

В.А. Фомичёв

**Математические
основы представления
содержания посланий
компьютерных
интеллектуальных
агентов**



В.А. Фомичёв

**МАТЕМАТИЧЕСКИЕ
ОСНОВЫ ПРЕДСТАВЛЕНИЯ
СОДЕРЖАНИЯ ПОСЛАНИЙ
КОМПЬЮТЕРНЫХ
ИНТЕЛЛЕКТУАЛЬНЫХ
АГЕНТОВ**

Москва
2007

СОДЕРЖАНИЕ

Введение	7
Глава 1. Проблема представления содержания посланий компьютерных интеллектуальных агентов и формализация семантики естественного языка	14
1.1. Проблема разработки языков общения компьютерных интеллектуальных агентов	14
1.2. Значение формальных методов представления содержания естественно-языковых текстов для разработки лингвистических процессоров ...	16
1.3. Подходы к формализации семантики естественного языка, разработанные в конце 1960-х – первой половине 1980-х гг.	20
1.4. Роль формальных систем семантических представлений с большими выразительными возможностями в проектировании лингвистических процессоров	24
Глава 2. Математическая модель для описания системы первичных единиц концептуального уровня, используемых многоагентной системой	30
2.1. Постановка задачи	30
2.2. Базовые обозначения и вспомогательные определения	33
2.2.1. Общематематические обозначения	33
2.2.2. Предварительные определения и обозначения из теории формальных грамматик и языков	33
2.2.3. Используемые определения из теории алгебраических систем ..	34
2.3. Краткая характеристика предлагаемой математической модели для описания системы первичных единиц концептуального уровня, используемых многоагентной системой	35
2.4. Основные идеи определения класса сортов систем	37
2.5. Формальное определение сортовой системы	39
2.6. Типы, порождаемые сортавыми системами, и конкретизации типов ..	40
2.6.1. Определение множества типов	40
2.6.2. Интерпретация определения множества типов	42
2.6.3. Отношение конкретизации на множестве типов	44

2.7. Концептуально-объектные системы	48
2.8. Системы кванторов и логических связей. Концептуальные базисы ...	50
2.9. Обсуждение разработанной математической модели для описания системы первичных единиц концептуального уровня	54
2.9.1. Особенности модели с математической точки зрения	54
2.9.2. Сравнение модели с другими подходами к описанию первич- ных единиц концептуального уровня	56

Глава 3. Математическая модель для описания структурированных значений предложений и связанных текстов на естественном языке

3.1. Постановка задачи	59
3.2. Краткая характеристика предлагаемого решения поставленной задачи	62
3.2.1. Краткая характеристика новых правил построения формул	62
3.2.2. Схема определения трех классов формул, порождаемых кон- цептуальными базисами	66
3.3. Использование интенциональных кванторов в формулах	68
3.4. Использование реляционных символов и разметка формул	72
3.4.1. Правила для применения реляционных символов	72
3.4.2. Правило, позволяющее пометать формулы	75
3.5. Использование логических связей «не», «и», «или»	76
3.6. Построение составных обозначений понятий и объектов	78
3.6.1. Правило для построения составных обозначений понятий	78
3.6.2. Построение составных обозначений объектов	79
3.7. Использование в формулах кванторов существования и всеобщ- ности. Построение обозначений упорядоченных наборов	81
3.7.1. Применение кванторов существования и всеобщности	81
3.7.2. Построение обозначений упорядоченных наборов	82
3.7.3. Сводная таблица правил $P[0]$ – $P[10]$	83
3.8. СК-языки. Математическое исследование их свойств	84
3.9. Обсуждение построенной математической модели	88

Глава 4. Исследование выразительных возможностей СК-языков

4.1. Удобный способ описания событий	92
4.2. Формализация предположений о структуре семантических пред- ставлений множеств	94
4.3. Построение семантических представлений вопросов с ролевыми вопросительными словами	96
4.4. Семантические представления вопросов о количестве предметов и о количестве событий	98

4.5. Семантические представления вопросов с формами вопросительно-относительного местоимения «какой»	99
4.6. Построение семантических представлений вопросов с ответом «Да»/«Нет»	99
4.7. Отображение смысловой структуры команд	100
4.8. Представление теоретико-множественных отношений и операций на множествах	101
4.9. Представление смысла фраз с придаточными предложениями цели и с косвенной речью	102

Глава 5. Исследование возможностей применения аппарата

СК-языков к построению онтологий	103
5.1. Онтологии и их значение для глобальных информационных сетей ..	103
5.2. Анализ некоторых возможностей представления знаний о предметных областях средствами СК-языков	105
5.2.1. Представление содержания фрагментов знаний о мире	106
5.2.2. Объектно-ориентированные представления фрагментов знаний	107
5.3. Моделирование выразительных механизмов основных языков дескриптивной логики	108
5.4. Построение сложноорганизованных фрагментов знаний о мире	109

Глава 6. Исследование выразительных возможностей основных подходов к представлению содержания посланий компьютерных интеллектуальных агентов

6.1. Использование СК-языков для представления содержания посланий, соответствующих отдельным фразам естественного языка	113
6.2. Явное представление причинно-следственных отношений, передаваемых дискурсами	116
6.3. Семантические представления дискурсов со ссылками на смысл фраз и более крупных частей текста	117
6.4. Проблема формирования контрактов и протоколов переговоров в области электронной коммерции	118
6.5. Анализ возможностей использования СК-языков для формирования контрактов и протоколов переговоров в области электронной коммерции	121
6.6. Краткие сведения о бесконтекстных грамматиках	125
6.6.1. Определения и примеры	125
6.6.2. Бесконтекстные грамматики в форме Бэкуса–Наура	128
6.7. Синтаксис языка представления содержания посланий компьютерных интеллектуальных агентов FIPA SL Content Language	129

6.8. Моделирование выразительных механизмов языка представления содержания посланий FIPA SL Content Language средствами СК-языков	132
6.9. Преимущества СК-языков по сравнению с языком FIPA SL Content Language	135
6.10. Сравнение выразительных возможностей СК-языков с возможностями основных известных подходов к формальному представлению содержания ЕЯ-текстов	136
6.10.1. Сравнение с основными подходами, разработанными в нашей стране	136
6.10.2. Сравнение с основными зарубежными подходами	139
Литература	143
Приложение. Доказательства Лемм 1, 2 и Утверждения 3.5 из главы 3 ...	165

ВВЕДЕНИЕ

И все мосты языков минуя,
что между несхожим встали,
нас поднимают светлые струны
над безраздельностью дали.

Райнер Мария Рильке

С начала 1990-х гг. интенсивно развивается область исследований, направленных на создание многоагентных систем (МАС). Следуя работе (Wooldridge, Jennings, 1995), под компьютерным интеллектуальным агентом (КИА) будем понимать аппаратную или программную сущность, способную действовать в интересах достижения целей, поставленных владельцем и/или пользователем. МАС является сетью КИА, совместно решающих такие задачи, которые выходят за рамки индивидуальных способностей или индивидуальной базы данных любого из агентов сети.

Таким образом, МАС должна обладать следующими свойствами:

- способностей любого агента системы недостаточно для решения задачи в одиночку,
- в МАС нет глобальной системы управления,
- данные децентрализованны,
- вычисления асинхронны.

В научной литературе наиболее часто формулируются следующие свойства, которыми должны обладать КИА, входящие в многоагентную систему (Wooldridge, Jennings, 1995; FIPA, 1998a):

- автономность (autonomy), определяемая как возможность действовать без вмешательства своего владельца, а также контролировать свое внутреннее состояние и свои действия;
- социальное поведение (social behaviour), т.е. возможность осуществлять взаимодействие с другими агентами;
- реактивность (reactivity), определяемая как адекватное восприятие среды и наличие соответствующих реакций на ее изменение;
- активность (pro-activity), определяемая как способность порождать цели и действовать рациональным образом для их достижения;

– наличие базовых знаний (basic knowledge), т.е. знаний агента о себе и об окружающей среде, в том числе о других агентах, которые не меняются в течение жизненного цикла агента;

– наличие убеждений (beliefs), т.е. той части базовых знаний, которая может со временем меняться;

– наличие постоянных целей (permanent goals), т.е. конечного множества состояний, на достижение которых ориентировано поведение агента в целом;

– наличие намерений (intentions), т.е. конечного множества состояний, на достижение которых ориентировано текущее поведение агента и которые согласуются с постоянными целями;

– наличие обязательств (commitments), т.е. задач, выполнение которых берет на себя агент по просьбе или поручению других агентов.

Взаимодействие компьютерных интеллектуальных агентов осуществляется через обмен посланиями, которые могут выражать сообщения, вопросы и команды. Для формирования таких посланий разрабатываются специальные языки общения интеллектуальных агентов (Agent Communication Languages, или ACL). Для координации деятельности исследовательских центров разных стран по разработке стандартных инструментальных средств в области МАС в 1996 г. образован международный Фонд интеллектуальных физических агентов (The Foundation for Intelligent Physical Agents, или FIPA), штаб-квартира которого находится в Женеве. В 1997–2000 гг. в рамках этого фонда был разработан стандарт языка общения КИА, который в дальнейшем будет называться FIPA ACL. Фондом поставлена задача разработки библиотеки языков представления содержания (в отличие от внешней информации – об отправителе, получателе и т.д.) посланий КИА (Content Languages), совместимых с этим языком и охватывающих весь спектр применений МАС. В 2002 г. в рамках фонда был разработан язык представления содержания посланий КИА, получивший название FIPA SL Content Language (FIPA SL CL, 2002).

В 2005 г. FIPA был официально принят Компьютерным обществом Института инженеров по электричеству и электронике (IEEE Computer Society) в качестве 11-го комитета по стандартам этой авторитетной международной организации. Данное событие говорит о возникновении необходимости вписывания стандартов МАС в более широкий контекст компьютерной науки в целом.

Многоагентные системы рассматриваются как ключевая технология для реализации электронной коммерции. Следовательно, выразительные возможности языка общения КИА должны быть достаточ-

ными для того, чтобы представить произвольных коммерческих переговоров и контрактов, заключенных в результате этих переговоров. Поэтому формальные языки для представления содержания коммерческих переговоров и контрактов являются предметами исследования в новых научных направлениях в области МАС, называемых электронными переговорами (e-negotiations) и электронным заключением контрактов (electronic contracting).

Между тем, как показано в заключительной главе этой книги, выразительные возможности языка представления содержания посланий КИА FIPA SL Content Language довольно далеки от того, чтобы быть удобными для решения этой задачи. Поэтому актуальна задача создания методов разработки более совершенных формальных языков – таких, которые были бы удобны для представления содержания любых посланий КИА, в том числе и для представления содержания произвольных коммерческих переговоров и контрактов.

Основная часть информации хранится и передается людьми с помощью естественного языка (ЕЯ), т.е. совокупности русского, английского, японского и других языков. Один из главных подклассов компьютерных систем с элементами искусственного интеллекта, или интеллектуальных систем (ИС), составляют программы, понимающие ЕЯ или синтезирующие выражения ЕЯ по некоторым внутренним представлениям. Такие программы называются системами обработки естественного языка (в англоязычной научной литературе: natural language processing systems), или лингвистическими процессорами (ЛП).

Анализ научной литературы показывает, что проблема создания языков представления содержания произвольных посланий КИА тесно взаимосвязана с проблемой разработки широко применимых методов формального описания содержания (смысла) предложений и дискурсов, обрабатываемых семантико-синтаксическими анализаторами лингвистических процессоров. Рассмотрим несколько неформальных понятий, являющихся базовыми для описания этих взаимосвязей.

Под семантикой ЕЯ будем понимать совокупность закономерностей передачи информации средствами ЕЯ. Связным текстом (или дискурсом) называется последовательность взаимосвязанных по смыслу выражений на ЕЯ.

