

# СЕМАНТИЧЕСКОЕ ПРОСТРАНСТВО В ФАКТОРИАЛЕ ТЕОРИИ АВТОМАТИЧЕСКОГО ПОРОЖДЕНИЯ АРХИТЕКТУРЫ ЗНАНИЙ (ТАПАЗ-2)

**А. Н. Гордей**

---

*Белорусский государственный университет*

*Минск, Беларусь*

*e-mail: [alieks2001@yahoo.com](mailto:alieks2001@yahoo.com)*

Комбинаторные возможности ТАПАЗ-2 [1] в компьютерной обработке данных на естественном языке, понимании естественного языка, решении изобретательских задач представлены в виде 696 192 модификаций 112 макропроцессов базового исчисления с умножением на 31 роль индивидов каждой предметной области. Таким образом, общее количество формализмов ТАПАЗ-2 составляет 21 581 952, тогда как «Словарь современного русского литературного языка» в 17 томах содержит 120 480 слов, заявленный объем «Большого академического словаря русского языка» – 150 000 слов, а имеющиеся электронные ресурсы Института лингвистических исследований РАН на 1,4 млрд словоупотреблений содержат около 5 млн русскоязычных слов XVIII – XXI вв. [2].

*Ключевые слова:* семантический код; семантический примитив; исчисление предметных областей; модификация акции; рекурсия; искусственный интеллект.

# THE SEMANTIC SPACE IN FACTORIAL OF THE THEORY FOR AUTOMATIC GENERATION OF KNOWLEDGE ARCHITECTURE (TAGKA-2)

**A. N. Hardzei**

---

*Belarusian State University*

*Minsk, Belarus*

Combinatorial possibilities of TAGKA-2 [1] in Natural Language Processing, Natural Language Understanding, Inventive Problem Solving are presented by 696 192 modifications of 112 classes of actions in base calculation multiplied by 31 roles of individuals in each Field of Knowledge. Thus, the total number of TAGKA-2 formalisms is 21 581 952, while the Dictionary of Modern Russian Literary Language (in 17 volumes) contains 120 480 words. The amount claimed by authors of Russian Large Academic Dictionary is 150 000 words and the available electronic resources of the Institute of Linguistic Studies of Russian Academy of Sciences contain about 5 million Russian-speaking words of XVIII – XXI centuries based on 1,4 billion word usages [2].

*Keywords:* semantic code; semantic primitive; calculus of fields of knowledge; modification of action; recursion; artificial intelligence.

## ВВЕДЕНИЕ

Т. Виноград в исследованиях по искусственному интеллекту поставил перед семантикой четкую задачу – в практическом плане создать «преобразователь, который может работать с синтаксическим анализатором и выдавать данные, приемлемые для логической дедуктивной системы. Если имеются синтаксический анализатор для грамматики английского языка и дедуктивная система на базе знаний о конкретном предмете, роль семантики сводится к тому, чтобы заполнить пространство между ними» [3]. Однако как добиться этого, никто не знал. К чести белорусской науки, прогресс в данной сфере во многом обязан работам В. Мартынова по семантическому кодированию [4]. Создатель УСК считал, что для оснащения компьютера энциклопедическими базами знаний необходимо:

1. «Исчислить примитивы, т. е. семантически неразложимые ключевые слова и правила их комбинаторики.
2. Установить необходимый и достаточный набор формальных характеристик, составляющих “словарную статью”.
3. Определить набор семантических операций, позволяющий исчислять предметные области любого вида.
4. Выработать эвристические правила обучения работе с такой системой.
5. Разработать систему взаимных отсылок на семантической основе» [5].

В 1994 г. нами под руководством В. Мартынова была впервые предложена процедура исчисления предметных областей в виде особо ориентированного графа ранжирования сложных цепочек [6]. Применение процедуры потребовало установления одно-однозначного (векторного) перехода между макропроцессами в базовом семантическом классификаторе и привело к созданию теории автоматического порождения архитектуры знания (ТАПАЗ), основу которой составили формализованная теория, семантический двойник, таблица семантических элементов (макропроцессов), алгоритм ролей индивидов и граф поиска гипонимов через гиперонимы [7]. В 2014 г. появилась 2-я версия ТАПАЗ. В ней значительно упрощен алгебраический аппарат, увеличены правила интерпретации типовых совмещений индивидов, а количество операций сокращено до двух. Теперь это была алгебра вида:  $A = \langle M, *, - \rangle$  [1].

