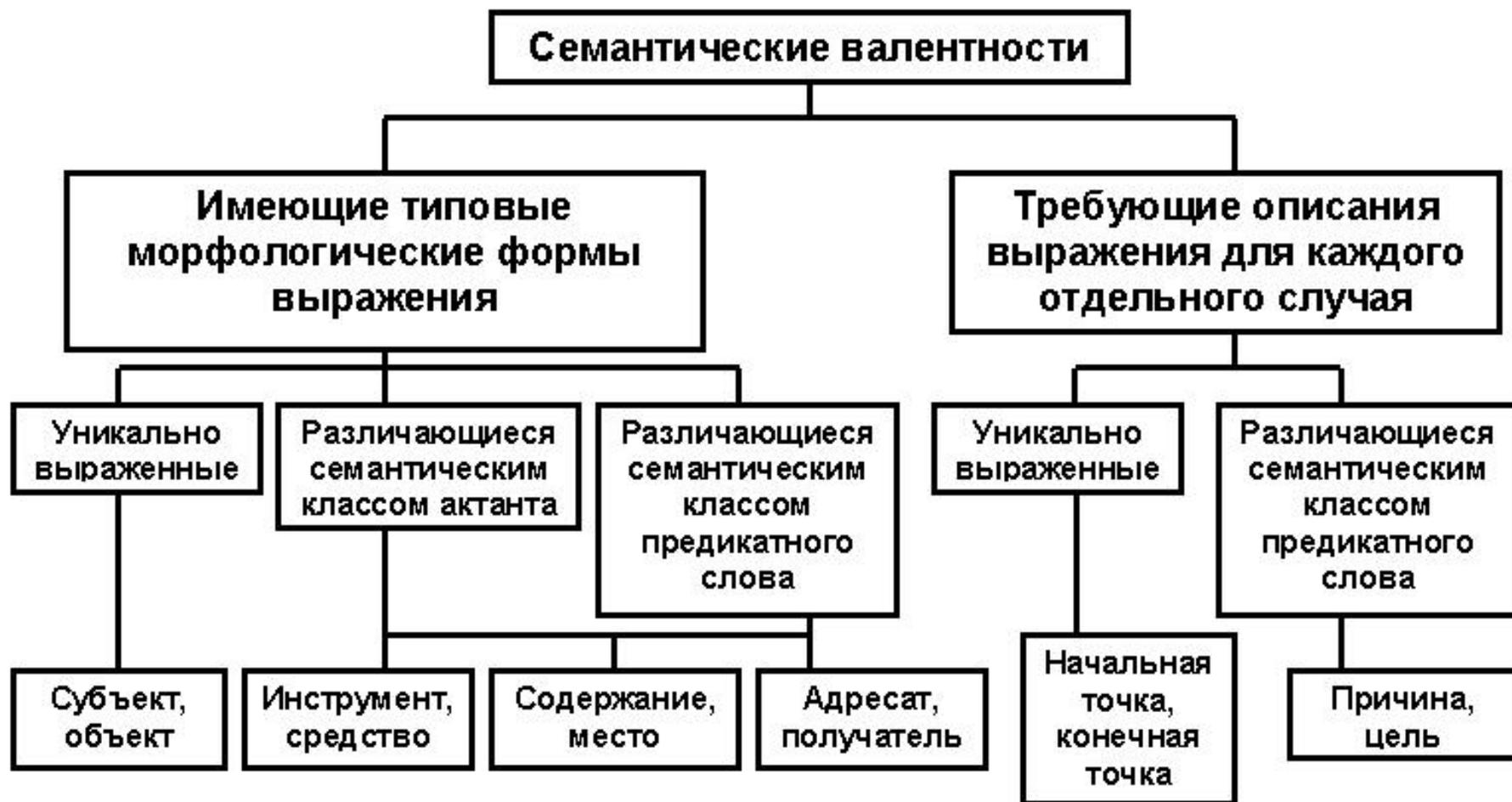


Рис.3.

Формирование множеств атрибутов и их значений : типовые формы морфологической реализации семантических валентностей.

Таблица 1.

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 2. 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 3. 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 4. 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 системы Семантических Классов.

Основную трудность при применении реализованных в **ToscanoJ** классических методов ФКА к исследуемой системе ролевых ориентаций представляет учет связей (прежде всего – родовидовых) между семантическими классами предикатных слов как значениями атрибутов и их визуализация непосредственно при построении контекста (4). В программе **ToscanoJ** отношение “надкласс-подкласс” может быть задано вручную выбором из списка названий семантических классов, причем для каждого класса должны быть отмечены как непосредственные подклассы, так и транзитивно связанные с ним. Для визуализации указанного отношения может быть построена следующая модель системы семантических классов в виде формального контекста :