Если T – некоторое выражение на ЕЯ (словосочетание, предложение, дискурс), то структурированным значением выражения T является информационная структура, строящаяся мозгом человека, владеющего данным подязыком ЕЯ (русским, английским или другим), независимо от контекста, в котором услышано или прочитано выраже-

ние Т, т.е. строящаяся на основе только знаний о значениях элементарных лексических единиц и правил их комбинирования в данном языке.

Под семантическим представлением (СП) ЕЯ-выражения Т понимается формальная структура, являющаяся либо образом структурированного значения этого выражения, либо отражением смысла (или содержания) данного выражения в определенном контексте – в конкретной ситуации диалога, в контексте знаний о мире или в контексте предшествующей части дискурса.

Таким образом, СП ЕЯ-выражения Т является формальной структурой, первичные элементы которой – это, в частности, обозначения понятий, конкретных объектов, множеств объектов, событий, имена функций и отношений, логические связки, обозначения чисел и цветов, а также обозначения смысловых отношений между значениями фрагментов текста или между объектами рассматриваемой предметной области. СП текстов могут являться, например, строками и размеченными ориентированными графами (семантическими сетями).

Алгоритм семантико-синтаксического анализа строит по тексту на ЕЯ его СП, используя для этого знания о морфологии и синтаксисе подязыка ЕЯ (русского, английского и др.), информацию о взаимосвязях лексических единиц с единицами семантического уровня и знания о мире. СП текста, построенное таким алгоритмом, интерпретируется прикладной интеллектуальной системой в зависимости от ее назначения, например, как задание на поиск ответа на вопрос, команда на выполнение физического действия автономным интеллектуальным роботом, фрагмент знаний о мире, предназначенный для пополнения базы знаний и т.д.

Благодаря бурному прогрессу компьютерной сети Всемирная Паутина (the World Wide Web, WWW, W3) пользователи сети во всем мире получили быстрый доступ к огромному количеству ЕЯ-текстов, относящихся к различным областям деятельности человека. С середины 1990-х гг. специалисты в самых разных предметных областях работают не только с публикациями и базами данных (БД) своих организаций, но и стремятся использовать информационные ресурсы Паутины. Поэтому чрезвычайно актуальна задача организации взаимодействия на ограниченном естественном языке из различных предметных областей с огромным объемом накопленных информационных ресурсов Всемирной Паутины.

ЕЯ-интерфейсы для взаимодействия с информационными ресурсами Паутины необходимы не только специалистам для решения профессиональных задач, но и конечным пользователям, перед которыми

стоят задачи получения медицинской или юридической информации, расширения культурного кругозора, получения дополнительного профессионального образования и т.д.

В феврале 2001 г. консорциум сети Всемирная Паутина, обозначаемый в большинстве документов сокращением W3C (the World Wide Web Consortium), официально объявил о широком развертывании исследований по преобразованию существующей сети в Семантическую Всемирную Паутину (Semantic Web). Один из наиболее важных аспектов реализации этого крупномасштабного проекта заключается в том, что компьютерные интеллектуальные агенты (КИА) смогут анализировать информацию, представленную на веб-сайтах, взаимодействуя между собой. Часть КИА сможет выполнять смысловой анализ естественно-языковых компонентов электронных документов, представленных в веб-сайтах. Это даст возможность конечным пользователям осуществлять поиск информации в Паутине не по ключевым словам, а по смыслу, с помощью КИА (Semantic Web, 2001; Java, Finin, Nirenburg, 2006).

Важные дополнительные возможности для пользователя предоставят речевые браузеры: они позволят использовать телефоны (в том числе мобильные) для взаимодействия с Семантической Паутиной на ЕЯ (Voice, 2001).

В данной монографии предлагается оригинальный, широко применимый (по гипотезе автора – универсальный) математический подход к представлению содержания посланий компьютерных интеллектуальных агентов в многоагентных системах, в частности, в системах электронной коммерции. Выразительные возможности определяемого в книге класса СК-языков (стандартных концептуальных языков) превышают возможности основных известных подходов к этой проблеме, в том числе возможности языка представления содержания посланий КИА, разработанного в рамках международного Фонда интеллектуальных физических агентов (FIPA). Обоснованы возможности использования аппарата СК-языков в проектировании лингвистических процессоров (компьютерных систем, осуществляющих смысловую обработку письменных текстов или устной речи на естественном языке) в произвольных предметных областях в качестве инструмента математического описания смысловой структуры не только предложений, но и сложных связных текстов (или дискурсов), относящихся к деловой прозе: текстов по медицине, технике, экономике, юриспруденции и т.д.

Результаты данной монографии дают не только продвижение вперед, но и *качественный скачок* в области разработки формальных

средств и методов проектирования информационного обеспечения многоагентных систем, в том числе систем электронной коммерции. Этот качественный скачок обусловлен тем, что разработчики информационного обеспечения многоагентных систем получили систему правил (причем компактную, состоящую всего из 10 основных правил), позволяющих, по гипотезе автора, отражать содержание, т.е. строить семантические представления, произвольных посланий компьютерных интеллектуальных агентов.

СОДЕРЖАНИЕ КНИГИ

В главе 1 анализируются взаимосвязи проблемы разработки языков представления содержания посланий компьютерных интеллектуальных агентов и проблематики проектирования семантически-ориентированных лингвистических процессоров.

В главе 2 описывается математическая модель, перечисляющая первичные единицы концептуального уровня, используемые многоагентной системой, а также описывающая информацию, связанную с такими единицами и необходимую для соединения этих единиц в составные единицы, отображающие структурированные значения (СЗ) сколь угодно сложных ЕЯ-текстов.

В главе 3 (в развитие результатов главы 2) построена математическая модель для описания СЗ предложений и сложных связных текстов (дискурсов) на естественном языке (в частности, на русском, английском, немецком, французском языках). Модель представляет собою определение нового класса формальных языков, названных стандартными концептуальными языками (стандартными К-языками, СК-языками), и может рассматриваться как формальная грамматика нового вида. Сущность этой модели в том, что она задает 10 операций на концептуальных структурах, с помощью которых за конечное число шагов можно построить семантическое представление предложения или дискурса из чрезвычайно широкого подязыка деловой прозы.

Проведено математическое исследование формальных объектов, задаваемых этой моделью, – выражений СК-языков. В частности, доказана однозначность структурного анализа таких выражений.

Глава 4 посвящена исследованию выразительных возможностей класса СК-языков с позиций описания смысловой структуры отдельных предложений естественного языка. Показано, что выражения СК-языков удобно использовать для построения СП предложений, выражающих высказывания, вопросы, команды, распоряжения, советы, включая предложения со сложными описаниями множеств, однород-

ными членами предложения, причастными оборотами, придаточными цели и придаточными определительными предложениями.

Содержание главы 5 говорит о том, что аппарат СК-языков предоставляет широкий спектр новых возможностей для построения онтологий предметных областей, в том числе для построения формальных определений понятий и объектно-ориентированных модулей знаний.

В главе 6 исследуются возможности использования аппарата СК-языков для разработки языков представления содержания посланий компьютерных интеллектуальных агентов, в частности, языков, предназначенных для формирования контрактов и протоколов переговоров в области электронной коммерции. Показано, что выразительная сила класса СК-языков превышает выразительную силу языка представления содержания посланий КИА FIPA SL Content Language, разработанного в рамках международного Фонда интеллектуальных физических агентов. Проведено сравнение выразительных возможностей СК-языков с выразительными возможностями других, наиболее часто используемых подходов к формальному представлению значений (смысловой структуры) ЕЯ-текстов: теории представления дискурсов, теории концептуальных графов, эпизодической логики, теории расширенных семантических сетей, теории неоднородных семантических сетей и компьютерной семантики русского языка. Показано, что выразительные возможности СК-языков значительно превосходят возможности перечисленных подходов и, в то же время, аппарат СК-языков позволяет моделировать механизмы представления информации, характерные для каждого из указанных подходов.

Приложение содержит доказательства двух лемм и базирующегося на них доказательства одного из утверждений из главы 3.

Данная монография представляет собою новый шаг в развитии разработанной автором теории К-представлений (Фомичев, 1981а–2007; Fomitchov, 1984; Fomichov, 1992–2007).

Материалы монографии используются в курсе «Агентные системы в Интернет» для магистерской программы по направлению «Электронный бизнес» Государственного университета – Высшей школы экономики (факультет бизнес-информатики) и курсе «Математическая лингвистика» для студентов кафедры «Информационные технологии» «МАТИ» – Российского государственного технологического университета им. К.Э. Циолковского.

Я признателен редактору книги Е.Л. Наумовой за чрезвычайно тщательную работу с рукописью книги.

Глава 1

ПРОБЛЕМА ПРЕДСТАВЛЕНИЯ СОДЕРЖАНИЯ ПОСЛАНИЙ КОМПЬЮТЕРНЫХ ИНТЕЛЛЕКТУАЛЬНЫХ АГЕНТОВ И ФОРМАЛИЗАЦИЯ СЕМАНТИКИ ЕСТЕСТВЕННОГО ЯЗЫКА

1.1. Проблема разработки языков общения компьютерных интеллектуальных агентов

Прогресс исследований в области искусственного интеллекта (ИИ) и компьютерных сетей привел к появлению теории многоагентных систем (МАС). Многоагентные системы являются одним из наиболее быстро развивавшихся в 1990-е гг. и первой половине 2000-х гг. направлений информатики.

Главная причина постоянно растущего интереса к этому направлению заключается в следующем. В настоящее время можно прогнозировать бурное развитие в ближайшие годы *электронной коммерции* (electronic commerce, E-commerce), базирующейся на широчайших возможностях сети Интернет (Thome и Schinzer, 1998); МАС рассматриваются как ключевая технология для конструирования систем электронной коммерции (Guilfoyle, Jeffcoate, Stark, 1997; Wooldridge, 1998).

В частности, одной из перспективных областей использования МАС является индустрия туристического сервиса. Задача проектирования компьютерных интеллектуальных агентов (КИА) для этой индустрии является весьма естественной, поскольку различные КИА многих поставщиков услуг (заказ авиабилетов и железнодорожных билетов, резервирование комнат в гостиницах, взятие напрокат автомобилей, организация культурной программы и т.д.) должны динамически обнаруживать друг друга и эффективно взаимодействовать для решения стоящих перед ними задач.

Многочисленные КИА (сконструированные разными научно-исследовательскими центрами, использующие различную аппаратуру

и программное обеспечение) смогут эффективно взаимодействовать в процессе решения своих задач только в том случае, когда эти КИА будут располагать общим языком для обмена информацией и руководствоваться едиными правилами общения. Поэтому в 1990-е гг. было разработано несколько языков общения компьютерных интеллектуальных агентов; два из них являются наиболее широко применимыми.

Во-первых, это язык KQML, разработанный в США в рамках проекта разделения знаний, осуществлявшегося национальным агентством ДАРПА (DARPA). Исследования по разработке этого языка отражены, в частности, в публикациях (Finin et al., 1993; Labrou, 1996; Finin, Labrou, Mayfield, 1997; Labrou, Finin, 1997, 1998). Значительную роль в этих исследованиях сыграли специалисты Стэнфордского университета.

Второй язык разработан в рамках международного Фонда интеллектуальных физических агентов (the Foundation for Intelligent Physical Agents, или FIPA), штаб-квартира которого находится в Женеве. Одним из важных результатов исследований, организованных этим фондом, стала разработка в 1997–2000 гг. стандарта для представления посланий (messages) КИА, получившего название Языка общения агентов (FIPA Agent Communication Language = FIPA ACL) (FIPA, 1998a). Теоретической основой для создания этого языка послужили принципы разработки языка KQML и языка KIF, или Knowledge Interchange Format (Genesereth, Fikes et al., 1992; Genesereth, 1999). Для использования в сочетании с языком FIPA ACL разработан язык представления содержания посланий компьютерных интеллектуальных агентов FIPA SL Content Language (FIPA SL CL, 2002).

