

СИНТЕЗ-ОБОСНОВАНИЕ ИДЕАЛА СТРУКТУРЫ ЯЗЫКА НАУЧНОЙ ТЕОРИИ

Abstract: Some approximation to semantic model of ideal structure of scientific theory language has been synthesized. This approximation combines consistently pithiness and formality. The results are coordinated with famous results that don't raise author's doubts. The language structure of scientific theory consists of three blocks – signs, syntax and semantics of scientific theory languages. The each block structure is a hierarchical net of identical modules. The blocks are intercommunicated with each other, with sign image source-receiver and real world. Each module functioning is determined by solution tasks of analysis and synthesis structure of some information presentation.

Key words: structure, scientific theory, language, morphology, syntax, semantics.

Анотація: Синтезовано деяке наближення до семантичної моделі ідеалу структури мови наукової теорії, що без протиріччя поєднує змістовність із формальністю та узгоджене з відомими результатами, які поділяє автор. Структура мов наукових теорій складається з чотирьох блоків (знаків, синтаксису мов, семантик і функцій наукових теорій) ідентичної структури в формі ієрархічної мережі ідентичних модулів і взаємопов'язаних між собою, з джерелами-приймачами знакових образів та з реальним світом. Функціонування всякого модуля визначається вирішенням задач аналізу і синтезу структури деякого представлення інформації.

Ключові слова: структура, наукова теорія, мова, морфологія, синтаксис, семантика.

Аннотация: Синтезировано некоторое приближение к семантической модели идеала структуры языка научной теории, непротиворечиво сочетающей содержательность с формальностью, и согласовано с известными результатами, разделяемыми автором. Структура языков научных теорий состоит из четырех блоков (знаков, синтаксиса языков, семантик и функций научных теорий) идентичной структуры в форме иерархической сети идентичных модулей и взаимосвязанных между собой, с источниками-приемниками знаковых образов и с реальным миром. Функционирование всякого модуля определяется решением задач анализа и синтеза структуры некоторого представления информации.

Ключевые слова: структура, научная теория, язык, морфология, синтаксис, семантика.

1. Введение

Всякая научная теория доступна для использования и развития через ее язык, и всякому языку присуща некая теория. Согласно А. Черчу: "Язык обязывает делать именно те онтологические допущения, которые формулируются в аналитически истинных предложениях данного языка" [1, с.127, с.131]. Его дополняет У.В.О. Куайн: "говорим разумно о ...предложении как об истинном ..., когда мы обращаемся к положениям фактически существующей в данный момент теории, принятой хотя бы в качестве гипотезы" [2, с.47].

Всякая научная теория использует естественный язык и, развивая его, становится его частью. В лингвистике общепринято, что: "язык составляет одну из целого ряда систем знаний", "богатую и сложную систему со специфическими особенностями"; отличаясь от "концептуальной структуры" "реального мира" [3, с.111, с.129; 4, с.175; 5, с.74; 6, с.263], язык дополняет ее для функционирования всякой научной теории в обществе.

Наиболее существенные признаки языка сводятся к двум взаимнообратным функциям:

– в процессе анализа входной вполне упорядоченной информации синтез структуры составляющих всякого осмысленного высказывания;

– порождение, синтез выходной вполне упорядоченной информации согласно структуре составляющих всякого осмысленного высказывания.

Эти свойства – часть врожденных способностей людей – обеспечивают неосознаваемое владение всяким родным языком и исследуются в генеративной грамматике, ставящей цель синтеза теории языковой компетенции как сущностной основы языковой деятельности. "Хомский впервые

предпринял попытку формального описания языка как естественного явления" [7, с.17] в форме эмпирической теории. "Хорошая формализованная грамматика ... сопоставляет ... структурную характеристику каждому элементу бесконечного класса предложений и ... делает это совершенно формально" [8]. Теорию языка, грамматику считают процессором, выполняющим двусторонние формальные преобразования между представлением смысла всякого высказывания и образной (в частности, фонетической) формой его выражения.

Лингвистическая теория прошла сложный путь от Стандартной теории 60-х годов [9, 10] через Расширенную стандартную теорию [5], Теорию управления и связывания и Теорию принципов и параметров [3, 4] к Минималистской программе и ее развитию [7, 11], внося ясность в ряд проблем языка. Но определенность многих проблем языка все еще далека от требуемой для компьютерной реализации полноценной естественной языковой системы, среди них центральной остается проблема синтеза функциональной структуры (архитектуры) языка.

Целью статьи является синтез некоторого приближения к идеалу структуры языка научной теории, согласованной с известными результатами, разделяемыми автором.

2. Структура системы научных теорий

Исходя из результатов логико-методологических и лингвистических исследований [7, 11 и др.], структуру системы научных теорий представим иерархией четырех уровней множеств разных эмпирических теорий (рис. 1): 1 – множество SV-блоков знаков, теорий выражения предложений s -языка ($s = 1 \div S$) разными v -образами ($v = 1 \div V$) знаков (фонологии, текста и др.); 2 – множество S -блоков синтаксиса, теорий синтаксиса разных s -языков ($s = 1 \div S$); 3 – блок семантики, множество K -теорий предметных k -миров ($k = 1 \div K$); 4 – блок функций научных теорий, связанных с источниками и приемниками образов, с предметными мирами и между собой через репрезентации: 5 – множество V -репрезентаций, анализируемых и порождаемых предложений s -языков, выраженных в знаковых v -образах; 6 – множество S -репрезентаций структур предложений в идеальных знаках алфавитов s -языков; 7 – множество структур логических форм высказываний; 8 – множество K -репрезентаций ситуаций анализа и порождения объектов предметных k -миров ($k = 1 \div K$).