$$K_{SEM} = (G_{SEM}, M_{SEM}, I_{SEM}), \quad (5)$$

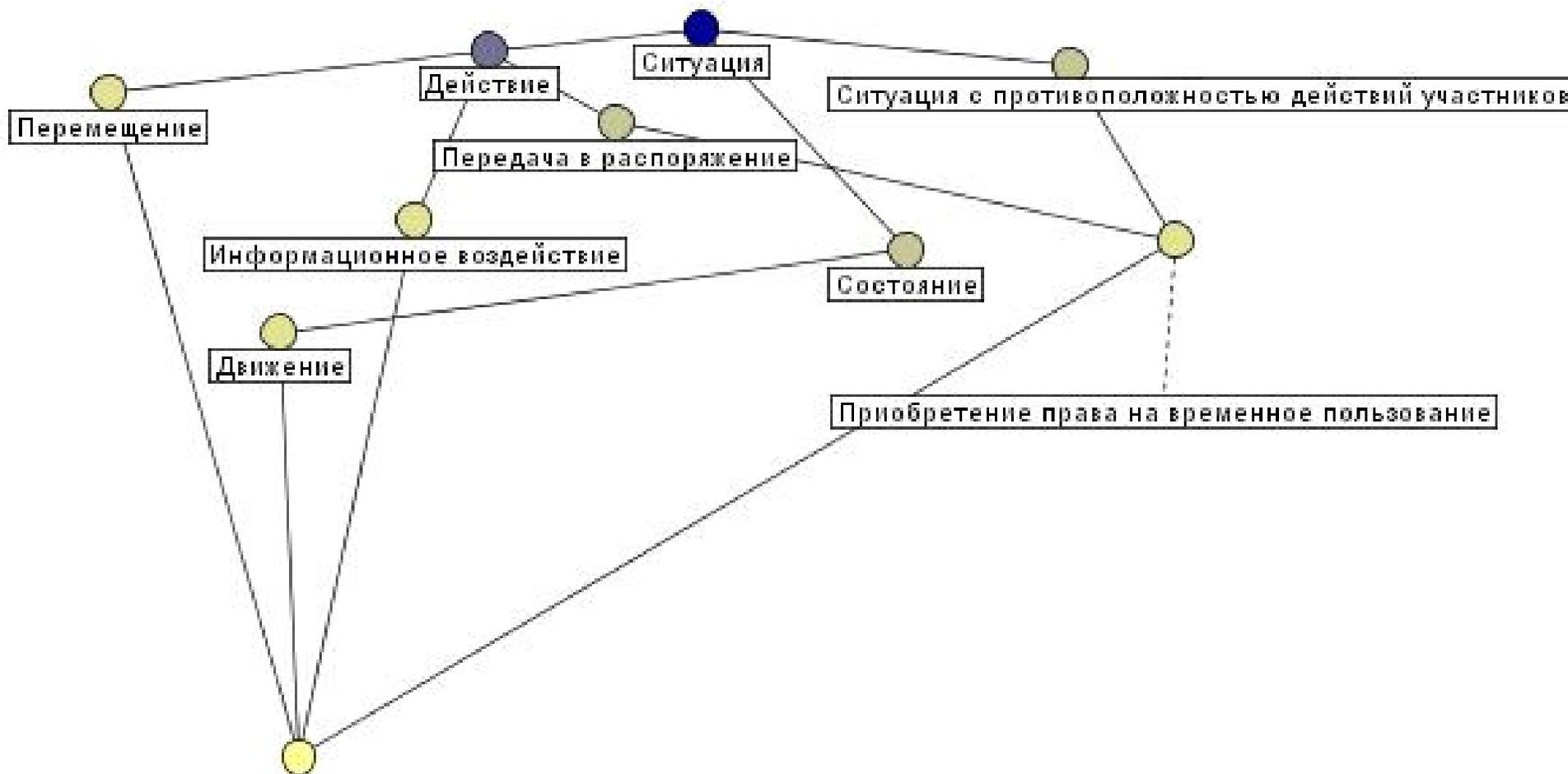
где $G_{SEM} = M_{SEM} = S_CL$, $S_CL \subset W$, см. (4), соответствует множеству

Семантических Классов (в нашем примере – Семантических Классов

предикатных слов, бинарное отношение I_{SEM} ставит в соответствие каждому

семантическому классу $g_{SEM} \in G_{SEM}$ множество его надклассов из M_{SEM} .

Визуальное представление системы семантических классов предикатных слов.



Заключение.

Визуальное представление контекста программой ToscanaJ дает возможность оценить адекватность выбранного набора характеристик семантических валентностей. Так, в ряде случаев валентность контрагента будет ошибочно идентифицироваться как одно из отличных от него отношений валентной структуры “ситуации с противоположностью действий участников” (“приобретение права на временное пользование”, “займ” и т.п.). Причина кроется в отсутствии типовых морфологических форм выражения и типовой семантической ориентации (семантического класса) актанта для этой валентности.

На основе рассмотренной модели для решения указанной проблемы множество M (4) следует расширить введением характеристики “Типовое название участника обозначаемой словом ситуации”, что позволяет также идентифицировать случаи наличия у симметричных предикатов второго субъекта.

Литература.

- 1. B. Ganter and R. Wille, Formal Concept Analysis - Mathematical Foundations. Berlin: Springer-Verlag, 1999.**
- 2. Апресян Ю.Д. Избранные труды. Т.1. Лексическая семантика. Синонимические средства языка. // М.: Школа “Языки русской культуры”, издат. фирма “Восточная литература”, 1995**