Проблема создания адекватных логико-информационных основ электронной коммерции предъявляет высокие требования к языку представления содержания посланий КИА. Этот язык должен позволять отображать содержание коммерческих переговоров. Однако язык FIPA SL Content Language обладает многими ограничениями в этом отношении. В связи с этим возникает проблема разработки математического описания такого класса языков, который был бы удобен для отображения содержания произвольных посланий КИА.

В нашей стране в последнее десятилетие проблематике много-агентных систем уделялось значительное внимание многими учеными. В частности, различные аспекты теории МАС и вопросы применения этой теории исследовались в работах В.И. Городецкого (Городецкий, 1998), В.В. Емельянова (Emelyanov, 2001), Э.С. Клы-

шинского (Клышинский, 1999), Г.С. Плесневича и В.Б. Тарасова (Тарасов, 1998; Plesniewicz, Tarassov, 2001), Д.А. Поспелова (Поспелов, 1998), Г.В. Рыбиной и В.Ю. Берзина (Рыбина, Берзин, 2002).

Однако следует отметить, что проблематика разработки формальных языков с широкими выразительными возможностями для представления содержания посланий КИА не получила в трудах ученых нашей страны такого же внимания, как и в серии зарубежных проектов, выполненных в рамках международного Фонда интеллектуальных физических агентов.

1.2. Значение формальных методов представления содержания естественно-языковых текстов для разработки лингвистических процессоров

Во введении к данной книге было показано, что проблема разработки языков представления содержания посланий компьютерных интеллектуальных агентов тесно взаимосвязана с проблематикой компьютерной лингвистики. В связи с этим проанализируем значение методов построения семантических представлений естественно-языковых текстов (ЕЯ-текстов) для конструирования лингвистических процессоров (ЛП).

Накопленный опыт исследований по созданию ЛП показал, что огромное влияние на проектирование анализаторов ЕЯ-текстов оказывают используемые методы формального отображения содержания (или смысла) текстов, а также методы формального представления промежуточных результатов смыслового анализа текстов. Особую актуальность в 1990–2000-е гг. приобрела проблема формального представления содержания связных текстов (или дискурсов).

Во-первых, основной объем информации в текстовых базах данных (БД) и сети Интернет представлен дискурсами. Во-вторых, сформулированная Э.В. Поповым современная концепция разработки систем общения с БД на ограниченном естественном языке (ОЕЯ) предполагает, что на вход системы поступают не только предложения, но и дискурсы (Попов, 2002). В-третьих, можно согласиться с высказанной Э.В. Поповым гипотезой о том, что повышению эффективности общения на ОЕЯ с большими БД будет способствовать реализация таких систем общения, когда активную роль в диалоге будет играть не только конечный пользователь, но и компьютер, располагающий моделью базы знаний, причем инициатива будет на протяжении диалога неоднократно переходить от одного участника общения к другому.

Последовательность выражений на ОЕЯ (с указанием авторов выражений), сформированных участниками общения, образует дискурс.

Можно выделить несколько наиболее важных аспектов проблемы формального представления содержания (или смысла) ЕЯ-текстов в компьютерных системах.

Идея использования в системах машинного перевода искусственного языка-посредника для представления смысла ЕЯ-текстов была высказана еще в 1960-м г. А.К. Жолковским, Н.Н. Леонтьевой и Ю.С. Мартыняновым. В 1960–1970-е гг. эта идея получила значительное развитие в работах А.К. Жолковского и И.А. Мельчука по лингвистической модели «Смысл – Текст» (Жолковский, Мельчук, 1969; Мельчук, 1974). В 1970-е гг. усилению внимания к идее семантического языка-посредника способствовала теория смысловой зависимости в ЕЯ Р. Шенка, нашедшая применение в нескольких экспериментальных системах компьютерной обработки ЕЯ (Schank, 1972; Schank et al., 1975).

Использование языка-посредника для представления содержания (смысла) ЕЯ-текстов позволяет перейти от неформализованного объекта, каким является ЕЯ-текст, к формальной структуре, что открывает возможности обработки этой структуры различными процедурами – «семантическими экспертами» в рамках базы знаний, представленных записями на формальном языке (языке представления знаний). На протяжении 1980–2000-х гг. в проектировании ЛП наиболее часто использовались языки-посредники, предоставляемые теорией семантических сетей, теорией фреймов, теорией концептуальных графов и эпизодической логикой. В нашей стране использовались также язык-посредник, разработанный в рамках компьютерной семантики русского языка, расширенные семантические сети, неоднородные семантические сети (см. раздел 6.10), СК-языки (Fomichov, Chuykov, 2000; Fomichov, Kochanov, 2001; Fomichov, Lustig, 2001; Люстиг, Фомичев, 2002; Fomichov, Akhromov, 2003; Fomichova, Fomichov, 2004).

В середине 1990-х гг. возникла новая проблема, усилившая внимание исследователей к проблеме разработки языка-посредника для отображения содержания ЕЯ-текстов. С целью устранения языкового барьера между пользователями сети Интернет из разных стран мира в монографии (Uchida, Zhu, Della Senta, 1999) был предложен новый язык-посредник, использующий слова английского языка для обозначения информационных единиц и несколько специальных символов. Этот язык, названный универсальным сетевым языком (UNL, the Universal Networking Language), базируется на идее отображения содержания фраз с помощью бинарных отношений. С конца 1990-х гг. ООН финансиру-

таким комплексный проект, направленный на разработку системы ЛП, преобразующих фразы на различных естественных языках в выражения языка UNL, а также преобразующих выражения языка UNL в предложения на различных естественных языках. Координатором проекта являлся Институт передовых исследований ООН Токийского университета. В проекте разрабатывались ЛП для шести официальных языков ООН и многих других языков. В течение нескольких последних лет исследования в этом направлении координируются Фондом универсального сетевого цифрового языка (Uchida, Zhu, Della Senta, 1999; Uchida, Zhu, 2001; Zhu, Uchida, 2002; UNL, 2005, 2006).

Проблема разработки формальных языков-посредников для отображения содержания (или смысла) ЕЯ-текстов (другими словами, языков семантических представлений, или семантических языков) исследуется специалистами разных стран в течение более трех десятилетий. В нашей стране ряд аспектов этой проблемы в различные периоды изучался Ю.Д. Апресяном, И.М. Богуславским, В.М. Брябриным, Б.Ю. Городецким, А.К. Жолковским, А.П. Ершовым, Ю.И. Клыковым, О.С. Кулагиной, Е.С. Кузиным, Л.Т. Кузиным, И.П. Кузнецовым, Д.Г. Лахути, Н.Н. Леонтьевой, Л.И. Литвинцевой, Ю.Я. Любарским, М.Г. Мальковским, А.Г. Мацкевичем, И.А. Мельчуком, Л.И. Микуличем, А.С. Нариньяни, Г.С. Осиповым, Г.С. Плесневичем, Э.В. Поповым, Д.А. Поспеловым, В.Ш. Рубашкиным, В.А. Тузовым, З.М. Шаляпиной, Г.С. Цейтиным, Л.Л. Цинманом и другими учеными.

За рубежом наибольший вклад в разработку методов математического описания содержания (смысла) ЕЯ-текстов внесли Р. Монтегю (грамматики Монтегю), Дж. Барвайз и Р. Купер (теория обобщенных кванторов, ситуационная теория), М. Кресвелл (теория структурированных значений предложений), Й. Гронендейк и М. Стокхоф (динамические грамматики Монтегю, динамическая предикатная логика), Дж. Сова (теория концептуальных графов), Л. К. Шуберт и Ч.Х. Хуан (эпизодическая логика), Г. Камп и У. Рейль (теория представления дискурсов).

Несмотря на усилия, предпринимавшиеся в течение многих лет учеными разных стран, до последнего времени многие существенные аспекты проблемы формального описания содержания ЕЯ-текстов оставались мало изученными. Одна из основных причин этой ситуации заключается в том, что внимание уделялось, главным образом, формализации смысловой структуры отдельных фраз, а не дискурсов. Кроме того, недостаточно изученной является проблема формального описания смысловой структуры отдельных фраз, обозначающих высказывания

и включающих описание ситуации множество и/или придаточные цели и/или слова «понятие», «термин», а также структуры фраз, выражающих команды и вопросы.

Наконец, сегодня ясно, что понимание ЕЯ-текста осуществляется в контексте системы знаний о мире и о целях интеллектуальных систем. Однако выразительные возможности большинства известных подходов к математическому описанию смысловой структуры ЕЯ-текстов (а именно, грамматик Монтегю, теории обобщенных кванторов, ситуационной теории, теории структурированных значений предложений, динамических грамматик Монтегю, динамической предикатной логики) недостаточны для построения теорий компьютерного понимания ЕЯ в контексте системы знаний о мире и о целях интеллектуальных систем. Например, исследования по дескриптивным логикам, выросшие из работ по терминологическим языкам представления знаний (ЯПЗ), показали полезность включения в состав ЯПЗ составных обозначений понятий. Однако перечисленные непосредственно выше подходы не предоставляют такой возможности.

Проблема автоматизации формирования баз знаний ИС посредством извлечения информации из ЕЯ-текстов с помощью ЛП, проблема разработки семантического языка-посредника для устранения языкового барьера между пользователями сети Интернет и ряд других актуальных научно-технических проблем требуют создания эффективных средств формального представления содержания произвольных ЕЯ-текстов, относящихся к *деловой прозе* (термин А.П. Ершова, ставший широко популярным в компьютерной лингвистике), т.е. ЕЯ-текстов, относящихся к юриспруденции, бизнесу, медицине, технике и т.д.

Между тем перечисленные наиболее популярные подходы к формальному представлению содержания ЕЯ-текстов имеют ограниченную сферу применения. В частности, эти подходы не предоставляют адекватных формальных средств для представления содержания произвольных предложений с описаниями множеств или составными обозначениями понятий, дискурсов со ссылками на смысл фраз и более крупных частей текстов, с обозначениями сложных целей, с косвенной речью.

Так, язык-посредник UNL ориентирован на представление содержания отдельных предложений, а не дискурсов. Кроме того, в языке UNL нет формальных средств описания множеств, средств формального различения описаний объектов и описаний понятий, квалифицирующих эти объекты, средств представления ссылок на смысл фраз и более крупных фрагментов дискурсов.

В связи с этим актуальна проблема разработки более мощных математических методов описания смысловой структуры реальных предложений и связных текстов, относящихся к юриспруденции, бизнесу, медицине, технике, экономике и т.д.

1.3. Подходы к формализации семантики естественного языка, разработанные в конце 1960-х – первой половине 1980-х гг.

Основные результаты в теории лингвистических процессоров были получены с конца 1960-х гг. В 1970-е гг. было достигнуто значительное продвижение вперед в отношении принципов «понимания» компьютерной системой естественного языка (ЕЯ). В большой степени этому способствовали работы Т. Винограда (Winograd, 1971), У. Вудса и Р. Каплана (Woods, Kaplan, 1971), Й. Уилкса (Wilks, 1973), Р. Шенка и его коллег Ч. Ригера, Н. Голдмана, Ч. Ризбека (Schank et al., 1975). Проекты, реализованные этими исследователями, показали, в частности, что (а) понимание ЕЯ-выражения осуществляется в рамках базы знаний (БЗ) о мире и (б) целесообразно использовать специальные формальные выражения для отображения смысла ЕЯ-выражений, причем выбор подхода к построению этих формальных выражений значительно влияет на процесс разработки ЛП.

Наиболее популярными подходами к построению таких формальных структур для отображения смысла ЕЯ-текстов в 1970-е гг. являлись теория семантических сетей (ТСС), импульс к появлению которой был дан работами Куиллиана (Quillian, 1968) и Р. Саймонса (Simmons, 1973), а также теория смысловой зависимости в естественном языке (ТСЗЕЯ) Р. Шенка (Schank, 1972; Schank et al., 1975). Как известно, семантические сети являются ориентированными графами со специальными метками вершин и ребер. Метки вершин обозначают понятия, реальные предметы, ситуации (в частности, события), числа, значения цветов и т.д., а метки ребер соответствуют смысловым отношениям между элементами текста и/или между понятиями. ТСЗЕЯ предложила способы построения диаграмм определенных видов для отображения смысла фраз и коротких связных текстов. Оба этих подхода не были математическими, но ТСС использовала математическое понятие ориентированного графа для иллюстрации способов представления содержания простых фраз и текстов в виде размеченного графа.

