

Шаронова Н.В., Канищева О.В.
Национальный технический университет
«Харьковский политехнический институт»

olya-kanisheva@rambler.ru, nvsharonova@mail.ru

ПЕРСПЕКТИВЫ ИСПОЛЬЗОВАНИЯ АЛГЕБРЫ КОНЕЧНЫХ ПРЕДИКАТОВ ПРИ ОПИСАНИИ СЕМАНТИЧЕСКОЙ СОЧЕТАЕМОСТИ

Введение. Актуальность работы. Автоматический семантический анализ является одной из актуальных и вместе с тем наиболее сложной задачей компьютерной лингвистики. Он является необходимым этапом создания любой системы, моделирующей человеческий интеллект. Исследования по семантике велись достаточно давно, однако каких-либо установившихся методов и приемов не было найдено. Можно утверждать, что имеющиеся решения практически пригодны только для ограниченных областей [1].

Очевидно, что синтаксический и семантический анализ непосредственно связаны и ни одна задача не может быть решена без помощи другой. Основной лингвистической единицей, вокруг которой разрабатывается тот или иной прием определения смысла или содержания, является слово. Слова в предложении связаны друг с другом грамматически и по смыслу. Грамматические связи обеспечивают грамматическую правильность речи. Примеры грамматических связей: согласование прилагательного-определения с определяемым существительным в роде, числе и падеже, глагольное управление существительными. Задача грамматической сочетаемости, благодаря морфологическому анализу и специфическим маркерам (например, изменение окончания), является практически идентифицируемой, особенно для часто используемых конструкций словосочетаний.

Смысловые связи обеспечивают правильность высказывания по смыслу, и смысловые отношения слов редко выражаются при их написании. Тем не менее, употребляя слово, мы должны согласовывать его по смыслу с другими словами. Это смысловое согласование выражается в двух типах словесной сочетаемости – семантической и лексической. Семантическая сочетаемость слова – это его способность вступать в сочетания с целыми классами слов, объединяемых общностью смысла. Например, глаголы думать, полагать, радоваться, смеяться, грустить и другие описывают различные состояния человека; значит, и сочетаться они могут лишь с такими словами, которые обозначают человека (это и есть один из семантических классов): мальчик, старик, прохожий, врач, учительница и т.п.

Лексическая сочетаемость слова – это его способность вступать в сочетания не с любым словом из какого-либо семантического класса, а только с некоторыми. Например, существует класс слов, объединяемых общим смыслом ‘множество, совокупность’: стадо, табун, стая, рой, косяк и т.п. При необходимости обозначить множество каких-нибудь животных мы не можем сочетать название любого животного с любым из этих слов. Говорят: стадо коров, табун лошадей, стая птиц, рой пчел, косяк рыбы (но не «стая рыбы или рыб», «стадо пчел», «рой лошадей» и т.п.).

Целью данной работы является исследование семантической сочетаемости слов и использование аппарата алгебры конечных предикатов, разработанной школой Ю.П. Шабанова-Кушнаренко [2] для моделирования семантических отношений в словосочетаниях.

Основной материал. Между словами в словосочетаниях существуют различные типы отношений (рис. 1). Отношения иерархии как разновидность сильных парадигматических отношений соответствуют отношениям подчинения или отношениям типа «выше–ниже». В пределах иерархических отношений различают отношения типа «род–вид» и «целое–часть» [3]. Отношение «род–вид» является одним из важнейших видов связей между понятиями. При этом родовым (подчиняющим) называется понятие, выражающее существенные признаки класса предметов, являющихся видами этого рода. Соответственно видовым (подчиненным) называется понятие, которое отображает существенные признаки класса предметов, являющегося видом какого-либо рода.

