

## О ГРАММАТИКЕ КОНЦЕПТУАЛЬНЫХ ОТНОШЕНИЙ

### ON A POSSIBLE GRAMMAR OF CONCEPTUAL RELATIONS

*Н. Н. Леонтьева (nnl@srcc.msu.ru)*

*Научно-исследовательский вычислительный центр МГУ им. Ломоносова, Москва*

Основу Грамматики концептуальных отношений, на наш взгляд, нужно искать в обобщении и правилах преобразований тех семантических отношений, которые используются в лингвистических работах.<sup>1</sup>

В рамках информационно-лингвистической модели автоматического понимания текста семантические единицы, полученные лингвистическим анализом, отличены от концептуальных единиц, которые мы можем видеть в разных таблицах баз данных (БД). Эти последние отличаются большим разнообразием для разных предметных областей (ПО), особенно если взяты узкие ПО. Узлы и отношения, построенные лингвистическим путем, обозначим СемУ и СемО, а единицы, используемые при описании разных ПО, назовем «специальными» и обозначим СпецУ и СпецО, соответственно. СпецОтношениями могут быть объявлены имена полей БД или заголовки столбцов таблиц и т.п., например, **Размер пласта**, **Объем добычи за год**, **Количество жертв**, **Средний размер зарплаты рабочего** и многие другие. Еще более разнообразны и специфичны имена объектов (СпецУзлов), заполняющих объявленные слоты таблиц БД (*5 тыс. руб./месяц, 80 тыс. дол./год, 100 тонн в сутки* и др.). СпецУ и СпецО в общем случае не совпадают с текстовыми единицами СемУ и СемО.

Прежде всего это различия в объеме понятий и их ранге, положении в той и другой структуре. Например, свободному словосочетанию текста соответствует СпецО, имя графы БД-таблицы (*количество жертв*, или *общее число работающих пенсионеров*) или значение графы (*3 тыс. руб.*); простому узлу СемП соответствует важный онтологический предикат (*угон, убийство, теракт*); свободное текстовое словосочетание (*в сутки перекачивается 10 тыс. тонн нефти*) распределится между именем отношения (*суточная норма*), значением параметра (*10 тыс. тонн*) и названием всей таблицы (*нефть*) и т.д.

В концептуальных структурах некоторых ИЕ-систем не различается статус узлов и отношений, например, для ситуации *пропажа автомобиля* (онтологический предикат), наряду с такими вполне лингвистическими отношениями как **Исходная\_точка** (откуда пропал), заданы спецотношения Угонщик (в лингвистическом СемП это **Агент** угона) и Место\_обнаружения (лингвистически **Конечная\_точка** - куда угнали автомобиль). Эти примеры демонстрируют произвольность задания спец-единиц и их формальную сложность по сравнению с лингвистическими узлами и отношениями.

Такие несоответствия в форме выражения и структуре двух типов структур не позволяют проводить прямое сравнение и взаимообмен между СемП текста и структурой знаний. Для привлечения сложных форм задания знаний к взаимодействию с текстом недостаточно простых приемов анализа и заведомо недостаточно только синтаксического уровня. Это означает, что среди процессоров системы понимания текста должен быть некий **спецтранслятор** - компонент, который отвечает за **спецпонимание**, т.е. приводит разнообразные источники знаний к виду, близкому или сравнимому с текстовыми семантическими структурами.

Д.А.Поспелов (1981:41) считает, что "Проблема установления совместимости двух записей не менее трудна, чем проблема построения СемП. По существу, она сводится к построению СемП второго уровня". Хотя автор относит эту мысль к сравнению СемП разных предложений текста, она тем более справедлива по отношению к сравнению СемП текста и встречного текста в виде структуры знаний. Однако мы видим возможное решение проблемы не в повышении, а в **понижении** уровня сравниваемых структур. Чтобы обеспечить сравнимость, нужно перевести естественные записи в базах знаний на тот язык, который принят в СемП текста (построив некое «встречное» СемП знаний), и задать правила синтеза элементарных СемО в сложные единицы, специфические для данной ПО. Так, если в ПО «Родства» часто встречаются СемО **Равно(А,В)** и **По отношению к (В,С)**, можно задать правило образования спец-отношения **В(А,С)**, например, *Иван=брат* и *Брат Петра* – образуют спец-предикат **БРАТ(Иван,Петр)**.

<sup>1</sup> Работа ведётся при поддержке гранта РГНФ № 01-04-00185а "Разработка грамматики концептуальных отношений для автоматического извлечения знаний из текстов" (2004-2006 гг.).