В нашей стране в 1970-е и 1980-е гг. в процессе развития теории семантических сетей И.П. Кузнецовым (Кузнецов, 1976, 1978, 1986)

влияния) теория расширенных семантических сетей (см. раздел 6.10 данной книги).

Успехи 1970-х гг. по реализации экспериментальных проектов понимания компьютером отдельных фраз на ЕЯ и текстов, состоящих из нескольких простых фраз, позволили выдвинуть в 1980-е гг. перед исследователями следующие новые задачи: (а) переход от обработки простых фраз к обработке связных текстов (дискурсов), включающих пропуски слов в отдельных фразах (явление эллипсиса), ссылки на ранее упомянутые объекты («для этого предприятия» и т.п.) и ссылки на смысл предыдущих фраз и более крупных частей текста («об этом», «этот метод» и т.п.); (б) эффективный учет прагматики общения, т.е. интерпретация очередного ЕЯ-выражения в контексте всего диалога; (в) создание формальных, предметно-независимых методов проектирования ЛП с целью получения возможности широкого тиражирования эффективных проектных решений.

Вопрос о необходимости эффективных формальных методов для проектирования ЛП возник совершенно естественно. Дело в том, что опыт реализации в 1970-х гг. экспериментальных проектов ЛП показал, что ЛП, обеспечивающие информационные потребности реального пользователя в той или иной предметной области, будут являться сложными программными комплексами. При разработке сложных технических систем в различных предметных областях широко используются математические методы. Например, для конструирования самолетов разработана и используется аэродинамика, а для проектирования кораблей и подводных лодок применяется гидромеханика.

Достаточно полное представление о том, какие формальные инструменты для изучения семантики ЕЯ были доступны к середине 1980-х гг. разработчикам ЛП за рубежом, дают учебник (Thayse et al., 1988) по логическим методам в научном направлении Искусственный Интеллект (ИИ), учебник (Partee, ter Meulen, Wall, 1990) по математической лингвистике и учебники (Grishman, 1986; Gazdar, Mellish, 1989) по компьютерной лингвистике. Анализ этих источников и целого ряда других публикаций показывает, что в этот период запас формальных методов для изучения семантики ЕЯ был довольно бедным в отношении математического описания смысловой структуры реальных связных текстов (или дискурсов) на ЕЯ и соответствия между текстами и их семантическими представлениями (СП) с учетом базы знаний о мире.

Общепринято считать, что история современных подходов к формализации семантики ЕЯ начинается с фундаментальных работ американского логика Р. Монтегю (Montague, 1970, 1974a, 1974b). Подход к

формализации семантики ЕЯ, изложенный в этих работах, впоследствии был назван Грамматикой Монтегю. Рядом исследователей были предложены различные расширения Грамматики Монтегю; в частности, такие расширения рассматриваются в работах (Partee, 1976; Thomason, 1980). В 1980-х гг. было предпринято несколько попыток использовать подход Монтегю в проектировании ЛП. Однако теория формализации семантики ЕЯ, получившая название грамматики Монтегю, (а) не является универсальной, (б) недостаточно удобна для практики с точки зрения вычислительной эффективности, (в) изложена ее автором в трудно воспринимаемой форме, что тормозило прямое использование этой теории на практике. Поэтому подход Монтегю был переработан и дополнен (иногда существенно) с целью использования его в проектировании ЛП. Так, Клиффорд (Clifford, 1983, 1988) разработал определение формального языка QE-III для представления содержания вопросов к историческим базам данных. Это было сделано на основе расширения Грамматики Монтегю.

Фреймоподобный семантический язык FRAIL (Hirst, 1988) отходит довольно далеко от Грамматики Монтегю. Язык FRAIL использовался в семантическом интерпретаторе ABSITY для представления результатов обработки ЕЯ-текстов с целью включения этих результатов в базу знаний. Джоуси (Jowsey, 1987) предложил упрощенную версию Грамматики Монтегю для построения СП текстов в прикладной интеллектуальной системе, выполняющей рассуждения общего вида. Сембок и Райсберген (Sembock, van Rijsbergen, 1990) применили язык Джоуси, близкий к языку логики первого порядка, в экспериментальной информационно-поисковой системе.

Обобщенные грамматики фразовых структур (Gazdar, Klein, Pullum, Sag, 1985) также могут рассматриваться как расширения Грамматики Монтегю, поскольку в этих грамматиках язык интенциональной логики Монтегю используется для построения СП предложений.

Другими основными подходами к формальному изучению семантики ЕЯ в первой половине 1980-х гг. были теория обобщенных кванторов (Barwise, Cooper, 1981; Gaerdenfors, 1987; Peres, 1991), ситуационная семантика (Barwise, Perry, 1983; Fenstad, Halvorsen et al., 1987; Cooper, 1991), теория представления дискурсов (Kamp, 1981; Kamp, Reyle, 1990). Все эти подходы имеют общую отправную точку – пионерские работы Р. Монтегю – и ряд взаимосвязей.

Анализ показывает, что глубокая связь с традициями математической логики является главной причиной очень большого разрыва между возможностями этих подходов и требованиями, предъявляемыми прак-

тикой проектирования ЛП. В частности, можно выделить следующие ограничения этих подходов к формальному изучению ЕЯ:

1. Неадекватность с точки зрения описания структурированных значений дискурсов на ЕЯ. В частности, отсутствие выразительных возможностей для описания семантической структуры дискурсов, содержащих ссылки на смысл фраз и более крупных частей текста (такие ссылки могут задаваться, например, словами и выражениями «поэтому», «об этом», «данный метод», «это распоряжение», «поставленный вопрос»).

2. Идущая от логики внутренняя ориентация на рассмотрение выражений, представляющих высказывания. Между тем еще существуют выражения, обозначающие вопросы, команды, цели, действия, пожелания, советы, обещания, назначения вещей, поэтому необходим формальный аппарат для описания смысловой структуры таких выражений. В этой связи следует отметить, что за рубежом первые шаги в этом направлении были сделаны теорией структурированных значений предложений (Cresswell, 1985; Chierchia, 1989), а в нашей стране первые шаги такого рода были сделаны в публикациях автора (Фомичев, 1981а, б, 1983).

3. Игнорирование или недостаточно глубокое рассмотрение многих важных особенностей структуры выражений, обозначающих высказывания. В частности, можно отметить отсутствие адекватных формальных методов описания: (а) множеств, операций над множествами и отношений на множествах; (б) назначений вещей; (в) семантической структуры фраз, содержащих причастные обороты и придаточные определительные предложения; (г) структуры фраз, в которых логические связки «и», «или» соединяют не обозначения высказываний, а обозначения различных объектов, понятий, множеств или назначений вещей («считывание или запись данных», «прием и отправка груза» и т.п.); (д) структуры предложений со словами «понятие», «термин».

4. Структура данных, позволяющих поставить в соответствие выражению на ЕЯ одно или несколько возможных семантических представлений (СП), либо не моделировалась, либо моделировалась нереалистично с точки зрения разработки ЛП, способных анализировать тексты, относящиеся к науке, технике, экономике, медицине или юриспруденции.

5. Хорошо известно, что понимание ЕЯ-текста человеком может существенно зависеть от знаний этого человека (другими словами, реципиента текста) о реальности. Между тем основные подходы к фор-

мализации семантики ЕЯ, популярные в 1980-е гг., не обладали выразительной силой, необходимой для эффективного описания знаний о реальности, для построения моделей концептуальной памяти и т.д.

6. Как следствие, за рубежом не разрабатывались модели соответствия «Текст – Система знаний – Семантическое представление (или представления) текста».

7. В 1980-е гг. ряд исследователей отмечал необходимость и важность моделирования процессов использования ЕЯ в общении, т.е. с учетом целей интеллектуальных систем, реализуемых в процессе общения, и их знаний о мире в целом и о другом участнике диалога (Попов, 1982; Fomichov, 1983, 1984; Narin'yani, 1984; Фомичев, 1988б; Fomichov, 1992). Однако наиболее популярные в 1980-е гг. формализмы, использовавшиеся для изучения семантики ЕЯ, не предоставляли такой существенной возможности.

Представляется, что перечисленные ограничения являются наиболее важными с точки зрения проектирования семантико-синтаксических анализаторов дискурсов, относящихся к науке, технике, экономике, медицине, а также для разработки ЕЯ-интерфейсов больших баз данных и знаний.

1.4. Роль формальных систем семантических представлений с большими выразительными возможностями в проектировании лингвистических процессоров

Совокупность задач, поставленных перед теорией ЛП в начале 1980-х гг., оказалась чрезвычайно трудной. Как следствие, развитие теории ЛП в эти годы сильно замедлилось. Несмотря на реализацию значительного количества проектов конструирования ЛП в разных странах мира, существенного продвижения вперед не удавалось достичь.

Главная причина этого замедления заключалась в следующем. В ЕЯ причудливым образом взаимодействуют многочисленные механизмы кодирования и декодирования информации. Поэтому часто для того чтобы «понять» даже довольно простые для человека фразы или дискурсы, компьютер должен привлекать знания о закономерностях различных уровней языка (морфологическом, синтаксическом, семантическом), а также знания о мире и о конкретной ситуации диалога. Например, для того чтобы узнать, какие из нескольких ранее упомянутых объектов обозначаются местоимением «их», может потребоваться проведение умозаключений здравого смысла и логических рассуждений. Аналогичная ситуация имеет место и для задачи восстановления смысловой

структуры фраз с пропусками слов (эллиптических фраз) в контексте всего дискурса или всего диалога.

Поэтому, пытаясь формализовать понимание компьютером даже довольно простых текстов, исследователи быстро убеждались в том, что для решения их частных задач необходимо предварительно иметь теоретические решения, относящиеся к произвольным текстам группы естественных языков (например, русского, английского, немецкого, французского). В итоге в 1980-е гг. в англоязычных публикациях даже возникла метафора «theory bottleneck» («узкое горлышко теории»), отражающая значительные трудности создания адекватной теории понимания компьютером ЕЯ.

Наконец, несколькими группами исследователей из разных стран (в том числе и автором данной работы) была предложена идея, позволяющая найти выход из охарактеризованной тупиковой ситуации. Суть этой идеи заключается в следующем. Необходимо разработать такие формальные языки для представления знаний о мире и построения семантических представлений (СП) ЕЯ-текстов, чтобы можно было конструировать СП в виде выражений, отражающих многие структурные особенности самих текстов. Другими словами, нужны формальные языки (или формальные системы, поскольку множество их правильно построенных выражений образует язык) для описания структурированных значений (или смыслов) ЕЯ-текстов, обладающие выразительными возможностями, близкими к возможностям ЕЯ. Тогда можно будет выполнять смысловой анализ текста в два этапа:

$$\begin{aligned} \text{ЕЯ-текст } T &\rightarrow \text{Недоопределенное СП текста } T \rightarrow \\ &\rightarrow \text{Целевое СП текста } T. \end{aligned}$$

Эту схему следует понимать следующим образом. Сначала должно быть построено промежуточное, предварительное СП текста, называемое недоопределенным семантическим представлением (НСП) рассматриваемого текста. Это выражение в большинстве случаев будет отображать смысл входного текста T лишь частично, неполно. Например, в НСП текста T может отсутствовать указание на конкретный объект, соответствующий конкретному вхождению в текст T местоимения «ей» или не выбрано конкретное значение слова «станция», входящего в T .

