

## **Формализация словообразовательного синтеза на основе семантических свойств формантов**

*Вороной С.М., к.т.н., доц., Егошина А.А., Донецкий государственный институт искусственного интеллекта, г. Донецк  
postmaster@iai.donetsk.ua*

Работа посвящена решению задачи словообразовательного синтеза на основе применения семантических свойств формантов. Проведен анализ соотношения компонентов производных слов и значения словообразовательного форманта. Предложена формальная модель словообразовательного синтеза, учитывающая заданные грамматические и лексические значения производного слова.

### **Введение**

Неотъемлемой частью современных интеллектуальных информационно-поисковых и обучающих систем с естественно-языковым интерфейсом является модуль словообразования. Описание словообразования является важной составной частью описания языковой системы в целом. Словообразование – одно из основных средств пополнения словарного состава языка новыми словами. Комплексные единицы словообразования как системно организованные группы становятся объектом внимания лингвистов приблизительно с середины 70-х годов XX века. Их изучение тесно связано с вопросами систематизации производной лексики и с проблемой словообразовательного гнезда и образующих его элементов [1].

Иногда в современной лингвистике словообразование рассматривают как часть лексикологии. Однако более характерен для лингвистической традиции грамматический подход к словообразованию, когда оно включается в состав грамматики. Основанием для этого является тот факт, что словообразовательные средства языка – это морфемные средства, и поэтому механизм словообразования во многом сближается с механизмом морфологического уровня языка, в результате чего проблемы словообразования тесно переплетаются с проблемами морфологии.

При первоначальном знакомстве с языком главное внимание обращается не столько на звуковой состав слова, сколько на морфемную структуру языка, т.е. на кратчайшие единицы, имеющие смысловое значение. Морфемы могут выражать три типа значения: вещественное, деривационное и релятивное. Носителем вещественного значения является корневая морфема, деривационного – приставки и суффиксы. Они выражают дополнительные признаки, ограничивающие и уточняющие значение корня. Реляционное значение обычно выражается окончанием, которое показывает отношение слова к другим словам в предложении [2].

В [3] предложена формализация словообразовательной семантики в интеллектуальной системе с морфемными словарями. Формализация особенностей организации семантических единиц словообразования позволяет автоматизировать процесс построения производных слов с учетом заданной специфики грамматического и лексического значений.

В данной работе рассматривается решение задачи словообразовательного синтеза с использованием семантических свойств формантов.

## Ярусное представление словообразовательного гнезда

Проведем анализ соотношения компонентов немотивированной части значения мотивированных слов и значения словообразовательного форманта. Под словообразовательным формантом понимается наименьшее в формальном и семантическом отношении словообразовательное средство, которым отличается какое-либо слово от других слов, находящихся с ним в отношениях мотивации [4].

Формально определим мотивационное отношение следующим образом. Пусть слова  $a$  и  $b$  представляют собой некоторые упорядоченные множества символов, т. е.

$$A = \{a_1, b_1, b_2, \dots\}; B = \{b_1, b_2, \dots\} \quad (1)$$

и

$$A - B \neq 0 \quad (2)$$

Установив, что слово  $A$  имеет некоторый формант, отсутствующий в слове  $B$ , находящемся в отношении мотивации со словом  $A$ , определяем, что слово  $A$  является мотивированным, а слово  $B$  – мотивирующим., т. е. установлено направление мотивации.

Формант может включать в себя одно или несколько словообразовательных средств. Таким образом, формально можно определить формант как множество словообразовательных средств:

$$R = [P, S], \quad (3)$$

где

$S = \{s_1, s_2, \dots\}$  - множество суффиксов, входящих в словообразовательный формант R;

$P = \{p_1, p_2, \dots\}$  - множество префиксов, входящих в словообразовательный формант R.

Производное слово может быть в свою очередь по отношению к другому слову производящим. Подобные слова, находящиеся в отношениях последовательной мотивированности, образуют словообразовательный ряд [4].