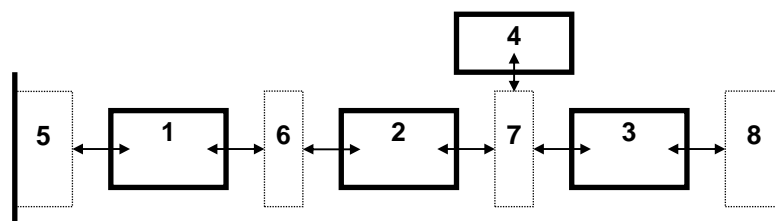


Рис. 1. Структура системы научных теорий и их языков

Здесь блок 4 функций – общий для всех научных теорий; логическая форма информации - общая для блоков синтаксиса, семантики и функций; у каждого s -языка свои блоки знаков и синтаксиса, но общая семантика ("концептуальная структура ... не зависит от конкретного языка" [7, с.170]), что обусловлено единством реальности и решаемых проблем [12, с.139]. Эта

структура уточняет предложенную в [7] "модель языка с тремя независимыми источниками дискретной бесконечности: фонологией, синтаксисом и семантикой". Две первые концептуальные необходимости минимализма Н. Хомского здесь выражены явно: репрезентацией 5 ("язык должен ... быть услышанным и озвученным") интерфейс вычислительной системы с мирами знаковых образов и репрезентацией 7 ("язык должен ... выражать мысль") интерфейс с "концептуально-интенциональной системой", а третья (интерфейс с лексиконом), в отличие от структуры Р. Джакендоффа, – неявно, репрезентацией 6 и распределением лексикона и его функций между блоками 1-3. Целесообразность и возможность разделения лексикона на морфонологию, морфосинтаксис и лексическую семантику обоснована в [7], а возможность деления функций "лексического лицензирования", закрепленных за "правилами соответствия", следует из представимости всякого интерактивного интерфейса двумя активными узлами, взаимосвязанными пассивной репрезентацией. Кроме того, предлагаемая структура в форме блока 4 дополнительно выражает четвертую концептуальную необходимость – управления функциями научной теории, а блока 3 и репрезентаций 8 – также новую (но общепринятую в программировании) – пятую концептуальную необходимость: смысл всякой мысли, выраженной в языке, должен быть выполнен конструктивно операциями с объектами предметного мира. Последнее – единственное естественное оправдание самого существования любого языка его необходимостью для жизнедеятельности и эволюции вида.

Эта структура вполне согласуется с традиционной лингвистикой, отдельно изучающей морфонологию, синтаксис и семантику всякого языка, и с заключением исследований языковой деятельности головного мозга об иерархии структуры сети языковых блоков и модулей [12, с.136]: "... обработка языковых данных ... распределяется между множеством разных функционально-анатомических модулей"; модулем является "интерактивная" "система операций по обработке информации". Наиболее существенными признаками семантического модуля являются [12]:

- "информационная инкапсуляция";
- "самодостаточность в отношении внимания";
- "когнитивная непроницаемость ... (модуль) следует своим собственным операционным правилам независимо от стратегических или волевых соображений, относящихся к более высоким иерархическим уровням".

Сложность детализации структуры (рис. 1) существенно снижает гипотеза Т. Гивона относительно объяснения языковой способности эволюцией доязыковых структур головного мозга:

- Первичным является семантический модуль, решающий проблемы реальности. Морфонология и синтаксис – самые "молодые" "лингво-специфичные модули", первый "вероятно, старше" [12].
- В процессе эволюции нейронных схем модулей семантический блок приобрел форму иерархической сети автономных модулей, адаптируемой на решение разных проблем.
- "Факт, что каждый специализирующийся на выполнении того или иного языкового задания модуль ... представляет собой очевидную прямую адаптацию соответствующей доязыковой функции", причем "большинство механизмов ..., существенных для человеческого

языка, не являются ни новыми, ни лингвоспецифичными, но были ... вовлечены в процесс выполнения новых для них языковых заданий" [12, с.145, с.159].

На основании этой гипотезы проблема детализации структуры (рис. 1) сводится к проблемам модульной структуры блока семантики, структуры его модуля и их адаптации к специфике лингвистических задач.

3. Структура блока семантики

Структура блока семантики представляется многоярусной иерархической сетью вложенных теорий, содержащей (рис. 2) ярус 3_0 из множества независимых 3_{ok} семантических модулей фундаментальных теорий k -миров, взаимосвязанных с репрезентациями 7 – логической формы высказываний и $8 \{8_k, k = 1 \div K\}$ ситуаций объектов k -миров, и высшие ярусы 3_m ($m = 1 \div M$) из разного числа (N, \dots, W) семантических модулей частных теорий.

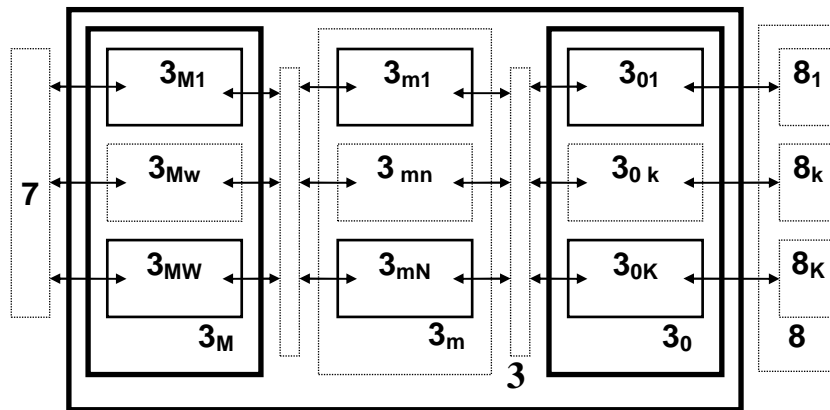


Рис. 2. Блок семантики

Структура семантического модуля 3_{ok} k -мира содержит (рис. 3): 9, 10 – эмпирические модели анализа и синтеза синтаксических структур; 11 – концептуальную модель синтаксиса структур логической формы высказываний; 12 – прагматику k -мира; 13 – концептуальный словарь; 14 – концептуальную модель семантики k -мира; 15, 16 – эмпирические модели действий и измерений в среде объектов k -мира. Словарь 13 – массив записей соответствия между ссылками на синтаксическую структуру (13_1) и на семантику (13_2) терминов концептуальных моделей.