Однако НСП текста T является формальным выражением, в отличие от исходного ЕЯ-текста T . Поэтому на втором этапе обработки T для снятия той или иной недоопределенности можно будет вызвать одну из многочисленных специализированных процедур-«экспертов» по кон-

кретным вопросам. Такие процедуры можно будет проектировать с применением формальных средств представления информации, поскольку базы знаний ЛП состоят из выражений формальных языков представления знаний, а исходное НСП – вход процедуры и преобразованное НСП (в частности, совпадающее с целевым СП) являются формальными выражениями. Впервые эта идея была высказана в работах (Фомичев, 1981а, 1981б; Fomichov, 1983, 1984).

С конца 1980-х гг. по настоящее время идея опоры при проектировании ЛП на формальные системы семантических представлений с широкими выразительными возможностями является центральной для развития теории понимания компьютером ЕЯ. Росту популярности этой идеи способствовало появление серии публикаций по эпизодической логике (ЭЛ) (Schubert, Hwang, 1989, 2000; Hwang, Schubert, 1993а–1995), реализация на основе ЭЛ проекта TRAINS, направленного на формализацию проблемно-ориентированного диалога на естественном (английском) языке (Allen, Schubert et al., 1995), осуществление проекта машинного перевода Core Language Engine (CLE) в Кембриджском отделении (Великобритания) Стенфордского исследовательского института (Alshawī и van Eijck, 1989; Alshawī, 1990, 1992), а также осуществление в 1980–1990-х гг. проекта SnepS в США (Shapiro, 1996).

Отправной точкой исследований для авторов перечисленных выше работ был язык логики предикатов первого порядка. Характер принимавшихся усилий по расширению выразительных возможностей этого языка (точнее, класса логики предикатов первого порядка) можно проиллюстрировать следующими двумя примерами.

Пример 1. В ЕЯ-текстах встречается большое количество таких обозначений различных объектов, которые являются сочетаниями «Прилагательное + Существительное» (для русского языка) или «Артикль + Прилагательное + Существительное» (для английского, немецкого и французского языков). ЭЛ позволяет рассматривать формальные аналоги значений таких словосочетаний. Например, сочетанию «маленький дом» может соответствовать фрагмент семантического представления (СП) текста, являющийся выражением $\exists y : [y((\text{атрибут, маленький})\text{дом})]$ (Hwang и Shubert, 1993а).

В ЕЯ-текстах часто встречаются и однородные члены предложения, например, «комнату или маленький дом». ЭЛ даст возможность строить и формальные аналоги выражений такого вида. В частности, в СП текста сочетанию «комнату или маленький дом» может соответствовать формальное выражение $\exists y : [[y \text{ комната}] \vee [y ((\text{атрибут маленький})\text{дом})]]$ (Hwang, Shubert, 1993а).

В логике предикатов первого порядка нет средств построения формальных аналогов словосочетаний вида «Прилагательное + Существительное» или «Существительное1 + 'или' + Прилагательное + Существительное2». Поэтому различные части СП текста могут отражать различные компоненты смысла словосочетаний подобного рода.

Пример 2. В проекте «Базовый Языковой Механизм» (Core Language Engine, или CLE) формулы, используемые для построения недоопределенных семантических представлений (НСП) предложений, называются квазилогическими формами (КЛФ). В частности, выражению «the three firms» (три фирмы) будет соответствовать КЛФ (см. Alshawi, 1990) $q_term(< t = quant, n = plur, l = all >), S, [subset, S, q_term(< t = ref, p = def, l = the, n = number(3) >, X, [firm, X])]$.

Таким образом, язык КЛФ позволяет строить формальные аналоги некоторых естественно-языковых обозначений множеств (Alshawi, 1990).

Сегодня наиболее популярным за рубежом подходом к формализации семантически-ориентированных компьютерных методов анализа дискурсов является эпизодическая логика (ЭЛ). Ее создатели, Л.К. Шуберт и Ч.Х. Хуан, в указанных выше работах предложили логику семантических представлений с широкими выразительными возможностями. Создание ЭЛ представляло собою значительный вклад в формальную теорию построения и понимания ЕЯ-дискурсов и, как следствие, в формальную теорию использования ЕЯ.

В то же время анализ показывает, что выразительная сила класса формул, рассматриваемых в ЭЛ, недостаточна с точки зрения представления основных принципов использования ЕЯ интеллектуальными системами. В первую очередь, выразительные возможности ЭЛ являются существенно ограниченными с точки зрения представления структурированных значений (СЗ) сложных целей, команд, дискурсов с описаниями множеств, дискурсов со ссылками на смысл фраз и более крупных частей текста. Кроме того, ЭЛ не предоставляет средств для рассмотрения составных обозначений понятий в качестве термов и для описания операций над понятиями. Те же ограничения (и ряд других) относятся к языкам проекта Core Language Engine и SnepS.

Логика предикатов первого порядка послужила в 1980-х гг. отправной точкой для создания не только ЭЛ, но и нескольких других направлений, относящихся к области логического программирования. Одним из наиболее интересных направлений, развивающимся более 15 лет, является атрибутная логика (АЛ), или логика со значениями свойств (Johnson, 1988; Carpenter, 1992; Carpenter, Penn, 2001). Для АЛ харак-

терен переход от рассматривания однородного множества логических формул, описывающих факты, ситуации, к группировке формул, характеризующих одну сущность (человека, фирму и т.д.). Перечисленные выше ограничения ЭЛ относятся и к атрибутной логике.

Проблема формализации семантики ЕЯ-текстов в течение многих лет привлекает внимание исследователей и в нашей стране. Наиболее часто это внимание было обусловлено задачей выявления и формализации таких явлений семантического уровня ЕЯ, которые можно было бы эффективно использовать для представления знаний в прикладных интеллектуальных системах. В этой связи можно отметить цикл работ Д.А. Поспелова, Ю.И. Клыкова и ряда других авторов, в которых выделяются бинарные смысловые отношения между элементами предложения для решения задач ситуационного управления (Поспелов, 1975, 1981, 1986; Клыков и Горьков, 1980), Г.С. Плесневича по теории логического вывода на ассоциативных сетях и понятийно-ориентированным языкам (Плесневич, 1997–2003), Г.С. Осипова по использованию неоднородных семантических сетей в интеллектуальных системах приобретения знаний (Осипов, 1990, 1997), В.Н. Вагина по проблеме обобщения знаний, представленных семантическими сетями (Вагин, 1988).

Другую часть исследователей интересовала разработка формальных средств отображения содержания (смысла) текстов, анализируемых ЛП. В работе Н.Н. Леонтьевой (Леонтьева, 1981), по-видимому, впервые в нашей стране был высказан тезис о том, что для формализации диалога, осуществляемого в ограниченных предметных областях с помощью довольно простых текстов, нужен семантический язык с выразительными возможностями, близкими к возможностям ЕЯ.

Отдельные аспекты формализации семантики ЕЯ нашли отражение в работах В.М. Брябрина, Б.Ю. Городецкого, А.К. Жолковского, А.П. Ершова, О.С. Кулагиной, Е.С. Кузина, Л.Т. Кузина, И.П. Кузнецова, Д.Г. Лахути, Н.Н. Леонтьевой, Л.И. Литвинцевой, Ю.Я. Любарского, М.Г. Мальковского, А.Г. Мацкевича, И.А. Мельчука, Л.И. Микулича, А.С. Нариньяни, Г.С. Осипова, Э.В. Попова, Д.А. Поспелова, В.Ш. Рубашкина, В.А. Тузова, З.М. Шаляпиной, Г.С. Цейтина, Л.Л. Цинмана и ряда других ученых.

Однако ни в какой из публикаций отечественных авторов, во-первых, не предлагается формального аппарата, удобного для отображения поверхностной смысловой структуры произвольных ЕЯ-текстов, относящихся к деловой прозе. Во-вторых, не предлагается лингвистической теории, которую можно было бы положить в основу построения математической модели, удобной для представления смысловой

структуры ЕЯ-текстов деловой прозы. Таким образом, вопрос о формальных языках, удобных как для построения семантических представлений произвольных ЕЯ-текстов, относящихся к деловой прозе, так и для моделирования ЕЯ-диалога интеллектуальных систем, является чрезвычайно актуальным и остается в доступной литературе открытым (исключение составляют публикации автора данной работы).

Прогресс в решении этого вопроса означал бы существенный шаг вперед в решении фундаментальной проблемы разработки модели русского языка, сформулированной А.П. Ершовым еще в 1986 г. следующим образом: «Мы хотим как можно глубже познать природу языка, и в частности русского. Одним из выражений этого познания должна стать модель русского языка. Это формальная система, которая должна быть адекватной и равнообъемной живому организму языка, но в то же время она должна быть анатомически отпрепарированной, разъятой, доступной для наблюдения, изучения и изменения» (Ершов, 1986, с. 12).

Глава 2

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ДЛЯ ОПИСАНИЯ СИСТЕМЫ ПЕРВИЧНЫХ ЕДИНИЦ КОНЦЕПТУАЛЬНОГО УРОВНЯ, ИСПОЛЪЗУЕМЫХ МНОГОАГЕНТНОЙ СИСТЕМОЙ

2.1. Постановка задачи

Проанализировав по доступной литературе состояние исследований в области формализации семантики ЕЯ, Перегрин (Peregrin, 1990) пришел к выводу о том, что существующие логические системы не позволяют формализовать все аспекты семантики ЕЯ, являющиеся важными для проектирования ЛП. Поэтому «мы не можем использовать имеющуюся форму логики как плавильную форму, в которую любой ценой необходимо втиснуть естественный язык», и для создания адекватной формальной теории семантики ЕЯ необходимо выполнить системное лингвистическое исследование всех компонентов ЕЯ и установить взаимосвязи между логическими подходами к формализации семантики ЕЯ и лингвистическими моделями смысла.

В сущности, к тому же выводу (но значительно раньше, на рубеже 1970–1980-х гг.) пришел автор данной монографии. Этот вывод явился отправной точкой для разработки излагаемой ниже постановки задачи, а также постановки задачи в главе 3.

Представляется, что границы традиционной математической логики слишком узки для того, чтобы предоставить адекватную основу для компьютерно-ориентированной формализации семантики ЕЯ. Поэтому задача создания логических основ проектирования интеллектуально мощных систем смысловой обработки ЕЯ требует не только расширения логики первого порядка, но, скорее, разработки новых математических систем, совместимых с логикой предикатов первого порядка и позволяющих формализовать логику использования ЕЯ интеллектуальными системами.

Мы будем исходить из гипотезы о том, что существует единственный ментальный уровень для представления смысла ЕЯ-выражений, который можно назвать концептуальным уровнем, но не семантический и концептуальный уровни отдельно. Эту гипотезу поддерживают многие ученые (см., например, Meyer, 1994).

Анализ показывает, что первым шагом на пути создания широко применимого и предметно-независимого математического подхода к описанию структурированных значений ЕЯ-текстов должна являться разработка формальной модели, перечисляющей первичные (т.е. не составные) единицы концептуального уровня, используемые многоагентной системой (МАС), а также описывающей информацию, связанную с такими единицами и необходимую для соединения таких единиц в составные единицы, отображающие структурированные значения сколь угодно сложных ЕЯ-текстов.

С целью построения формальной модели, обладающей указанным свойством, был, во-первых, проведен анализ лексического состава русского, английского, немецкого и французского языков.

Во-вторых, был изучен состав первичных информационных единиц, используемых в современных языках представления знаний в прикладных интеллектуальных системах, в частности, используемых в терминологических языках представления знаний.

На основании проведенного исследования в данной главе ставится задача разработки такой предметно-независимой математической модели для описания системы первичных единиц концептуального уровня, используемых МАС, и информации, связанной с такими единицами, которая, во-первых, конструктивно учитывает существование следующих явлений естественного языка:

1.1. На множестве понятий задана иерархия по степени их общности. Например, понятие «физический объект» является частным случаем понятия «пространственный объект».