Пусть  $k$  – корень данного ряда, а слово представляет собой множество, состоящее из корня, множества суффиксов и множества префиксов. Тогда словообразовательный ряд можно представить следующим образом

$$\{k\} \rightarrow \{k, s_1\} \rightarrow \{p_1, k, s_1\} \rightarrow \{p_1, k, s_1, s_2\} \rightarrow \{p_1, k, s_1, s_2, s_3\} \quad (4)$$

Обозначим текущий элемент словообразовательного ряда как  $A^j$ , а функцию, задающую переход к следующему элементу, -  $\varphi$ . Тогда

$$A^{j+1} = \varphi(A^j) \quad (5)$$

На основании приведенных определений можно установить, что отношения мотивации устанавливаются между членами одного и того же словообразовательного гнезда – совокупности слов с тождественным корнем. Также словообразовательное гнездо можно определить как совокупность словообразовательных рядов, имеющих одно и то же исходное слово.

Обозначим элемент словообразовательного гнезда  $A_{ik}^j$ , где  $i$  - порядковый номер элемента в  $j$ -том уровне, который образован от элемента предыдущего уровня с порядковым номером  $k$ . Элементы одного уровня не связаны друг с другом отношением мотивации, как элементы словообразовательного ряда. Поэтому каждый элемент словообразовательного гнезда можно вычислить как функцию от предыдущего элемента своего словообразовательного ряда. Учитывая (5), получим

$$A_{ik}^j = \varphi(A_{kn}^{j-1}) \quad (6)$$

Словообразовательное гнездо можно представить в виде ярусного графа, представленного на рисунке 1.

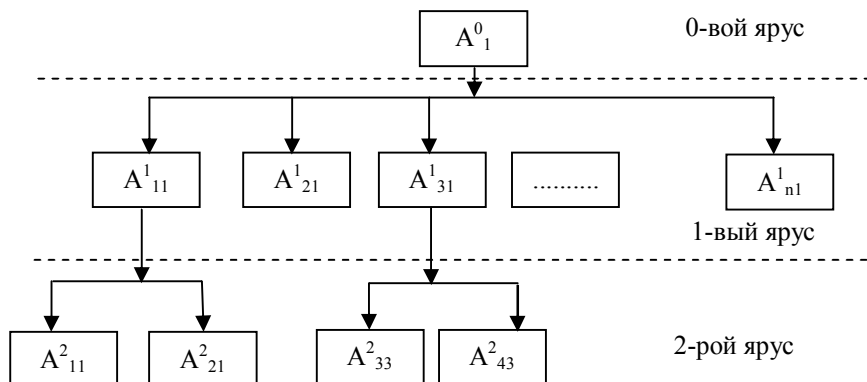


Рис. 1. Ярусный граф словообразовательного гнезда

## Формальная модель словообразовательного синтеза

В [3] семантика каждого слова  $X$  языка определена как функция  $f(X)$ . Слово  $X$  представляет собой упорядоченное множество морфем, т. е.

$$X = (X_1, X_2, X_3), \quad (7)$$

где  $X_1$  – множество приставок, входящих в слово.  $X_1 = (x_1^1, x_1^2, \dots)$ .

Так как слово может не содержать приставку, то  $|X_1| \geq 0$  – мощность множества  $X_1$ ;

$X_2$  – множество корней, входящих в слово.  $X_2 = (x_2^1, x_2^2, \dots)$ . Так как слово обязательно содержит хотя бы один корень,  $|X_2| > 0$ ;

$X_3$  – множество суффиксов, входящих в слово.  $X_3 = (x_3^1, x_3^2, \dots)$ . Так как слово может не содержать суффикс, то  $|X_3| \geq 0$ .

Производные слова определенной части речи связаны между собой многообразными отношениями словообразовательного характера. Лишь некоторые префиксы и единичные суффиксы выступают одновременно в разных частях речи. Остальные аффиксы присущи словам только одной части речи. Разного вида семантические взаимодействия (совмещение значений и их отдельных компонентов), взаимоотношения между производными словами разных способов словообразования осуществляются только в пределах одной части речи. Таким образом,

семантическое значение аффикса есть функция двух переменных: аффикса и части речи, т. е.  $f(x_i^i, z_k)$ , где  $z_k$  – элемент множества всех частей речи  $Z$ .