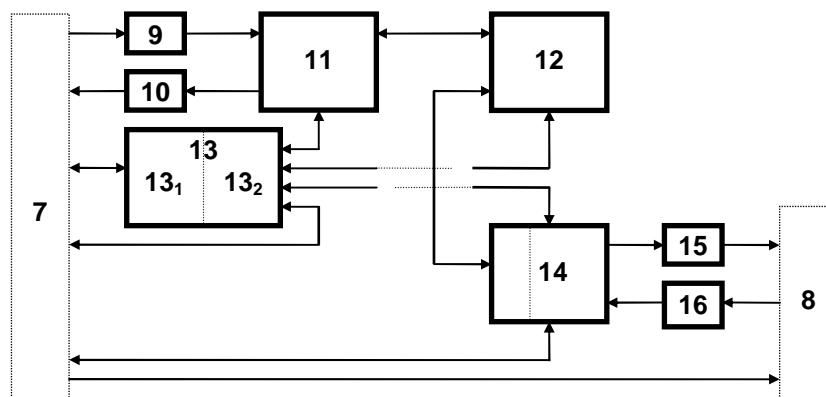


Рис. 3. Семантический модуль

Модуль Z_{ok} при решении задач k -ого мира функционирует под управлением концептуальной модели прагматики 12 модуля: синтаксические преобразования структуры высказывания во взаимодействии узлов 9, 10, 11, 12; модификация концептуальной модели 14 с использованием словаря 13; интерпретация концептуальной модели задачи во взаимодействии узлов 12, 14, 15, 16.

4. Структура блока знаков

Структура блока знаков ("артикуляторно-перцептивной системы" [11]) для всякого s -языка ($s = 1 \div S$) идентична модульной структуре блока семантики (рис. 2).

Структура фундаментального модуля знаков 1_{ov} ($v = 1 \div V$) всякой v -образной формы предложений всякого s -языка ($s = 1 \div S$) идентична структуре (рис. 3) семантического модуля Z_{ok} при специфичном содержании элементов: 7 – репрезентация структур v -образов знаков; 9, 10, 11 – эмпирические и концептуальная модели синтаксиса структур v -образов знаков; 12 – прагматика анализа и синтеза v -образов знаков; 13 – словарь, содержащий записи соответствий синтаксической и семантической структур v -образов знаков; 14, 15, 16 – концептуальная и эмпирические модели семантики v -образов знаков (применительно к фонологии 15 содержит процедуры управления артикуляторной, а 16 – перцепторной материальными системами); 8 – информационная репрезентация предложений v -образами предметного v -мира.

Задачи, решаемые модулями 1_{ov} ($v = 1 \div V$) всякого s -языка, являются вариантом задач анализа и синтеза объектов реальности, решаемых модулем Z_{ok} блока 3. Их специфика в отношении к объектам v -мира, как к v -образам идеальных знаков x_s алфавита s -языка.

Суть решения задач анализа модулей 1_{ovs} ($v = 1 \div V, s = 1 \div S$) представляется идентичной для всех v -образов и s -языков итерационной структурой (рис. 4): 1 - для текущего v -образа X_{ovs} предметного мира синтез протокола вычисления рекурсивной функции X_{vs} , удовлетворяющей минимуму оценки $\varepsilon_1 = F(X_{vs}, X_{ovs})$ нормы отклонения от v -образа X_{ovs} и допустимой оценке ε качества X_{vs} согласно ограничивающему предикату $P(X_{vs}) \equiv \varepsilon_1 \leq \varepsilon$. При успехе - к 2, иначе - к 5; 2 – исходя из протокола вычисления X_{vs} , синтез синтаксической формы X_{vs} с использованием словаря 13 знаков, установка на следующий v -образ и к 3; 3 – анализ знака X_{vs} : если это знак конца предложения, то к 4, иначе – к 1; 4 – вывод структуры предложения в идеальных x_s знаках, передача структуры вышестоящему модулю и к 1; 5 – продвижение к следующему v -образу и к 1.

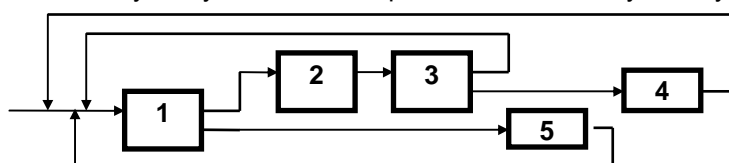


Рис. 4. Анализ образной формы предложений языка

Модель решения модулями 1_{ovs} ($v = 1 \div V, s = 1 \div S$) задач синтеза v -образов предложений имеет итерационную структуру (рис. 5): 1 – анализ репрезентации 7 на наличие очередного предложения: если есть, то переход к 2, иначе – возврат к 1; 2 – начиная со структуры предложения s -языка в идеальных x_s знаках – вывод структуры предложения в реальных X_{vs} знаках, переход к 3; 3 – установка на первый реальный X_{vs} знак предложения, переход к 4; 4 – для очередного знака X_{vs} посредством словаря установить структуру (протокол) вычисления соответствующего v -образа X_{ovs} , вычисление v -образа предметного v -мира, переход к следующему знаку X_{vs} предложения и к 5; 5 – анализ наличия знака X_{vs} : если есть, то к 4, иначе – возврат к 1.

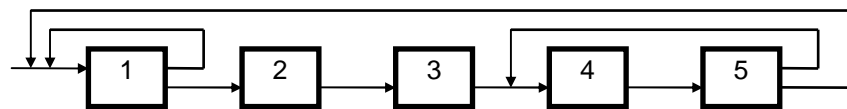


Рис. 5. Синтез образной формы предложений языка

Рассмотренные структура и функционирование модулей знаков яруса 1_{ovs} согласованы с результатами сегментной фонологии, исследующей внутреннюю структуру фонем "в виде многоуровневой иерархической структуры артикуляторных признаков" [13, с.186] и реализацию их контекстной взаимосвязи посредством синтаксических правил (типа правил ассимиляции) преобразования дифференциальных признаков фонем [13].

Структура модулей знаков высших ярусов и их функционирование 1_{mvs} ($m = 1 \div M, v = 1 \div V, s = 1 \div S$) в основном те же, что и для яруса 1_{ovs} , и решают они те же задачи анализа и синтеза (рис. 4, рис. 5), но по отношению не к внутренней структуре знаков, а к сочетаниям знаков. В фонологии это просодическая иерархия супrasegmentных теорий, изучающих структуры слога [13], слов и словосочетаний. Каждый уровень этой иерархии теорий представляется своим модулем знаков со специфичными моделями.

