

## Кластеризация семантических знаний в задаче распознавания ситуаций смысловой эквивалентности

*Михайлов Д. В., Емельянов Г. М.*

mdv@novsu.ac.ru

Великий Новгород, ГОУ ВПО НовГУ им. Ярослава Мудрого

Центральной задачей анализа смысла высказывания Естественного Языка (ЕЯ) является выделение класса Семантической Эквивалентности (СЭ). Наиболее известная система классов СЭ в ЕЯ определяется множеством  $\Pi^R$  правил синонимических преобразований ЕЯ-высказываний в рамках стандартных Лексических Функций (ЛФ) [1]. При этом для каждого правила  $\pi \in \Pi^R$  задается условие  $r(\pi)$  его применимости. Условие  $r(\pi)$  есть совокупность требований к синтаксическим и семантическим свойствам лексических единиц исходного ЕЯ-высказывания, заменяемых посредством  $\pi$ . В содержательном плане такие требования определяются Смысловыми Отношениями (СО) между некоторым ЕЯ-словом, относительно которого задается Ситуация СЭ (ССЭ), и его лексико-семантическими производными (лексическими коррелятами), которые входят в заменяемый комплекс лексических единиц. В лексической семантике такие СО описываются стандартными ЛФ. Ставится задача: на основе признаков слов пары  $T$  сравниваемых ЕЯ-высказываний  $T_1$  и  $T_2$  отнести  $T$  к одному из известных классов  $\pi \in \Pi^R$ , либо образовать с помощью  $T$  новый класс. Данная задача есть классическая задача распознавания образов. В настоящей работе мы рассмотрим метод, основанный на теории Анализа Формальных Понятий (АФП) [1, 2], для автоматизации формирования  $r(\pi)$  как прецедента класса  $\pi$ .

Наибольший интерес для описания  $\{r(\pi)\}$  представляют ситуации с ЛФ-параметрами. Посредством этих ЛФ описываются Расщепленные Значения (РЗ). Фактически РЗ есть конструкция, непосредственно задающая СО в рамках  $r(\pi)$ . Это позволяет поставить задачу его выявления и обобщения по аналогии с описанием семантики Именных Групп на основе формализованного представления толкований Лексических Значений (ЛЗ) слов в виде теорий [1]. Сказанное подтверждается наработками по Русскому общесемантическому словарю (РОСС): ЛФ используются в качестве Семантических Характеристик отдельных слов в РОСС [2]. Следовательно, такие слова могут быть и названиями отношений в утверждениях теорий других слов. Пример — значение Лексической Функции  $Oper_1$  для ЛЗ «эксперимент» (т. е. «осуществлять»). Как видно из Рис. 1, оно присутствует в одном из утверждений теории ЛЗ «экспериментировать». Данное ЛЗ эквивалентно РЗ «осуществлять эксперимент», где значением ЛФ  $Oper_1$  задается СО типа «операция с...» между 1-м участником ССЭ (кто осуществляет эксперимент) и её названием.

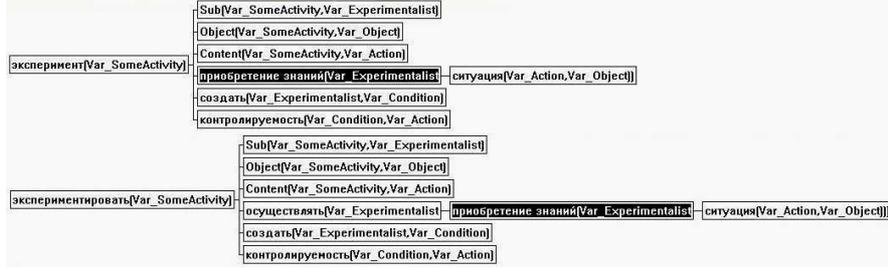


Рис. 1. Теории ЛЗ «эксперимент» и «экспериментировать».

Пусть для слов  $w_1 \in T_1$  и  $w_2 \in T_2$  мы имеем описания теорий ЛЗ. При этом утверждение (постулат значения [1]) каждой теории либо задает бинарное отношение  $R_2$  между понятиями  $C_1$  и  $C_2$  посредством тройки:

$$M_p = (R_2, C_1, C_2), \quad (1)$$

либо представляет собой описание отношения  $R_n$  произвольной арности:

$$M'_p = (R_n, C, L^M). \quad (2)$$

В этом случае постулат значения определяется рекурсивно на основе списка  $L^M$  утверждений вида (1) и (2). Посредством  $L^M$  задается связь понятия  $C$  с другими словами и понятиями. При этом смысл слова  $w_i \in T_1 \cup T_2$  определяется множеством функций [1], которые задаются утверждениями теории ЛЗ  $w_i$  и составляют для него набор признаков.

