# ТЕОРЕТИКО-МНОЖЕСТВЕННОЕ МОДЕЛИРОВАНИЕ СЕМАНТИЧЕСКИХ ПОЛЕЙ СЛОВ

Черепанова Ю. Ю.
Харьковский государственный технический университет радиоэлектроники

The article deals with set-theoretical modeling of semantic fields. The thesaurus of semantic fields is one of tools for the description of semantics of natural language that can be used in the systems of natural language understanding for solving some problems. The received set-theoretical model of the thesaurus of semantic fields can be used as a basis for the further modeling of semantic fields.

Тезаурус семантических полей является одним из инструментов для описания семантики и анализа естественного языка, позволяющим решить некоторые проблемы, возникающие в процессе обработки естественно-языковых текстов. Семантические поля представляют собой структуры, включающие в себя лексические единицы различных уровней и связывающие их семантические отношения. Такая структура семантических полей дает возможность моделирования их с помощью математического аппарата теории множеств, теории графов и теории категорий.

Формализуем лингвистические модели семантических полей с помощью теоретико-множественного аппарата.

Определим следующие множества:

– $\mathbf{U=\{v_1,v_2,...,v_M\}}$ – множество слов языка;

– $\mathbf{A=\{a_1,a_2,...,a_K\}}$ – множество семантических признаков;

– $\mathbf{T=\{\tau_1,\tau_2,...,\tau_N\}}$ – множество семантических типов.

Множества **U**, **A**, **T** представляют собой лексические единицы различных уровней. Для отражения существующих между ними семантических связей введем следующие отношения.

Определим отображение $\boldsymbol{f_{код}: U \to A,}$ где $\mathbf{U=\{v_1,v_2,...,v_M\}}$ – множество слов языка, $\mathbf{A=\{a_1,a_2,...,a_K\}}$ – множество семантических признаков. Данное отображение представляет собой кодирование слов в квазиосновы, семантические признаки. Это отображение является всюду определенным, сюръективным. Так как один признак может соответствовать нескольким словам (отображение не является инъективным), то $\mathbf{к \le м}$.

Отношение $\mathbf{R_{см} \subseteq A \times A}$ содержательно означает «иметь семантический множитель». Это отношение всюду определенное, рефлексивное, транзитивное. Тогда сечения $\mathbf{R_{см}(a_i)}$ будут представлять собой семантические объемы слов $\mathbf{v_i = f_{код}^{-1}(a_i)}$.

При построении семантических полей все множество слов **U** разбивается на множество семантических полей $\mathbf{U_1,U_2,...,U_N}$. Семантическое поле $\mathbf{U_i}$ – некоторое множество слов $\mathbf{\{v_{1i},v_{2i},...,v_{Mi}\}}$, связанных между собой парадигматическими и синтагматическими семантическими отношениями, соответственно разделяются парадигматические и синтагматические семантические поля. К парадигматическим полям относятся различные классы лексических единиц, имеющие общие признаки в своих значениях. То есть если каждому слову $\mathbf{v_i}$ поставить в соответствие множество $\mathbf{W_i = R_{см}(a_i) =\{a_{i1},a_{i2},...,a_{iK}\}}$ (семантический объем), где $\mathbf{a_{iL}}$ – семантический признак слова $\mathbf{v_i}$, то слова $\mathbf{v_i}$, $\mathbf{v_j}$ объединяются в одно парадигматическое поле $\mathbf{U_z}$ при условии, что $\mathbf{W_i \cap W_j \neq \varnothing}$.

Таким образом, можно задать отношение $\mathbf{R_п \subseteq U \times U}$, определяющее наличие парадигматических связей между словами. Отношение $\mathbf{R_п}$ всюду определенное, сюръективное, рефлексивное, симметричное. Сечения $\mathbf{R_п(v_i)}$ будут представлять собой парадигматические поля слов $\mathbf{v_i}$.

Синтагматическими называются отношения между сочетаемыми единицами языка. Таким образом, синтагматические семантические поля слова представляют собой связи слов либо их абстрактных категорий в текстах. Для отражения этих связей необходимо ввести несколько отношений.

Введем всюду определенное, сюръективное отношение $\mathbf{R_{ст} \subseteq U \times T}$ означающее «сочетаться с семантическим типом».

Отношение $\mathbf{R_{тт} \subseteq T \times T}$ означающее «сочетаться», всюду определено, сюръективно, симметрично, антирефлексивно.

Введем отображение $\mathbf{f_т : U \rightarrow T}$, содержательно означающее «относиться к семантическому типу».

Тогда отношение $\mathbf{R_{сп} \subset R_{ст}\, f_т^{-1}}$ будет определять синтагматические связи между словами. Сечения $\mathbf{R_{тт}(т_j)}$ будут представлять собой синтагматические связи категории $\mathbf{т_j}$, сечения $\mathbf{R_{ст}(v_i)}$ – синтагматические связи слов $\mathbf{v_i}$ с абстрактными категориями в текстах, а сечения $\mathbf{R_{сп}(v_i)}$ – связи слов $\mathbf{v_i}$ с другими словами.

Разумеется, данная модель является очень упрощенной. Отношения $\mathbf{R_п}$ и $\mathbf{R_{сп}}$ отражают лишь наличие соответствующих связей между словами. На самом деле парадигматические и синтагматические связи должны отражаться множеством отношений. Кроме того, данная модель отражает многозначность лексических единиц лишь в том случае, если разные значения слов будут отображаться как разные слова $\mathbf{v_i}$, $\mathbf{v_j}$.

Полученную теоретико-множественную модель тезауруса семантических полей будем использовать как основу для дальнейшего моделирования семантических полей.