При решении задач анализа для физического представления *перцептивных* признаков синтезируется структура предложения в форме размеченной конечной последовательности слов, каждое из которых есть конечная последовательность идеальных знаков языка (итерация итерации знаков, завершающихся точкой). При решении задач порождения, исходя из структуры предложения в идеальных знаках, синтезируется в системе *артикуляторных* признаков физическое представление последовательности знаковых образов.

Подчеркнем, что в процессе анализа унифицируются разные исполнения образов (в частности, для произнесения слов "качество голоса говорящего, его тон и артикуляция" [7, с.165]) осуществляется контроль правильности слога- и словообразования, восстанавливается знаковый состав слов, передающий на высшие уровни языка только смысло-различительную информацию, необходимую и достаточную для синтаксического анализа и семантической интерпретации дискурса. И, наоборот, в процессе порождения смысло-различительная

информация семантики и синтаксиса дискурса преобразуется в информацию, необходимую и достаточную для реализации однозначно воспринимаемой ν -образной формы дискурса.

5. Структура блока синтаксиса

Структура блока синтаксиса многоязычной системы идентична модульной структуре блока семантики, а структура его модулей идентична структуре модулей блока семантики. Ее элементами являются лексический синтаксис, синтаксис предложений и семантика синтаксиса.

Модуль морфологии занимает фундаментальный, самый нижний ярус иерархической структуры теорий синтаксиса языка, реализует лексический синтаксис, выполняя двустороннее преобразование между словами (точнее, словоформами в идеальных знаках) языка и их парадигмальной структурой в системе морфологических свойств (признаков).

Известны разные структуры и формализации морфологии: массив парадигмальных статей для всех словоформ языка (типа используемой в Прологе [14]), таблично-алгоритмическое представление [15–19], *сетевая морфология* [20] и др., ни одна из них не адекватна лингвистической практике и каждой присущи свои недостатки. Первая – простейшая, использует минимум лексических знаний, но требует недопустимо больших затрат памяти и времени на решение лингвистических задач для морфологически сложных языков; вторая есть сочетание программистской эвристики с пассивным лингвистическим словарем, не удовлетворяет критерию простоты; сетевой морфологии присущи недостатки семантических сетей. На этом фоне остается продуктивным развитие грамматической морфологии, намеченной Н. Хомским еще в [10].

Структура модуля 2_{os} всякого s -языка ($s = 1 \div S$) для грамматической морфологии идентична структуре семантического модуля (рис. 3) со специфичным содержанием элементов: 7 – репрезентация последовательности структур парадигмальных разложений слов; 9, 10, 11 – эмпирические и концептуальная модели синтаксиса структур парадигмальных разложений слов; 12 – прагматика контекстного анализа и синтеза слов; 13 – словарь соответствий синтаксической и семантической структур морфологических признаков непосредственных составляющих слов; 14 – концептуальная модель морфологических признаков слов и их непосредственных составляющих; 15, 16 – эмпирические модели порождения и анализа непосредственных составляющих.

Задачи морфосинтаксиса, решаемые модулем 2_{os} всякого s -языка ($s = 1 \div S$), являются вариантом задач анализа и синтеза объектов реальности, решаемых модулями 3_{ok} блока 3. Их специфика в отношении к предметному миру как к линейному упорядочению идеальных знаков конечного алфавита s -языка. Задача анализа – синтез парадигмального разложения всякого слова из идеальных знаков, а синтеза – по всякому осмысленному парадигмальному разложению породить слово в идеальных знаках. Обе задачи из класса задач перевода и используют общие лексические знания (словообразования – словоизменения) и механизмы решения.

Морфоанализ – это первая задача синтаксиса, цель которой *"идентифицировать слова и отнести их к соответствующим категориям, используя ресурсы лексикона"* [21, с.134]. Из-за эффектов омонимии это недетерминированная задача, неразрешимая на уровне морфологии. В связи с этим, оправдывая экономией вычислительного ресурса, предлагают [16] ее решать в два

этапа: на первом строить комбинированную морфологическую структуру (кортеж множеств всех омонимов словоформ) предложения, а на втором – выбрать структуру, допустимую для понимания. Более естественным и экономным будет итерационный синтез разных морфологических структур до первой, согласованной синтаксисом языка. Тогда морфоанализ есть часть всей задачи синтаксического анализа (рис. 6): 1 – морфоанализ; 2 – высшие уровни синтаксического анализа; 3 – анализ результата: если успех, то передача результата следующему уровню языка, иначе – возврат к 1 за следующим вариантом морфоанализа.

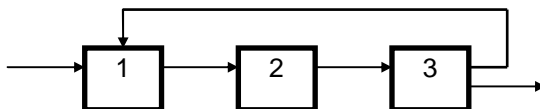


Рис. 6. Синтаксический анализ

Ниже в качестве примера, иллюстрирующего сложность и стиль описания задач анализа, дано в метаязыке [22] определение представительного фрагмента морфоанализа, формализующего часть общеизвестных знаний русской лексики [23], где метазнаки =_{dif}, ;, пробел, /, (), t, ' ', ?, !, # означают "есть по определению", конец определения, конкатенация, альтернативный выбор, итерационные скобки, истина, скобки константы знаков, анализ, порождение и анализ со следом (синтез C-маркера) соответственно. Дополнительно к выразительным средствам из [22] здесь использовано спаренное H, означающее переключение на использование в качестве входного массива (и обратно) линейного представления C-маркера.