1.2. Нередко один и тот же предмет может быть охарактеризован с помощью нескольких понятий, ни одно из которых не является частным случаем другого; такие понятия как бы дают значения «координат объекта» по разным «семантическим осям». Например, каждый человек является физическим объектом, способным перемещаться в пространстве. С другой стороны, каждый человек является интеллектуальной системой, поскольку люди могут решать задачи, читать, сочинять стихи и т.д.

1.3. В русском языке есть такие слова, как «некоторый», «определенный», «каждый», «какой-нибудь», «все», «несколько», «большин-

и ряд других, которые в предложениях всегда присоединяются к словам и словосочетаниям, обозначающим понятия. Например, мы можем построить выражения «каждый человек», «какой-нибудь автомобиль», «все люди», «несколько книг» и т.д. Аналогичные слова есть в английском, немецком, французском и многих других языках.

Во-вторых, модель должна позволять различать формальным образом обозначения первичных единиц концептуального уровня, соответствующих:

(2.1) объектам, ситуациям, процессам в реальном мире и понятиям, квалифицирующим (характеризующим) эти объекты, ситуации, процессы;

(2.2) объектам и множествам объектов;

(2.3) понятиям, квалифицирующим объекты, и понятиям, квалифицирующим множества объектов тех же видов («корабль» и «эскадра» и т.д.);

(2.4) упорядоченным n -местным наборам различных сущностей, где $n > 1$ («упорядоченная пара» и т.д.) и множествам.

В-третьих, модель должна учитывать, что совокупность первичных единиц концептуального уровня включает:

(3.1) единицы, соответствующие логическим связкам «и», «или», «не» и логическим кванторам существования и всеобщности;

(3.2) имена нетрадиционных функций с аргументами и/или значениями, являющимися: (3.2.1) множествами предметов, ситуаций (событий); (3.2.2) понятиями, (3.2.3) множествами понятий; (3.2.4) семантическими представлениями (СП) ЕЯ-текстов, (3.2.5) множествами СП ЕЯ-текстов;

(3.3) единицу, соответствующую слову «понятие» и отличающуюся от концептуальной единицы «понятие»; первая из упомянутых единиц вносит, например, вклад в формирование значения выражения «важное понятие, используемое в физике, химии и биологии».

Итогом решения поставленной задачи станет определение в разделе 2.8 класса формальных объектов, называемых концептуальными базисами. Описание класса концептуальных базисов является математической моделью, перечисляющей первичные единицы концептуального уровня, используемые МАС, а также описывающей информацию, связанную с такими единицами и необходимую для соединения этих единиц в составные единицы, отображающие структурированные значения сколь угодно сложных ЕЯ-текстов. Данная модель является первой частью теории К-представлений (концептуальных представлений).

2.2. Базовые обозначения и вспомогательные определения

2.2.1. Общематематические обозначения

$x \in Y$	элемент x принадлежит множеству Y
$x \notin Y$	элемент x не входит в множество Y
$X \subset Y$	множество X является подмножеством множества Y
$Y \cup Z$	объединение множеств Y и Z
$Y \cap Z$	пересечение множеств Y и Z
$Y \setminus Z$	теоретико-множественная разность множеств Y и Z , т.е. совокупность всех таких элементов x из Y , что x не входит в Z
$Z_1 \times \dots \times Z_n$	декартово произведение множеств $Z_1, \dots, Z_n$, где $n > 1$
$\emptyset$	пустое множество
$\forall$	для любого, для любых
$\exists$	существует
$\rightarrow$	следует, влечет за собой
$\Leftrightarrow$	тогда и только тогда

2.2.2. Предварительные определения и обозначения из теории формальных грамматик и языков

Определение. Алфавитом называется конечное множество символов. Если A – произвольный алфавит, то A^+ – это множество всех последовательностей вида $d_1, \dots, d_n$, где $n \geq 1$, для $i = 1, \dots, n$ $d_i \in A$.

Обычно вместо $d_1, \dots, d_n$ для упрощения пишут $d_1 \dots d_n$.

Пример. Если $A = \{0, 1\}$, то $011, 11011, 0, 1 \in A^+$.

Определение. Элементы множества A^+ называются непустыми цепочками (или непустыми строками) в алфавите A (над алфавитом A).

Пусть A – произвольный алфавит, d – символ из A , тогда $d^1 = d$, для $n > 1$ $d^n = dd \dots d$ (n раз).

Определение. Пусть $A^* = A^+ \cup \{e\}$, где e – пустая цепочка. Тогда цепочками (или строками) в алфавите A (или над алфавитом A) называются элементы множества A^* .

Определение. Для каждого $t \in A^*$ определено значение функции *Длина*(t) (обозначаемой также через $|t|$) следующим образом:
(1) $|e| = 0$; (2) если $t = d_1 \dots d_n$, $n \geq 1$, для $i = 1, \dots, n$ $d_i \in A$, то $|t| = n$.

Определение. Пусть A – произвольный алфавит. Тогда формальным языком (или, для краткости, языком) в алфавите A (или над алфавитом A) называется произвольное подмножество L множества A^* .

Пример. Пусть $A = \{0,1\}$, $L_1 = \{0\}$, $L_2 = \{e\}$, $L_3 = \{0^{2k}1^{2k} \mid k \geq 1\}$, тогда L_1, L_2, L_3 – языки в алфавите A .

2.2.3. Используемые определения из теории алгебраических систем

Определение. Пусть $n \geq 1$, Z – произвольное непустое множество. Тогда декартовой n -степенью множества Z называется (и обозначается через Z^n) множество Z при $n = 1$ и множество всех упорядоченных наборов вида $(x_1, x_2, \dots, x_n)$, где $x_1, x_2, \dots, x_n$ – элементы множества Z , при $n > 1$.

Определение. Пусть $n \geq 1$, Z – произвольное непустое множество. Тогда n -арным (или n -местным) отношением на множестве Z называется произвольное подмножество R множества Z^n – декартовой n -степени множества Z . При $n = 1$ отношение R называется *унарным отношением* (в этом случае R является произвольным подмножеством множества Z), а при $n = 2$ отношение R называется *бинарным отношением* (Ершов, 1970).

Пример. Пусть Z_1 – множество всех целых чисел, и Odd – подмножество всех нечетных чисел. Тогда Odd является унарным отношением на Z_1 . Пусть $Less$ является множеством всех упорядоченных пар вида (x, y) , где x, y – произвольные элементы множества Z_1 , и число x меньше числа y . Тогда $Less$ – бинарное отношение на множестве Z_1 .

Очень часто вместо записи $(b, c) \in R$, где R – бинарное отношение на произвольном множестве Z , b, c – произвольные элементы из Z , используется запись $b R c$.

Определение. Пусть Z – произвольное непустое множество, R – бинарное отношение на Z . Тогда

- (а) если для любого $a \in Z$ $(a, a) \in R$, то R – рефлексивное отношение;
- (б) если для любого $a \in Z$ $(a, a) \notin R$, то R – антирефлексивное отношение;
- (в) если для любых $a, b, c \in Z$ из $(a, b) \in R$, $(b, c) \in R$ следует, что $(a, c) \in R$, то R называется транзитивным отношением;
- (г) если для любых $a, b \in Z$ из $(a, b) \in R$ следует $(b, a) \in R$, то R – симметричное отношение;

- (д) если для любых $a, b \in Z$ из $a \neq b$ и $(a, b) \in R$ следует, что $(b, a) \notin R$, то R – антисимметричное отношение;
- (е) если R – рефлексивно, транзитивно и антисимметрично, то R – частичный порядок на Z (Johnsonbaugh, 2001).

Пример. Бинарное отношение *Less* из предыдущего примера является транзитивным, антирефлексивным и антисимметричным.

Пример. Пусть $Z1$ – множество всех целых чисел, и *Eqless* является множеством всех упорядоченных пар вида (x, y) , где x, y – произвольные элементы множества $Z1$, и число x равно числу y или меньше числа y . Тогда *Eqless* – бинарное отношение на множестве $Z1$. Это отношение является рефлексивным, транзитивным и антисимметричным. Таким образом, отношение *Eqless* является частичным порядком на $Z1$.

Пример. Пусть $Z2$ – множество всех понятий, которые обозначают транспортные средства, и *Genrel* является множеством всех упорядоченных пар вида (x, y) , где x, y – произвольные элементы множества $Z2$, и понятие x совпадает с понятием y или является обобщением понятия y . Например, понятие *корабль* является обобщением понятия *ледокол*; следовательно, пара $(\text{корабль}, \text{ледокол})$ входит в множество *Genrel*. Очевидно, что *Genrel* – бинарное отношение на множестве $Z2$. Это отношение является рефлексивным, транзитивным и антисимметричным. Таким образом, отношение *Genrel* является частичным порядком на $Z2$.

Определение. Пусть Z – произвольное непустое множество, R – бинарное отношение на Z . Тогда элементы $a, b \in Z$ называются сравнимыми для R , если либо $(a, b) \in R$, либо $(b, a) \in R$.

2.3. Краткая характеристика предлагаемой математической модели для описания системы первичных единиц концептуального уровня, используемых многоагентной системой

С математической точки зрения, решение задачи, поставленной в разделе 2.1, является определением нового класса формальных объектов, называемых *концептуальными базисами* (к.б.). Отдаленным прообразом этого понятия является понятие сигнатуры алгебраической системы (Ершов, Палютин, 1979).

Каждый к.б. B является упорядоченным набором вида

$$((c_1, c_2, c_3, c_4), (c_5, \dots, c_8), (c_9, \dots, c_{15}))$$

с компонентами $c_1, c_2, \dots, c_{15}$, являющимися (главным образом) конечными или счетными множествами символов и выделенными элементами

ими таких множеств. В частности, $c_1 = St$ – конечное множество символов, называемых сортами и обозначающих наиболее общие рассматриваемые понятия, $c_2 = P$ – выделенный сорт «смысл сообщения», $c_5 = X$ – счетное множество цепочек, используемых как «строительные блоки» для формирования модулей знаний и семантических представлений (СП) текстов, $c_6 = V$ – счетное множество переменных, $c_8 = F$ – подмножество множества X , элементы которого называются функциональными символами.

Компонент $c_3 = Gen$ является таким бинарным отношением (частичным порядком) на St , что если пара (s, u) входит в Gen , то либо $s = u$, либо понятие, соответствующее сорту u , является конкретизацией понятия, соответствующего сорту s . Компонент $c_7 = tp$ является отображением из объединения множеств X и V в некоторое счетное множество Tps цепочек, называемых типами и характеризующих элементы из X и V .

Предположим, например, что X включает элементы *интс*, *дин.физ.об.*, *чел.*, *редсовет*, *Д.И.Менделеев*, обозначающие сорт «интеллектуальная система», сорт «динамический физический объект», понятия «человек», «редакционный совет» и конкретного человека – выдающегося химика Дмитрия Ивановича Менделеева. Будем рассматривать символ $\uparrow$ как индикатор почти всех типов, связанных с понятиями. Тогда значениями отображения tp для элементов *чел.*, *редсовет*, *Д.И.Менделеев* будут элементы $\uparrow_{интс}$, $\uparrow_{\{интс*дин.физ.об.\}}$ и $\uparrow_{интс*дин.физ.об}$ соответственно. Если же в качестве элемента множества X мы рассматриваем обозначение редакционного совета конкретного издания, то для такой информационной единицы отображение tp примет значение $\{интс*дин.физ.об.\}$. Таким образом, типы помогают различать (а) объекты и понятия, характеризующие эти объекты, (б) множества и понятия, характеризующие эти множества.