## ЭВОЛЮЦИЯ СЕМАНТИЧЕСКОГО КОДИРОВАНИЯ

### *Запись процесса сжатия*

- в версии УСК-1:  $S(a)A_1\bar{S}(a)A O/O$  [8];
- в версии УСК-6:  $((XY)Z)((ZW)W')$  [9];
- в версии ТАПАЗ-2:  $Z((\bar{Z}\bar{W})W)$  [1].

## ЭВОЛЮЦИОННЫЕ ОТЛИЧИЯ ТАПАЗ-2 ОТ УСК

- Введена геометрическая модель и верифицирована непротиворечивость алгебры. Запрещены построения, приводящие к математическим и семантическим парадоксам:

$XX$   
 $XYX$   
 $XYZZWZ$   
 $XYZZWZ_{(X)}$

и установлено строгое правило правостороннего наращивания множителей:

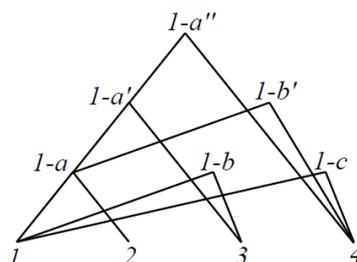
$$X \rightarrow Y \rightarrow Z \rightarrow W.$$

- Уточнены алгебраические операции.
- Представлены в явном виде правила построения, ограничения, сокращения и преобразования алгебраических выражений.
- Количество правил преобразования сведено к одному – *транспозиционному*.
- Установлены некоммутативные одно-однозначные (векторные) переходы между алгебраическими выражениями и их аргументами:

$$((X * \bar{X}) * \bar{Y}) * \bar{Y} \rightarrow (X * (\bar{X} * \bar{Y})) * \bar{Y} \rightarrow X * ((\bar{X} * \bar{Y}) * \bar{Y}) \rightarrow X * (\bar{X} * (\bar{Y} * \bar{Y})) \rightarrow$$

притягивание	скапливание	ужимание	присоединение
$Y_2 \text{ imp } X_2$			

- Предложена процедура исчисления предметных областей и семантики предметных областей в виде особо ориентированного графа:



- Построен алгоритм ролей индивидов (макрообъектов).
- Семантические примитивы (макропроцессы) исчислены в количестве **112** (в УСК-6 их 108) и сведены в макрогруппы и макроряды таблицы семантических элементов.
- Все микропроцессы рассматриваются как модификации макропроцессов; количество микропроцессов только на первой ступени модификации составляет  $112 \cdot 111 = \mathbf{12\ 432}$ , на второй убывает в прогрессии  $n-1$  по второму множителю, т. е.  $112 \cdot 110 = \mathbf{12\ 320}$ , что одновременно приводит к почти двукратному увеличению общего числа микропроцессов:  $12\ 432 + 12\ 320 = \mathbf{24\ 752}$ , **таким образом, мощность системы оказывается в интервале факториалов (9!; 10!), т. е. чрезвычайно большой и вместе с тем конечной**<sup>3</sup>.
- Сформулированы правила интерпретации типовых совмещений индивидов для реконструкции семантических двойников алгебраических выражений, например:

$\bar{\bar{\alpha}}_1 * \bar{\alpha}_1$  после  $\alpha_1 * \bar{\alpha}_1$  (совмещение среды  $\alpha_1$  с оболочкой  $\alpha_1$  как результат физического воздействия на среду  $\alpha_1$ ) – *формование*  $\alpha_1$ ;