Пример

- 1) морфоанализ =_{dif} слово ^тчк? Рun# (слово ^тчк? Рun#)? тчк#;
- 2) Рun =_{dif} пробел / зпт / тчкзпт / тчк;
- 3) слово =_{dif} функциональные_категории# / основные_категории / цепь_знаков;
- 4) функциональные_категории =_{dif} Conj / ADV / Part / Det / ... ;
- 5) Conj =_{dif} 'в_связи_с_тем,_что' / 'в_то_время_как' / 'после_того,_как' / 'а_также' / 'при_условии,_что' / ...;
- 6) ADV =_{dif} 'легко' / 'резко' / 'довольно' / 'совсем' / 'очень' / 'вместе' / 'отчасти' / 'вероятно' / 'почти' / ...;
- 7) Part =_{dif} 'а' / 'было' / 'бы' / 'б' / 'ведь' / 'вишь' / 'вон' / 'вот' / 'давай' / 'даже' / 'дай' / 'да' / 'же' / 'за' / 'едва' / ...;
- 8) основные_категории =_{dif} Prep# / Noun / Adj / Verb;
- 9) Prep =_{dif} 'благодаря' / 'для' / 'до' / 'за' / 'из' / 'ко' / 'к' / 'над' / 'о' / 'перед' / 'под' / 'по' / 'с' / 'у' / ...;
- 10) Noun =_{dif} имя_N# род_одуш_число_падеж;
- 11) имя_N =_{dif} N1 / N2 / N3 / ...;
- 12) N1 =_{dif} ДОМ / ГОРОД / ПОЕЗД / ... ;
- 13) ДОМ =_{dif} 'дом';
- 14) ГОРОД =_{dif} 'город';
- 15) ПОЕЗД =_{dif} 'поезд';
- 16) род_одуш_число_падеж =_{dif} усл_N11 N11 / усл_N21 N21 / ...;
- 17) усл_N11 =_{dif} H имя_N(N1(H;
- 18) N11 =_{dif} муж_N1#;
- 19) муж_N1 =_{dif} неод_N1;
- 20) неод_N1 =_{dif} ед_N1 / мн_N1;
- 21) ед_N1 =_{dif} ед_им_N1 / ед_род_N1 / ед_дат_N1 / ед_вин_N1 / ед_тв_N1 / ед_пр_N1;
- 22) мн_N1 =_{dif} мн_им_N1 / мн_родит_N1 / мн_дат_N1 / мн_вин_N1 / мн_тв_N1 / мн_пр_N1;
- ...
- 23) Adj =_{dif} имя_A# степень_форма_число_одуш_род_падеж;
- 24) имя_A =_{dif} A1 / A2 / A3 / ...;
- 25) A1 =_{dif} префикс основа_A1 / основа_A1;
- 26) префикс =_{dif} 'по' / 'наи';
- 27) основа_A_1 =_{dif} ВЕСЕЛЫЙ / ВЕРНЫЙ / КРАСИВЫЙ / ... ;
- 28) ВЕСЕЛЫЙ =_{dif} 'весел';

- 29) ВЕРНЫЙ =_{dif} 'верн' / 'верен' ^буква;
 30) КРАСИВЫЙ =_{dif} 'красив';
 31) степень_форма_число_одуш_род_падеж =_{dif} усл_A11 A11 / усл_A21 A21 /...;
 32) усл_A11 =_{dif} Н имя_A(A1(Н;
 33) A11 =_{dif} степень_A1# / A12;
 34) степень_A1 =_{dif} сравн_A1 / прев_A1 прев_число_род_падеж_A1;
 35) A12 =_{dif} Н основа_A1(Н форма;
 36) форма =_{dif} A13 / A14;
 37) A13 =_{dif} краткая_A1#;
 38) A14 =_{dif} полная_A1#;

...

Текст примера использует метазнаки и сокращения имен терминов: "усл", "одуш", "од", "неод", "кр", "полн", "ед", "мн" есть сокращения имен – "условие", "одушевленность", "одушевленное", "неодушевленное", "краткое", "полное", "единственное", "множественное" и т.п. В описании примера правила 1, 3, 8, 10, 16–18, 31–35 составляют прагматику анализа (часть из них 16, 17, 31, 32, 35 – контекстно зависимые), остальные – контекстно свободный базовый компонент. Базовый компонент классифицирует признаки составляющих частей речи.

В целом это контекстная грамматика, интерпретация которой дает парадигмальное разложение частей речи (существительных Noun, прилагательных Adj, союзов Conj, наречий ADV, частиц Part, предлогов Prep) и др., С-маркер которых в линейной скобочной записи имеет вид:

- Pup(тчк('.'))
- функциональные_категории(Conj('в_то_время_как'))
- функциональные_категории(ADV('довольно'))
- функциональные_категории(Part('ведь'))
- Prep('в_качестве')
- имя_N(N1(ГОРОД('город')) муж_N1(неод_N1(ед_N1(ед_дат_N1('у')))))
- имя_A(A1(основа_A1(КРАСИВЫЙ('красив')))) краткая_A1(кр_мн_A1('ы')).

Синтезированная последовательность С-маркеров слов предложения передается для последующего анализа вышестоящему уровню синтаксиса языка.

Морфосинтез реализует преобразование последовательности парадигмальных структур слов, синтезированной высшими уровнями синтаксиса, в последовательность идеальных знаков языка. Это преобразование выполняется как операция порождения над последовательностью парадигмальных структур слов с использованием базового компонента грамматики морфоанализа.

Высшие ярусы синтаксиса представлены системой теорий *фразовой структуры* [24], описывающих формальные преобразования между морфологической структурой и структурой смысла всякого естественногоязыкового предложения. Каждая из них представляется своим модулем со специфичными моделями разных языковых явлений. Ее элементами являются [11]:

Х-штрих-теория (X-Bar Theory), задающая (единую для врожденной языковой способности человека) рекурсивную структурную форму представления разных составляющих и их категорий.

Теория ограничиваения (Bounding Theory), ограничивающая возможности перемещения элементов предложения.

Теория управления (Government Theory), определяющая отношение между вершиной составляющей и ее зависимыми элементами.

Теория падежа (Case Theory), задающая принципы абстрактного (структурного) и морфологического падежного управления.

Тета-теория (Theta-Theory), определяющая тематические (= семантические) отношения между глаголом (или другим предикатом) и его аргументами (актантами).

Теория связывания (Binding Theory) задает принципы связи возвратных и взаимных местоимений (анафор), личных местоимений (прономиналов) и R-выражений (имен, прямо называющих предмет) с антецедентами (референтами).

Эти теории дополняют теорией уровня **функциональной формы** (FF), объясняющей эффекты скрэмблинга (свободного порядка слов) "стилистическим" перемещением фокуса [25].