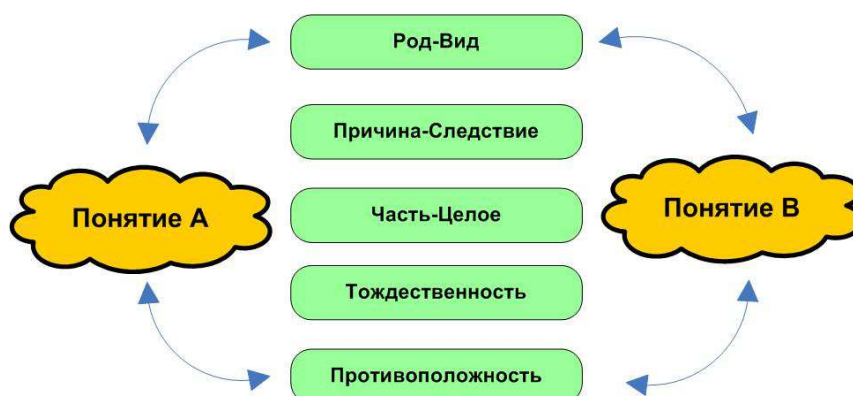


Рис. 1 Виды логических связей между понятиями

В качестве математического аппарата для моделирования семантики свободных словосочетаний предлагается использовать алгебру конечных предикатов (АКП) – универсальный математический аппарат для описания детерминированных, дискретных и конечных объектов, он позволяет описывать различные языковые закономерности в рамках единой математической модели. АКП является обобщением аппарата булевых функций и многозначной логики, дает возможность перехода от алгоритмического описания лингвистических закономерностей к описанию их в виде уравнений. Словоформы реализуют свои сочетания в контексте, и в нашем случае это словосочетание в определенной предметной области. Значение словоформы определяется лишь теми признаками, которые характеризуют ее связь со значениями слов в контексте. Семантическая сочетаемость словоформ зависит от знаний предметной области и от родо-видовой структуры [4, 5].

Родо-видовое отношение в русском языке – это бинарное отношение R , заданное на множестве словоформ D такое, что для любого $d \in D$, всегда существует $d^1 R d$ и, кроме того, R не рефлексивно ни для одного $d \in D$, антисимметрично и транзитивно. Отношение R называют еще отношением древовидного порядка. Возможен переход от двухместного родо-видового отношения к n -местному. Каждому n -местному отношению R , заданному на n -той декартовой степени D^n множества словоформ D , можно поставить в соответствие каноническое уравнение вида: $f_R(x_1, x_2, \dots, x_n) = 1$, которое связывает переменные x_1, x_2, \dots, x_n точно так же, как и отношение R . Предикат f_R выбирается следующим образом:

$$f_R(x_1, x_2, \dots, x_n) = \begin{cases} 1, & \text{если } (x_1, x_2, \dots, x_n) \in R \\ 0, & \text{если } (x_1, x_2, \dots, x_n) \notin R, \end{cases} \quad (1)$$

где $(1, 2, \dots, n)$ – количество уровней иерархии. Записав предикат f_R в виде формулы АКП, получаем аналитическую запись отношения R в форме уравнения (1). Слова, связанные родо-видовыми отношениями, представляют собой лексико-семантические классы с иерархической структурой, описываемой в виде родо-видовых деревьев [4, 5].

Литература:

1. Марчук Ю. Н. Компьютерная лингвистика / Ю. Н. Марчук. – М. : АСТ Восток – Запад, 2007. – 317 с.
2. Бондаренко М. Ф. Теория интеллекта / М. Ф. Бондаренко, Ю. П. Шабанов-Кушнаренко : учебник – Х. : ООО «Компания СМИТ», 2006. – 576 с.
3. Касаткин Л.Л. Русский язык / Л. Л. Касаткин, Е. В. Клобуков, Л. П. Крысин: учебник – М. : Издательский центр "Академия", 2001. – 768 с.
4. Замаруева И.В. Математические модели семантики свободных словосочетаний с родо-видовыми компонентами и их применение в АИС. – Дис. ... кан. техн. нак. – Харьков: ХТУРЭ, 1990. – 170 с.
5. Булкин В.И., Шаронова Н.В. Математические модели знаний и их реализация с помощью алгебропредикатных структур. – Монография. – Харьков, НТУ «ХПИ», Донецк, МЭГИ, 2010. – 304 с.