Определение класса концептуальных базисов будет использовано в главе 3 следующим образом. Каждому $k.b. B$ будут поставлены в соответствие три множества формул $Ls = Ls(B)$, $Ts = Ts(B)$, $Ys = Ys(B)$ (l -формулы, t -формулы, y -формулы). Множество $Ls(B)$ будет названо *стандартным K -языком в базисе B* . Его цепочки подходят для построения семантических представлений (СП) текстов на естественном языке. Каждая формула из $Ts(B)$ имеет вид $d \& t$, где $d \in Ls(B)$, t – тип из $Tps(B)$. Формулы из $Ys(B)$ имеют вид $a_1 \& \dots \& a_n \& d$, где $a_1, \dots, a_n, d \in Ls(B)$, n имеет разные значения для разных d , цепочка d строится из $a_1, \dots, a_n$ как из элементарных информационных единиц (некоторые из них мо-

гут быть немного преобразованы) однократным применением некоторого правила построения.

Например, к.б. B можно определить так, чтобы выполнялись соотношения

$$Ls(B) \ni \text{страна, нек страна, нек страна} : x1,$$

$$\text{Столица (нек страна} : x1),$$

$$(\text{Столица (нек страна} : x1) \equiv \text{Москва});$$

$$Ts(B) \ni \text{страна} \& \uparrow \text{простр.об, нек страна} \& \text{простр.об,}$$

$$\text{нек. страна} : x1 \& \text{простр.об, Столица (нек страна} : x1) \& \text{простр.об,}$$

$$(\text{Столица (нек страна} : x1) \equiv \text{Москва}) \& \text{сообщ,}$$

$$Ys(B) \ni \text{нек} \& \text{страна} \& \text{нек страна, нек страна} \& x1 \& \text{нек страна} : x1,$$

$$\text{Столица} \& \text{нек страна} : x1 \& \text{Столица (нек. страна} : x1),$$

$$\text{Столица (нек страна} : x1) \& \equiv \& \text{Москва} \& (\text{Столица (нек страна} : x1) \equiv \text{Москва}),$$

где $\text{сообщ} = P(B)$ – выделенный сорт «смысл сообщения» для рассматриваемого к.б. B , нек – информационная единица, соответствующая словам «некоторый» «некоторая», «некоторое».

2.4. Основные идеи определения класса сортовых систем

Начнем решать поставленную задачу. Будем предполагать, что необходимо построить формальное описание некоторой предметной области (ПО), и рассмотрим первые шаги в этом направлении.

Шаг 1. Введем в рассмотрение конечное множество символов, обозначающих наиболее общие понятия ПО: пространственный объект, физический объект, интеллектуальная система, натуральное число и т.д. Будем считать, что каждое такое понятие характеризует сущность, не рассматриваемую как упорядоченный набор других сущностей или как множество, состоящее из каких-то других сущностей. Обозначим это множество символов через St и будем называть его элементы сортами.

Шаг 2. Выделим в St некоторый сорт, который будем связывать с семантическими представлениями (СП) ЕЯ-текстов, выражающих отдельные высказывания либо являющихся связными повествовательными текстами. Обозначим такой сорт через P и назовем его сортом «смысл сообщения». Например, для каких-то применений роль выделенного сорта P может играть цепочка сообщ . Часть формул, которые

мы будем рассматривать в этой работе, представляема в виде $F \& t$, где F – СП ЕЯ-выражения, а t – цепочка, классифицирующая данное выражение. Тогда, если $t = P$, то подформула F интерпретируется как СП простого или сложного высказывания (другими словами, сообщения). В частности, так может интерпретироваться формула $(Вес(нек\ блок1 : x3) \equiv 4/тонна) \& сообщ.$

Шаг 3. Введем иерархию понятий на множестве сортов St с помощью некоторого бинарного отношения Gen на St , т.е. выделим некоторое подмножество $Gen \subset St \times St$. Например, могут выполняться соотношения $(цел, нат), (вещ, цел), (физ. об, дин. физ. об), (простр. об, физ. об) \in Gen$.

Шаг 4. Многие объекты могут быть охарактеризованы с разных точек зрения, у них есть «координаты» по разным «семантическим осям». Например, к конкретному университету можно подъехать или подойти, поэтому каждый университет имеет семантическую координату «пространственный объект». У университета есть руководитель (ректор), поэтому университеты имеют семантическую координату «организация». Наконец, университет может разработать некоторую технологию или некоторый прибор; следовательно, представляется разумным считать, что университеты имеют семантическую координату «интеллектуальная система».

Учитывая эти соображения, введем бинарное отношение совместимости (толерантности) Tol на множестве St . Это отношение интерпретируется следующим образом: если $(s, u) \in Tol \subset St \times St$, то существует такая сущность x в рассматриваемой ПО, что с x можно связать сорт s по одной семантической оси и сорт u по другой оси, причем сорт s и сорт u не являются сравнимыми для отношения Gen .

Например, множества St и Tol могут быть определены так, что Tol включает упорядоченные пары $(простр.объект, организация), (простр.объект, интел.система), (организация, интел.система), (организация, простр.объект), (интел.система, простр.объект), (интел.система, организация)$.

Из рассмотренной интерпретации отношения Tol вытекают следующие свойства: (1) $\forall u \in St (u, u) \notin Tol$, т.е. Tol – антирефлексивное отношение; 2) $\forall u, t \in St$ из $(u, t) \in Tol$ следует, что $(t, u) \in Tol$, т.е. Tol – симметричное отношение.

Сортовой системой (с.с.) будем называть произвольную четверку S вида (St, P, Gen, Tol) , компоненты которой удовлетворяют определенным условиям.

Определение. *Сортовой системой (с.с.)* будем называть произвольную упорядоченную четверку S вида

$$(St, P, Gen, Tol), \quad (2.5.1)$$

где St – конечное множество символов, $P \in St$, Gen – непустое бинарное отношение на St , являющееся частичным порядком на St (т.е. рефлексивным, транзитивным и антисимметричным), Tol – бинарное отношение на St , являющееся антирефлексивным и симметричным, и выполняются следующие условия:

- (1) St не включает символы ' $\uparrow$ ', '{', '}', '(', ')', ',', [сущн], [пон], [об], [$\uparrow$ сущн], [$\uparrow$ пон], [$\uparrow$ об];
- (2) $St \setminus \{P\} \neq \emptyset$ и $\forall u \in St \setminus \{P\}$ u, P – несравнимы как для отношения Gen , так и для отношения Tol ;
- (3) $\forall t, u \in St$ из $(t, u) \in Gen$ или $(u, t) \in Gen$ следует, что t, u несравнимы для отношения Tol ;
- (4) $\forall t_1, u_1 \in St, t_2, u_2 \in St$ из $(t_1, u_1) \in Tol, (t_2, t_1) \in Gen, (u_2, u_1) \in Gen$ вытекает, что $(t_2, u_2) \in Tol$.

Элементы множества St называются сортами; P – сортом «смысл сообщения»; $Gen \subset St \times St$ – отношением общности; $Tol \subset St \times St$ – отношением толерантности (совместимости). Если $(u, t) \in Gen$, то будем использовать эквивалентную запись $u \rightarrow t$ и говорить, что t – конкретизация сорта u , а u – обобщение сорта t . Если $(s, u) \in Tol$, то будем использовать запись $s \perp u$ и говорить, что сорт s совместим с сортом u .

Символы ' $\uparrow$ ', '{', '}', '(', ')', ',', [сущн], [пон], [об], [$\uparrow$ сущн], [$\uparrow$ пон], [$\uparrow$ об] будут играть особые роли при построении из сортов цепочек, называемых типами и классифицирующих сущности, рассматриваемые в выбранной предметной области (см. раздел 2.6).

Пример. Пусть $St_0 = \{nat, цел, вещь, простр.об, физ.об, дин.физ.об, вообр.об, интс, орг, сит, соб, мом, сообщ\}$. Элементы множества St_0 обозначают понятия и интерпретируются следующим образом: *nat* – «натуральное число», *цел* – «целое число», *вещь* – «вещественное число», *простр.об* – «пространственный объект», *физ.об* – «физический объект», *дин.физ.об* – «динамический физический объект», *вообр.об* – «воображаемый пространственный объект» (орбиты небесных тел, геометрические фигуры), *интс* – «интеллектуальная система», *орг* – «организация», *сит* – «ситуация», *соб* – «событие» (т.е. динамическая ситуация), *мом* – «момент времени», *сообщ* – «семантическое представление сообщения».

Пусть $P_0 = \text{сообщ.}$, $Ge1 = \{(u, u) \mid u \in St_0\}$,

$Ge2 = \{(\text{цел}, \text{нат}), (\text{вещ}, \text{цел}), (\text{вещ}, \text{нат}), (\text{простр.об.}, \text{физ.об.}),$
 $(\text{простр.об.}, \text{вообр.об.}), (\text{физ.об.}, \text{дин.физ.об.}), (\text{простр.об.}, \text{дин.физ.об.}),$
 $(\text{сум}, \text{соб})\}$,

$Gen_0 = Ge1 \cup Ge2$,

$T1 = \{(\text{интс}, \text{дин.физ.об.}), (\text{интс}, \text{физ.об.}), (\text{интс}, \text{простр.об.}), (\text{орг},$
 $\text{интс}), (\text{орг}, \text{физ.об.}), (\text{орг}, \text{простр.об.})\}$,

$T2 = \{(u, s) \mid (s, u) \in T1\}$, $Tol_0 = T1 \cup T2$.

Пусть $S_0 = (St_0, P_0, Gen_0, Tol_0)$. Тогда легко проверить, что S_0 является сортовой системой, и сорт *сообщ.* является выделенным сортом «смысл сообщения» этой системы. Из определения множества Gen_0 вытекают, в частности, следующие соотношения:

$\text{вещ} \rightarrow \text{цел}, \text{цел} \rightarrow \text{нат}, \text{вещ} \rightarrow \text{нат}, \text{простр.об.} \rightarrow \text{физ.об.},$
 $\text{простр.об.} \rightarrow \text{вообр.об.}, \text{физ.об.} \rightarrow \text{дин.физ.об.}; \text{интс} \perp \text{физ.об.},$
 $\text{интс} \perp \text{дин.физ.об.}, \text{интс} \perp \text{орг}, \text{физ.об.} \perp \text{интс}, \text{дин.физ.об.} \perp \text{интс}.$

2.6. Типы, порождаемые сортовыми системами, и конкретизации типов

2.6.1. Определение множества типов

Предположим, что нам необходимо описать некую предметную область (ПО), и мы решили рассматривать некоторые сущности как элементарные сущности (люди, фирмы, числа, факты, понятия и т.д.). Тогда определим составные сущности для данной области как такие сущности, которые рассматриваются как упорядоченные наборы других сущностей или как множества, состоящие из каких-то других сущностей. Будем интерпретировать понятия (другими словами, концепты) как общие описания сущностей, относящихся к некоторым различаемым людьми классам сущностей. Объекты определим как такие сущности, которые не рассматриваются как понятия. Класс объектов включает, в частности, семантические представления (СП) текстов, множества СП текстов и множества понятий.

Определим для каждой с.с. S множество цепочек $Tr(S)$, элементы которого назовем типами системы S и будем понимать их как характеристики сущностей, рассматриваемых в рассуждениях о данной области. При построении типов используются сорта из S и специальные

символы [сущн], [пон], [об], [$\uparrow$ сущн], [$\uparrow$ пон], [$\uparrow$ об], ' $\uparrow$ ', '{', '}', '(', ')', ',' (запятая). Символы [сущн], [пон], [об] будем называть соответственно типом «сущность», типом «концепт» (это наиболее общая характеристика понятий) и типом «объект» (это наиболее общая характеристика сущностей, не рассматриваемых как понятия). Символ «*» будет использоваться для соединения нескольких совместимых сортов (т.е. сравнимых для отношения толерантности *Tol*) при построении цепочек из множества $Tr(S)$. Символ ' $\uparrow$ ' будем интерпретировать как индикатор типа понятия.