Обширные исследования задач синтаксического анализа и порождения отражены в [26, 27].

Модуль логической формы занимает самый высший ярус синтаксиса языка и реализует модель двустороннего преобразования между структурой предложений дискурса и структурой логической формы выраженных ими высказываний, т.е. перевода между предложениями и высказываниями, выражающими их значение (смысл). Все низшие уровни языка осуществляют лишь подготовку для этих преобразований – "*... все в языке прямо ... или косвенно ... существует только для того, чтобы передавать... какой-то смысл, какую-то информацию*" [28, с.28].

Главная проблема синтеза адекватной логики здесь решена в пользу силлогистики: *"естественные языки не подчиняются принципам, с которыми имеет дело современная логика. Скорее, они соответствуют аристотелевской классической концепции о том, что предложение имеет субъект и предикат, где предикат может быть сложным: он может состоять из глагола и его прямого дополнения ... или из глагола и клаузуального компонента"* [3, с. 125]. *"... традиционная формальная логика остается пока наилучшим инструментом для анализа умозаключений в формах естественного языка"* [28, с.48]. Остальные проблемы перевода решаются синтезом понятия структуры высказываний, сопоставлением ее терминалам компонент грамматической структуры предложений и выяснением (для анализа) контекстных условий синтеза структур высказываний. Эти проблемы исследуют в формальной семантике, где достигнут определенный *"прогресс в анализе значения таких языковых выражений, как неопределенные выражения, анафоры, условные кванторы, модальные слова и др."* [29, с.212].

Важнейшим открытием формальной семантики является выявление факта трехчленной структуры элементарных высказываний, суждений: скрытый или явный оператор квантификации, рестриктивный терм (логическое подлежащее), ограничивающий объекты квантификации, и утверждаемый предикат (логическое сказуемое) в ядерной сфере действия [29, с.215-216]. В формальной записи на метаязыке [22] структура системы высказываний произвольной сложности имеет вид (рис. 7), где терминальными являются термины: \leftrightarrow (тождественности), \rightarrow (следования), \vee (дизъюнкции), $\&$ (конъюнкции), \neg (отрицания), предикатор, переменная, истина, \forall (квантор всеобщности), \exists (квантор существования) и др. кванторы. Им ставят в соответствие части синтаксической структуры предложений: *"материал глагольной группы VP соотносится с ядерной сферой действия, а материал NP – с рестриктивным термом"* [29, с.216]. Высказываниям с несколькими операторами ставят в соответствие предложения с несколькими именными группами и различают определенные и неопределенные (специфичные или

неспецифичные) именные группы. Все определенные (сильные) именные группы ("с определенным артиклем, с притяжательными и указательными местоименными прилагательными и с кванторами общности") представляют известные объекты, включенные в дискурс ранее (совпадают с ними), а в логической формуле являются рестриктивными терминами, связанными соответствующими кванторами. Неопределенные (слабые) именные группы вводят в дискурс новые объекты и для формирования логической формы высказывания о них переменные связывают *экзистенциальным закрытием*. При этом "предмет, на который указывает неспецифичное неопределенное выражение, не ... связан с ранее установленным объектом в дискурсе ..., а для специфичных неопределенных выражений достаточно, чтобы была установлена связь между ними и подмножеством предыдущих объектов" [29, с. 222].

```

система_высказываний =dif высказывание (высказывание);
высказывание =dif дизъюнкт следует дизъюнкт / дизъюнкт;
следует =dif  $\rightarrow$  /  $\leftrightarrow$  ;
дизъюнкт =dif конъюнкция ( $\vee$  конъюнкция);
конъюнкция =dif инверсия конъюнкт (& инверсия конъюнкт);
конъюнкт =dif оператор рестриктивный_терм (' дизъюнкт ') / оператор рестриктивный_терм
    ядерная_сфера_действия;
оператор =dif  $\forall$  /  $\exists$  / ...;
рестриктивный_терм =dif сложный_предикат;
ядерная_сфера_действия =dif сложный_предикат;
сложный_предикат =dif кон_предикат ( $\vee$  кон_предикат);
кон_предикат =dif инверсия сомн_пред (& инверсия сомн_пред);
сомн_пред =dif (' сложный_предикат ') / пред;
инверсия =dif  $\neg$  / истина;
пред =dif предикатор (' переменные ');
переменные =dif переменная (переменная);

```

Рис. 7. Структура системы высказываний

Эта же трехчленная логическая структура высказывания присуща и условным предложениям (с "если"): посылка образует рестриктивный терм, связанный "скрытым наречием универсальной квантификации, а заключение представляет собой ядерную сферу действия" [29, с. 227]. Эти и другие вопросы формальной семантики продолжают обсуждаться [28, 29].

6. Структура блока семантики естественного языка

Структура блока семантики естественного языка в целом совпадает со структурой блока семантики научной теории (рис. 2), отличаясь от нее дополнительным высшим уровнем, специфичным для умозаключений естественного языка.

Только традиционная логика придает необходимую формальность умозаключениям естественного языка, сохраняя присущую им содержательность отношения следования. Известны разные аксиоматизации силлогистики [1, 30–33]. Поэтому главная проблема не в том, чтобы предложить еще одну аксиоматизацию и/или систему вывода, а поиск на их множестве простейшей, поддерживающей модель вывода, близкую естественному мышлению людей.

Унификация суждений. В определении категорических суждений, кроме метапеременных a, b , есть пять логических терминов: "Всякое", "Ни одно", "Некоторое", "есть", "не есть" [34]:

"Aab означает (имеет смысл, есть по определению) Всякое a есть b ...,

Eab означает Ни одно a не есть b ...,

Iab означает Некоторое a есть b ...,

Oab означает *Некоторое a не есть b*".

Вполне очевидно, здесь избыток логических терминов, который и явился одной из двух причин разноречия переводов категорических суждений и высказываний (вторая - попытки решить логическими методами нелогическую проблему устранения пустых терминов [1, п.4.4; 30, §79]).