<sup>3</sup> Точное количество микропроцессов подсчитано Ч. Рутковским по формуле

$$S_n = \frac{2a_1 + d(n-1)}{2} \cdot n, \text{ где } L = 112, a_1 = L-1, d = -1, n = L-1;$$

$$S_n = \frac{2(L-1) + (-1)(L-1-1)}{2} \cdot (L-1) = \frac{2L-2-L+2}{2} \cdot (L-1) = \frac{L(L-1)}{2};$$

$$\Sigma = 112 \cdot S_n = \frac{L^2(L-1)}{2} = 696\ 192.$$

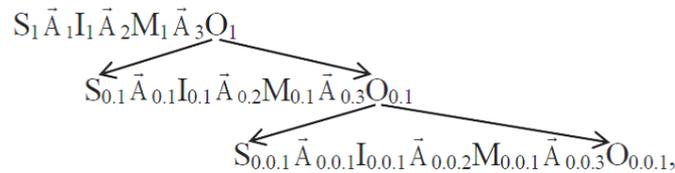
С учетом 31 роли индивидов каждой предметной области общее количество формализмов ТАПАЗ-2 составляет  $696\ 192 \cdot 31 = \mathbf{21\ 581\ 952}$ .

$\bar{\alpha}_1 * \bar{\alpha}_1$  после  $(\alpha_1 * \bar{\alpha}_1) * \bar{\alpha}_1 \approx \alpha_1 * \bar{\alpha}_1$  (совмещение среды  $\alpha_1$  с оболочкой  $\alpha_1$  как результат информационного воздействия на среду  $\alpha_1$ ) – *предрасположение*  $\alpha_1$ ;

$\bar{\alpha}_1 * \alpha_1$  после  $\alpha_1 * \bar{\alpha}_1$  (совмещение оболочки  $\alpha_1$  с  $\alpha_1$  как результат физического воздействия) – *формирование*  $\alpha_1$ ;

$\bar{\alpha}_1 * \alpha_1$  после  $(\alpha_1 * \bar{\alpha}_1) * \bar{\alpha}_1 \approx \alpha_1 * \bar{\alpha}_1$  при  $\alpha_1 = \alpha, \beta$  (совмещение оболочки  $\alpha_1$  с  $\alpha_1$  как результат информационного воздействия  $\beta$ ) – *воспитание*  $\alpha_1$ ;

- Введена процедура ступенчатого наращивания множителей в алгебраических выражениях при рекурсивном расширении геометрической модели:



где S – субъект,  $\bar{A}$  – акция, I – инструмент, M – медиатор, O – объект.

- Исключены составные цепочки, а описание события сведено либо к некоммутативной суперпозиции макропроцессов, либо к их некоммутативной препозиции, причем суперпозиция или препозиция макропроцессов означает такую же некоммутативную суперпозицию или препозицию микропроцессов в рекурсии<sup>4</sup>.

<sup>4</sup> Например: процесс *носить* не считается комбинацией процессов *фиксировать* и *перемещать*, как в УСК-6, а рассматривается как модификация процесса перемещения в *микросистеме составного инструмента*, тогда как процесс *возить* определяется в качестве модификации процесса перемещения в *микросистеме составного медиатора*. При таком подходе процесс фиксации оказывается в суперпозиции к процессу перемещения, т. е. **можно фиксировать, но не перемещать, однако нельзя перемещать, не зафиксировав прежде**. Невозможно сдвинуть с места автомобиль, сколько его не толкай, если он ручным тормозом зафиксирован относительно земли, его придется снять с тормоза и зафиксировать относительно рук путем прижатия их к автомобилю. Другой пример: нельзя сдвинуть с места стену, сколько ее не толкай, потому что она зафиксирована относительно земли, для этого придется подложить под стену, скажем, салазки, расфиксировав ее тем самым относительно земли и создав условия для фиксации руками. **Таким образом, фиксация является предварительным условием любого перемещения, суперпозиция процессов очевидна, сперва фиксировать, затем перемещать:**

***фиксировать* → *перемещать*.**

Пример препозиции процессов: во время приготовления пищи не важно, что делать первым – включать конфорку и ставить на нее кастрюлю с водой или наоборот. Однако по технике безопасности предпочтительнее вначале ставить на плиту кастрюлю с водой, а потом зажигать конфорку:

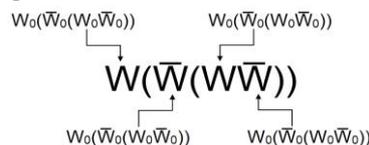
***перемещать* → *зажигать*.**

Пример суперпозиции процессов в рекурсии: чтобы просверлить титановое покрытие, необходимо для охлаждения лить на сверло масло, иначе сверло не выдержит, т. е. вначале литье масла, а потом сверление. Иными словами, **процесс сверления развивается на фоне литья масла:**

***лить* → *сверлить*.**

**Таким образом, для возникновения и развития события необходим векторный переход между процессами.**

Пример рекурсии процесса нормализации в ТАПАЗ-2:



Канонизированная запись текста на русском языке: «*Врач посредством лекарства вылечил больного*».

## ИЗОБРЕТАТЕЛЬСКАЯ ЗАДАЧА

Как очистить яйцо от скорлупы? В подгруппе *среда – оболочка* таблицы семантических элементов (макропроцессов) – *раздробить*; в подгруппе *оболочка-ядро* – ввести расщепляющий реагент – *расщепить*; в подгруппе *ядро – оболочка* – *разжать* изнутри расширением объема ядра.

## ЗАКЛЮЧЕНИЕ

ТАПАЗ-2 – один из возможных проектов исчисления семантики. Несмотря на то что этот проект по ряду показателей превосходит аналоги, а в исчислении предметных областей их не имеет, он не претендует на исключительность. Языковая семантика многогранна и допускает различные способы формализации. Однако все способы, подобно евклидовым и неевклидовым геометриям, должны быть непротиворечивы и эффективны в решении стоящих перед ними задач, и те, кто спорят с этим, по меткому выражению Г. фон Рейхенбаха, лишь «путают строгость метода с ограниченностью цели» [10].

## БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Гордей А. Н. Теория автоматического порождения архитектуры знаний (ТАПАЗ-2) и дальнейшая минимизация семантических исчислений // Открытые семантические технологии проектирования интеллектуальных систем = Open Semantic Technologies for Intelligent Systems (OSTIS-2014) : материалы IV Междунар. науч.-техн. конф., Минск, 20–22 февр. 2014 г. / редкол.: В. В. Голенков (отв. ред.) [и др.]. Минск : БГУИР, 2014. С. 49–64.
2. Кругликова Л. Е. «Большой академический словарь русского языка» как продолжатель традиций русской академической лексикографии = «El gran diccionario académico de la lengua rusa» como continuador de las tradiciones de la lexicografía académica rusa // Cuadernos de Rusística Española. 2012. № 8. С. 177–198.
3. Виноград Т. Программа, понимающая естественный язык. М. : Мир, 1976. С. 51.
4. Мартынов В. В. Кибернетика. Семиотика. Лингвистика. Минск : Наука и техника, 1966. В центре сознания человека. Минск : БГУ, 2009.
5. Мартынов В. В. Основы семантического кодирования. Опыт представления и преобразования знаний. Минск : ЕГУ, 2001.
6. Гордей А. Н. Процедуральная семантика и исчисление предметных областей // Язык: семантика, синтактика, прагматика : материалы I Междунар. науч. конф., Минск, 21–22 июня 1994 г. / Минск. гос. лингв. ун-т ; редкол.: Д. Г. Богушевич [и др.]. Минск, 1995. Ч. I. С. 16–17.
7. Гордей А. Н. Дедуктивная теория языка. Минск: Беларус. навука, 1998; Принципы исчисления семантики предметных областей. Минск : БГУ, 1998.
8. Мартынов В. В. Семиологические основы информатики. Минск : Наука и техника, 1974. С. 179.
9. Мартынов В. В. В центре сознания человека. Минск : БГУ, 2009. С. 91.
10. Рейхенбах Г. Философия пространства и времени. М. : Прогресс, 1985. С. 16.