Пусть  $M$  – множество всех слов языка, тогда справедливо следующее

$$\forall x \in M \quad f(x) = f(x_1^i, z_k) \cup f(x_2^j, z_m) \cup f(x_3^h, z_n) \quad (8)$$

Определим процесс словообразования как

$$G = (I, O, \varphi), \quad (9)$$

где  $I$  – множество входных воздействий;

$O$  – множество мотивированных слов, образующих словообразовательное гнездо;

$\varphi$  – функция перехода (закон словообразования), задающая отображение  $\varphi: I \rightarrow O$ .

Множество входных воздействий определяется следующим образом

$$I = (C, z, X), \quad (10)$$

где  $C$  – множество ключевых слов, определяющих семантику производного (мотивированного) слова;

$z$  – часть речи;

$X$  – исходное (мотивирующее) слово.

При выборе аффикса, необходимого для данного закона словообразования, учитывается его семантическое значение, ключевые слова которого должны соответствовать множеству  $C$ , т. е., если  $x^i$  – некоторый аффикс, то  $f(x^i, z) \cap C \neq \emptyset$ .

Таким образом, задача словообразования сводится к поиску пути к вершине ярусного графа словообразовательного гнезда, соответствующей образуемому слову. Требуется определить именно путь, так как дуги графа нагружены аффиксами, соответствующими закону словообразования, применяемому на этом переходе.

Для сокращения числа возможных переходов из текущей вершины яруса целесообразно использовать параметр  $z$ , задающий часть речи образуемого слова.

Дальнейшее сокращение числа возможных путей проводится с помощью множества ключевых слов  $C$ , определяющих семантику образуемого слова. Для каждой вершины текущего яруса определяем меру близости.

Значение производящего слова всегда в той или иной мере организует семантику производного. Как правило, лексическое значение производящего слова входит в лексическое значение производного в полном объеме, всеми своими семантическими

компонентами. Введем понятие меры семантической близости  $\delta$  слов  $A$  и  $B$

$$\delta(A, B) = f(A) \cap f(B) \quad (11)$$

Следующей вершиной пути на очередном ярусе считается вершина, обладающая максимальной мерой близости

$$\max_k \delta(A_{ik}^j, C) \quad (12)$$

Условием окончания работы алгоритма словообразования является достижение вершины  $A_{ik}^j$ , для которой  $f(A_{ik}^j) = C$ .

Для проверки соответствия полученного слова заданной семантике (множеству ключевых слов  $C$ ), решается обратная задача, состоящая в морфологическом разборе слова по алгоритму [5] и определении его семантического значения с помощью формулы (8).

## Заключение

Предложенная формальная модель словообразовательного синтеза, учитывающая семантические свойства формантов, а также заданные грамматические и лексические значения производного слова, используется при разработке интеллектуальных обучающих систем, разрабатываемых в институте проблем искусственного интеллекта. Планируется дальнейшая детализация модели, учитывающая схемы словообразования определенных частей речи.

## Литература

1. *Фатхутдинова В.Г.* О принципе сопоставимости комплексных единиц словообразования разноструктурных языков // Русская и сопоставительная филология: Системно-функциональный аспект / Казан. гос. ун-т; филол. фак-т. Казань, 2003. – С. 172-178.
2. *Улукханов И.С.* Словообразовательная семантика в русском языке. М.: Наука, 1977. – 255 с.
3. *Егошина А.А.* Формальная модель словообразовательной семантики// Радиоэлектроника. Информатика. Управление. Запорожье: ЗНТУ, 2008. №1. С. 54-57
4. *Милославский И.Г.* Вопросы словообразовательного синтеза. М.: Наука, 1980. – 292 с.
5. *Егошина А.А.* Алгоритмические методы доступа к словарям морфологического процессора//Научные труды Донецкого национального технического университета. Серия «Информатики, кибернетика и вычислительная техника» (ИКВТ-2007). Выпуск 8(120) – Донецк: ДонНТУ. – 2007. С. 107-117