Сведем все разнообразие форм суждений к универсальным, используя превращения для вывода определений *Aab* и *Eab* и диагонали логического квадрата для *Iab* и *Oab*. В формальных символах "означает" (\equiv), "Всякое" (\forall), "не" ($\bar{}$), "есть" ("*a* есть терм *b*" - обозначим выделяющими скобками) и "равнозначно" (\equiv) получим переводы:

$$Aab = \forall a b(a) \equiv \forall a \bar{b}(\bar{a}) \text{ (Всякое } a \text{ есть } b \equiv \text{Ни одно } a \text{ не есть не } b),$$

$$Eab = \forall a \bar{b}(a) \equiv \forall a b(\bar{a}) \text{ (Всякое } a \text{ есть не } b \equiv \text{Ни одно } a \text{ не есть } b),$$

$$Iab = \bar{\forall a} \bar{b}(a) \equiv \bar{\forall a} b(\bar{a}) \text{ (Не всякое } a \text{ есть не } b \equiv \text{Не всякое } a \text{ не есть } b),$$

$$Oab = \bar{\forall a} b(a) \equiv \bar{\forall a} \bar{b}(\bar{a}) \text{ (Не всякое } a \text{ есть } b \equiv \text{Не всякое } a \text{ не есть не } b).$$

В этих формах квантор \forall (как это общепринято в логике) можно удалить, после чего все они получают окончательный вид в логике отрицаний - а "есть" ("не есть") терм *b*, отрицание (или нет) предиката *b* и отрицание всего суждения в *Iab* и *Oab*:

$$Aab = b(a) \equiv \bar{b}(\bar{a}),$$

$$Eab = \bar{b}(a) \equiv b(\bar{a}),$$

$$Iab = \bar{\bar{b}(a)} \equiv \bar{b}(\bar{a}),$$

$$Oab = \bar{b}(a) \equiv \bar{\bar{b}(\bar{a})},$$

где *a*, *b* – возможно, структурно сложные метаимена терма и предиката.

В естественном языке типов кванторов существенно больше двух, но все их можно считать ограниченными, сводимыми к двум ограничениям области изменения термов [35], и, следовательно, к одному – универсальному. Полученные результаты обобщим формальным определением структуры всякого высказывания (рис. 8), где термины *общее_имя*, *единичное_имя*, *присущ* (как вариант *есть*) и *связки* являются терминальными.

высказывание =_{def} суждения следует суждение / суждение;
суждения =_{def} суждение (& суждение);
суждение =_{def} структура присущ субъект;
субъект =_{def} единичное_имя;
структура =_{def} элемент (\vee элемент);
элемент =_{def} инверсия сомножитель (& инверсия сомножитель);
сомножитель =_{def} '(' структура ')' / общее_имя;

Рис. 8. Структура высказывания

Эта структура предельно проста, не зависит от естественного языка, не усваивается в процессе усвоения языка и является одной из основных *языковых универсалий*, присущих врожденному мышлению (возможно, не только людей).

Вывод в естественном языке. Хотя в художественных произведениях и выявлены умозаключения только трех-четырех модусов – *Barbara*, *Celarent* первой и *Camestres* (*Cesare*) второй фигуры [28], однако их нельзя считать охватывающими *все множество* правильных умозаключений, поскольку еще Аристотель выявил 14 сильных модусов первых трех

исследованных им фигур. Полная же система вывода в естественном языке должна включать все известные 19 сильных модусов четырех фигур.

В традиционной логике, как правило, используют аксиоматический метод, применявшийся "уже Аристотелем при построении своей теории силлогизма" в сочетании с исчислением высказываний: "утверждениями именно этого исчисления интуитивно пользуется Аристотель в доказательствах модусов силлогизма" [32, с.220, с.221]. Общеизвестны полнота и непротиворечивость системы аксиом из четырех сильных модусов первой фигуры:

$Amp \& Asm \rightarrow Asp = p(m) \& m(s) \rightarrow p(s)$ - Barbara,

$Emp \& Asm \rightarrow Esp = \neg p(m) \& m(s) \rightarrow \neg p(s)$ - Celarent,

$Amp \& Ism \rightarrow Isp = p(m) \& \neg(m(\neg s)) \rightarrow \neg(p(\neg s))$ - Darii,

$Emp \& Ism \rightarrow Osp = p(\neg m) \& \neg(\neg m(s)) \rightarrow \neg(p(s))$ - Ferio.

Эти аксиомы в качестве фундаментальных правил вывода обеспечивают дедуктивный вывод в произвольных рассуждениях, структура которых ограничена (рис. 8).

Сильные модусы остальных трех фигур сводятся к модусам первой фигуры при помощи использования перестановок посылок и обращений:

$Amp \rightarrow Ipm = p(m) \rightarrow \neg(m(\neg p)) \equiv p(m) \rightarrow \neg(\neg m(p))$,

$Emp \leftrightarrow Epm = \neg p(m) \leftrightarrow \neg m(p)$,

$Imp \leftrightarrow Ipm = \neg(p(\neg m)) \leftrightarrow \neg(m(\neg p))$.

Тем самым логика естественного языка рекурсивно сводит составные, сложные умозаключения к элементарным, но структурно сложным относительно прикладной теории.

7. Выводы

Обоснованные современными достижениями логико-гносеологических и лингвистических исследований результаты синтеза наиболее существенных признаков идеала структуры языков системы научных теорий обобщим в форме следующих утверждений:

– структура языков системы научных теорий в развернутой (в пространстве или во времени) форме содержит четыре блока (знаков, синтаксиса языков, семантик и функций научных теорий, взаимосвязанных между собой, с источниками-приемниками знаковых образов и с реальным миром), каждый из которых представляется иерархической сетью идентичных модулей, чья структура содержит словарь и интерпретаторы концептуальных и эмпирических моделей синтаксиса, семантики и прагматики некоторой подтеории;

– в свернутой в пространстве (но способной к развертыванию во времени) форме структура системы научных теорий представима структурой в составе памяти словаря, интерпретатора концептуальных моделей с памятьми, содержащими концептуальные модели разных компонент подтеорий, и интерпретатора разных эмпирических моделей;

– функционирование всякого модуля определяется решением задач анализа и синтеза (порождения) структуры некоторого представления информации. Всякая задача всякой подтеории научной теории представляется итерационно-рекурсивной структурой.