**Утверждение 1.** *Смысловое отношение  $F$ , значимое для формирования  $r(\pi)$ , между некоторым словом  $w_1 \in T_1$  и его лексическим коррелятом  $w_2 \in T_2$ , входящим в РЗ, будет иметь место тогда, когда*

$$\begin{aligned} L_1^M &= L_{11}^M \cup \{(F, C, L_{22}^M)\} \cup L_{12}^M; \\ L_2^M &= L_{11}^M \cup L_{22}^M \cup L_{12}^M; \\ L_{11}^M \cap L_{22}^M &= \emptyset, \quad L_{11}^M \cap L_{12}^M = \emptyset, \quad L_{12}^M \cap L_{22}^M = \emptyset; \end{aligned}$$

где  $L_1^M$  — набор утверждений теории ЛЗ для  $w_1$ , а  $L_2^M$  — для  $w_2$ .

Правильность выявления  $F$  зависит от корректности ЕЯ-толкований  $w_1$  и  $w_2$ . Рассмотрим решение этой задачи методами АФП.

Лексическими Функциями описывается в первую очередь лексическая сочетаемость. Последняя определяется Лексическим Значением

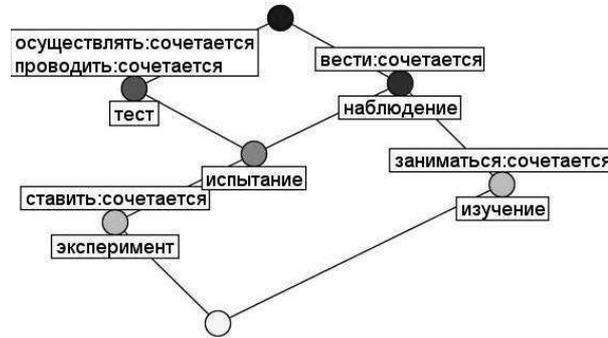


Рис. 2. Слова-аргументы Лексической Функции  $Oper_1$  из верхней окрестности для ЛЗ «эксперимент».

ключевого слова ЛФ-синонимической замены. Более узкое ЛЗ (в терминологии АФП — гипоним) включает более широкие ЛЗ (гиперонимы), которые упоминаются в толковании рассматриваемого ЛЗ, а, следовательно, и в его теории. Таким образом, слово-гипоним в большинстве случаев будет иметь в качестве значений ЛФ-параметра значения этой же ЛФ тех слов-гиперонимов, которые упоминаются в его толковании (теории). Для предикатных слов отношение гипонимии может быть выявлено анализом содержания их семантических валентностей [2].

Применением специализированного ПО ToscanaJ (<http://toscanaj.sourceforge.net>) для заданной ЛФ строится модель системы слов — ее аргументов, Рис. 2.

При этом слова-аргументы ЛФ выступают в качестве объектов, а её значения — в качестве атрибутов. На основе полученной модели критерий адекватности  $r(\pi)$  формулируется следующим образом.

**Утверждение 2.** Теории ЛЗ  $w_1$  и  $w_2$  адекватно задают  $r(\pi)$  при условии существования отношения  $F$ , если  $F$  принадлежит множеству формальных атрибутов того ЛЗ, которое является наименьшей верхней гранью (супремумом) множества слов верхней окрестности [2] ЛЗ  $w_2$ .

Требования к РЗ с ЛЗ-супремумом определяются аналогично.

Работа поддержана УНИК НовГУ и РФФИ, проект № 06-01-00028.

### Литература

- [1] Михайлов Д. В., Емельянов Г. М. Модель сортовой системы языка в задаче построения семантического образа высказывания на уровне глубокого синтаксиса // Таврический Вестник Информатики и Математики. — Симф., 2006. — № 1. — С. 79–90.

- [2] Михайлов Д. В., Емельянов Г. М. Применение семантических полей словаря РОСС в задаче построения Модели Управления предикатного слова // Всеросс. конф. ММРО-12. — Москва: Макс Пресс, 2005. — С. 382–385.