СПИСОК ЛИТЕРАТУРЫ

1. Смирнов В.А. Логические методы анализа научного знания. – М.: Наука, 1987. – 256с.
2. Куайн У.В.О. Слово и объект // Новое в зарубежной лингвистике. – М.: Прогресс, 1986. – Вып. 18: Логический анализ естественного языка. – С. 24–98.
3. Хомский Н. Язык и проблемы знания // Вестник Московского университета. Сер. 9. Филология. – 1995. – № 6. – С. 110–134.
4. Хомский Н. Язык и проблемы знания // Вестник Московского университета. Сер. 9. Филология. – 1996. – № 6. – С. 157–185.
5. Хомський Н. Роздуми про мову: Пер. з англ. – Львів: Ініціатива, 2000. – 352 с.
6. Аврутин С. Усвоение языка // Фундаментальные направления современной американской лингвистики / Под ред. А.А. Кибрика, И.М. Кобозевой и И.А. Секериной. – М.: Изд-во МГУ, 1997. – С. 261–274.
7. Джакендофф Р. Модель языковой способности: неоминималистская перспектива // Вестник Московского университета. Сер. 9. Филология. – 2002. – № 2. – С. 157–189.
8. Хомский Н. Объяснительные модели в лингвистике // Математическая логика и ее применения: Пер. с англ. – М.: Мир, 1965. – С. 245–272.
9. Хомский Н. Синтаксические структуры // Новое в лингвистике. – М.: Ин. лит., 1962. – Вып. 2. – С. 412–527.
10. Хомский Н. Аспекты теории синтаксиса: Пер. с англ. – М.: Изд-во МГУ, 1972. – 260 с.
11. Бейлин Дж. Краткая история генеративной грамматики // Фундаментальные направления современной американской лингвистики / Под ред. А.А. Кибрика, И.М. Кобозевой и И.А. Секериной. – М.: Изд-во МГУ, 1997. – С. 13–57.
12. Гивон Т. Система обработки визуальной информации как ступень в эволюции человеческого языка // Вестник Московского университета. Сер. 9. Филология. – 2004. – № 3. – С. 117–169.
13. Зубрицкая Е. Фонология // Фундаментальные направления современной американской лингвистики / Под ред. А.А. Кибрика, И.М. Кобозевой и И.А. Секериной. – М.: Изд-во МГУ, 1997. – С.168–206.
14. Стерлинг Л., Шапиро Э. Искусство программирования на языке ПРОЛОГ: Пер. с англ. – М.: Мир, 1990. – 235 с.
15. Кулагина О.С. Исследования по машинному переводу. – М.: Наука, 1979. – 320 с.
16. Лингвистический процессор для сложных информационных систем / Ю.Д. Апресян, И.М. Богуславский, Л.Л. Иомдин и др. – М.: Наука, 1992. – 256 с.
17. Мельчук И.А. Морфологический анализ при машинном переводе (преимущественно на материале русского языка) // Проблемы кибернетики. – 1961. – Вып. 6. – С. 207–276.
18. Иорданская Л.Н. Морфологические типы основ русского языка (для различения омонимии морфем в процессе анализа при машинном переводе) // Проблемы кибернетики. – 1961. – Вып. 6. – С. 281–287.
19. Анно Е.И. Система морфологического анализа с синтезом словоформ // Семиотика и информатика. – М.: ВИНТИ, 1978. – Вып.10. – С. 168–187.
20. Браун Д. "Сетевая морфология" и русская глагольная система // Вестник Московского университета. Сер. 9. Филология. – 1995. – № 6. – С. 91–109.
21. Хомский Н. Язык и проблемы знания // Вестник Московского университета. Сер. 9. Филология. – 1996. – № 4. – С. 133–162.
22. Кургаев А.Ф. Метаязык представления знаний // УСИМ. – 1998. – № 4. – С. 79–86.
23. Русский язык: справочник школьника / Под ред. В. Славкина. – М.: Филол. общ. "Слово", 1995. – 512 с.
24. Хомский Н. Язык и проблемы знания // Вестник Московского университета. Сер. 9. Филология. – 1996. – № 2. – С. 103–121.
25. Кондрашова Н. Генеративная грамматика и проблема свободного порядка слов // Фундаментальные направления современной американской лингвистики / Под ред. А.А. Кибрика, И.М. Кобозевой и И.А. Секериной. – М.: Изд-во МГУ, 1997. – С.110–141.
26. Секерина И.А. Американские теории синтаксического анализа предложения в процессе понимания // Вопросы языкознания. – 1996. – № 3. – С. 100–138.
27. Секерина И.А. Психолингвистика // Фундаментальные направления современной американской лингвистики / Под ред. А.А. Кибрика, И.М. Кобозевой и И.А. Секериной. – М.: Изд-во МГУ, 1997. – С.231–260.
28. Кривоносов А.Т. Естественный язык и логика. – Москва – Нью-Йорк: Wickerrsham Printing Company, Inc. 1993. – 318 с.
29. Изворска Р. Формальная семантика // Фундаментальные направления современной американской лингвистики / Под ред. А.А. Кибрика, И.М. Кобозевой и И.А. Секериной. – М.: Изд-во МГУ, 1997. – С.207–230.
30. Войшвилло Е.К., Дегтярев М.Г. Логика как часть теории познания и научной методологии (фундаментальный курс). – М.: Наука, 1994. – Кн. II. – 332с.
31. Логико-философские труды В.А. Смирнова / Под ред. В.И. Шалака. – М.: Эдиториал УРСС, 2001. – 592с.
32. Лукасевич Я. В защиту логики // Философия и логика Львовско-Варшавской школы: Пер. с польск. – М.: РОССПЭН, 1999. – С. 219–232.
33. Субботин А.Л. Теория силлогистики в современной формальной логике. – М.: Наука, 1965. – 124с.
34. Лукасевич Я. Аристотелевская силлогистика с точки зрения современной формальной логики: Пер. с англ. – М.: ИЛ, 1959. – 290 с.
35. Слупецкий Е., Борковский Л. Элементы математической логики и теория множеств: Пер. с польск. – М.: Прогресс, 1965. – 368 с.