Предположим, что мы используем сортовую систему S_0 , построенную в примере из раздела 2.5. Тогда мы сможем связать с понятием «человек» тип $\uparrow_{интс*дин.физ.об}$ из $Tr(S_0)$, с каждым конкретным человеком – тип $интс*дин.физ.об$, с понятием «студенческая учебная группа» – тип $\uparrow_{\{интс*дин.физ.об\}}$, с конкретной студенческой группой № 172 факультета бизнес-информатики ГУ-ВШЭ – тип $\{интс*дин.физ.об\}$.

Рассмотрим интерпретацию специальных символов [сущн], [пон], [об]. Формализуя рассуждения, условимся исходить из следующих рекомендаций. Если природа сущности z , рассматриваемой в рассуждении, не играет роли, то поставим в соответствие z тип [сущн] в ходе рассуждения. Если же важно то, что z представляет собой объект, то поставим в соответствие z тип [об]. Если, напротив, в отношении z важно то, что z является понятием, то поставим в соответствие z тип [пон]. Назначение типов [сущн], [пон], [об] станет понятным из следующих примеров.

Пусть E_1 и E_2 соответствуют выражениям «первая сущность, упомянутая на странице 12 выпуска газеты “The Moscow Times”, опубликованного 1 октября 1994 г.», и «первый объект, упомянутый на странице 12 выпуска газеты “The Moscow Times”, опубликованного 1 октября 1994 г.». Тогда можно связать типы [сущн] и [об] с сущностями, на которые ссылаются в E_1 и E_2 соответственно, в случае, если мы не читали страницу 12 указанного выпуска.

Однако, прочитав эту страницу, мы узнаем, что первая сущность и первый объект, упомянутые на этой странице, – город Мадрид. Следовательно, теперь мы можем связать с упомянутой сущностью (объектом) более информативный тип *простр.об* («пространственный объект»).

Пусть E_3 – выражение «понятие с меткой AC060, определенное в Longman Dictionary of Scientific Usage (Moscow, Russky Yazik Publishers, 1989)». Не читая словаря, мы можем связать с понятием, упомянутым в E_3 , только тип [пон]. Но после того, как мы найдем определение с пометкой AC060, мы узнаем, что это определение понятия «трубка»

(полый цилиндр с длиной много больше диаметра). Следовательно, мы можем связать с понятием, упомянутым в ЕЗ, более информативный тип $\uparrow \text{физ.об}$ (обозначение понятия «физический объект»).

Будем предполагать, что цепочки $[\uparrow \text{сущн}]$, $[\uparrow \text{пон}]$, $[\uparrow \text{об}]$ – это типы семантических единиц, соответствующих словам «сущность», «понятие», «объект». Эти типы образуют множество специальных типов *Spectp*.

Условимся в последующих определениях считать символами как цепочки $[\text{сущн}]$, $[\text{пон}]$, $[\text{об}]$, $[\uparrow \text{сущн}]$, $[\uparrow \text{пон}]$, $[\uparrow \text{об}]$, так и элементы сортовых множеств.

Определение. Пусть S – сортовая система вида (1), $\text{Spectp} = \{[\uparrow \text{сущн}], [\uparrow \text{пон}]\}$, $[\uparrow \text{об}]$, $\text{Toptp} = \{[\text{сущн}], [\text{пон}]\}$, $[\text{об}]$. Тогда через $\text{Tp}(S)$ обозначим наименьшее множество T , удовлетворяющее следующим условиям:

- (1) $\text{Spectp} \cup \text{Toptp} \cup \text{St} \cup \{-s \mid s \in \text{St}\} \subseteq T$; элементы множеств *Spectp* и *Toptp* называются специальными типами и верхними типами соответственно;
- (2) Если $k > 1$, для $\forall i = 1, \dots, k$ $s_i \in \text{St}$, $\forall i, j = 1, \dots, k$ из $i \neq j$ следует, что $s_i \perp s_j$ (т.е. цепочки s_i, s_j сравнимы для отношения совместимости ToI), то цепочка $s_1 * s_2 * \dots * s_k$ и цепочка $\uparrow s_1 * s_2 * \dots * s_k$ входят в T ;
- (3) Если $n > 1$, для $i = 1, \dots, n$ $t_i \in T \setminus \text{Spectp}$, то цепочка вида $(t_1, \dots, t_n)$ входит в T ;
- (4) Если $t \in T \setminus \text{Spectp}$, то цепочка $\{t\}$ входит в T ;
- (5) Если $t \in T \setminus (\text{Spectp} \cup \text{Toptp})$, и t начинается с символа '(' или '{', то цепочка $\uparrow t$ входит в T .

Множество $\text{Tp}(S)$ называется множеством типов, порождаемых с.с. S .

Определение. Если S – сортовая система, то $\text{Mtp}(S) = \text{Tp}(S) \setminus \text{Spectp}$; элементы множества $\text{Mtp}(S)$ называются основными типами (обозначение этого множества происходит от английского словосочетания main types).

2.6.2. Интерпретация определения множества типов

Сформулируем принципы установления соответствия между сущностями, рассматриваемыми в предметной области с с.с. S и типами из множества $\text{Mtp}(S)$. Типы понятий, в отличие от типов объектов, начинаются с символа ' $\uparrow$ '. С понятием, обозначаемым сортом s , свяжем тип $\uparrow s$. Тип $\{t\}$ соответствует любому множеству сущностей типа t .

Если $x_1, \dots, x_n$ – сущности типов $t_1, \dots, t_n$, тогда тип $(t_1, \dots, t_n)$ соответствует n -местному упорядоченному набору $(x_1, \dots, x_n)$. Как следствие, множества, состоящие из упорядоченных наборов с типом $(t_1, \dots, t_n)$, будут иметь тип $\{(t_1, \dots, t_n)\}$.

Пример 1. Можно связать типы из $Mtp(S)$ с некоторыми понятиями и объектами с помощью следующей таблицы:

ПОНЯТИЕ	ТИП
понятие «множество»	$\uparrow\{\{\text{сущн}\}\}$
понятие «множество объектов»	$\uparrow\{\{\text{об}\}\}$
понятие «множество понятий»	$\uparrow\{\{\text{пон}\}\}$
понятие «человек»	$\uparrow\text{интс}^*\text{дин.физ.об}$
Д.И. Менделеев	$\text{интс}^*\text{дин.физ.об}$
понятие «студенческая группа»	$\uparrow\{\text{интс}^*\text{дин.физ.об}\}$
Группа № 172	$\{\text{интс}^*\text{дин.физ.об}\}$
понятие «пара целых чисел»	$\uparrow(\text{цел}, \text{цел})$
пара (12, 144)	$(\text{цел}, \text{цел})$

Можно также связать с отношением «Меньше» на целых числах тип $\{(\text{цел}, \text{цел})\}$, с отношением «Принадлежать множеству» – тип $\{(\{\text{сущн}\}, \{\{\text{сущн}\}\})\}$, с отношением «Объект Y характеризуется понятием C » – тип $\{(\{\text{об}\}, [\text{пон}])\}$, а с отношением «Понятие D является обобщением понятия C » – тип $\{([\text{пон}], [\text{пон}])\}$.

Основная цель введения типов заключается в том, чтобы задавать семантические ограничения на атрибуты отношения, в частности, на аргументы и значения функций. Идею такого использования типов можно пояснить следующим образом. Пусть c – обозначение понятия (в частности, $c \in St$). Тогда через $Dt(c)$ будем обозначать множество всех сущностей, которые могут быть охарактеризованы понятием c , и называть $Dt(c)$ *денотатом* понятия c . Например, $Dt(\text{книга})$ = множество всех книг, $Dt(\text{человек})$ = множество всех людей.

Пусть R – обозначение n -арного отношения, $n > 1$. Тогда ограничение $(x_1, \dots, x_n) \in R \Leftrightarrow x_1 \in Dt(s_1), \dots, x_n \in Dt(s_n)$, где $s_1, \dots, s_n \in St$, будем указывать с помощью значения некоторого отображения tp , определенного для R , следующим образом: $tp(R) = \{(s_1, \dots, s_n)\}$. Аналогично, ограничение $(x_1, \dots, x_n) \in R \Leftrightarrow x_1 \in Dt(c_1), \dots, x_n \in Dt(c_n)$ будем представлять в виде $tp(R) = \{(tc_1, \dots, tc_n)\}$, где $tc_1, \dots, tc_n$ – типы, характеризующие сущности из рассматриваемой предметной области.

Пример 2. Семантические ограничения на атрибуты отношений *Брат*, *Расстояние* (последнее отношение является функцией, ставящей в соответствие двум пространственным объектам некоторое значение длины) можно представить следующим образом:

$$\begin{aligned} tp(\text{Брат}) &= \{(\text{интс} * \text{дин. физ. об}, \text{интс.} * \text{дин. физ. об})\}, \\ tp(\text{Расстояние}) &= \{(\text{простр. об}, \text{простр. об}, \text{дин})\}. \end{aligned}$$

2.6.3. Отношение конкретизации на множестве типов

Пусть S – произвольная сортовая система. Зададим на множестве типов $Tr(S)$ некоторое бинарное отношение, обозначаемое символом $\vdash$ и называемое *отношением конкретизации*. На множестве сортов St отношение $\vdash$ совпадает с отношением общности $\rightarrow$. Следующая система примеров демонстрирует требования к отношению $\vdash : [сущн] \vdash [об]$, $[сущн] \vdash [пон]$, $физ. об \vdash дин. физ. об$, $дин. физ. об \vdash интс * дин. физ. об$, $[пон] \vdash \uparrow интс$, $[пон] \vdash \uparrow интс * дин. физ. об$, $[об] \vdash физ. об$, $[об] \vdash \{физ. об\}$, $[об] \vdash \{(вещ, вещь)\}$.

Основная идея определения отношения конкретизации заключается в следующем. Мы хотим, чтобы расстояние могло быть определено и между неподвижными физическими объектами, и между динамическими физическими объектами, и между воображаемыми динамическими физическими объектами. Все объекты таких видов являются частными случаями пространственных объектов. Учитывая это, будем использовать отношение конкретизации $\vdash$ следующим образом.

Пусть R – обозначение n -арного отношения, где $n > 1$, и некоторое отображение tp ставит в соответствие R описание семантических ограничений на атрибуты $\{(t_1, \dots, t_n)\}$, т.е. $tp(R) = \{(t_1, \dots, t_n)\}$, где $n > 1$, $t_1, \dots, t_n \in Tr(S)$.

Будем предполагать, что выражение $R(x_1, \dots, x_n)$ выражает тот же смысл, что и выражение $(x_1, \dots, x_n) \in R$. Тогда будем считать выражение $R(x_1, \dots, x_n)$ допустимым $\Leftrightarrow$ существуют такие $u_1, \dots, u_n \in Tr(S)$, что $\forall k = 1, \dots, n \ t_k \vdash u_k$ и $x_k \in Dt(u_k)$, т.е. x_k входит в денотат понятия u_k .

Пример 3. Так как выполняются соотношения $простр. об \rightarrow вообр. простр. об$, $простр. об \rightarrow физ. об$, $физ. об \rightarrow дин. физ. об$, то $простр. об \vdash вообр. простр. об$, $простр. об \vdash дин. физ. об$.

Поэтому допустимыми будут являться выражения $Расст(x_1, x_2, l_1)$, $Расст(z_1, z_2, l_2)$, где x_1, x_2 – обозначения двух автомобилей, z_1, z_2 – обозначения орбит двух конкретных небесных тел, и l_1, l_2 – обозначения некоторых значений длины.

Пример 4. Студенческие учебные группы являются частными случаями множеств. Семантические ограничения на аргумент и значение

функции «Количество элементов множества», обозначаемой символом *Колич-эле*, можно задать соотношением $tp(\text{Колич-эле}) = \{(\{[сущн]\}, \text{нат})\}$. Предположим, что база знаний интеллектуальной системы включает идентификатор студенческой группы $Гр\#172$, и отображение tp связывает с этим идентификатором тип $\{интс*дин.физ.об\}$. Таким образом, данная гипотетическая интеллектуальная система рассматривает объект, обозначаемый идентификатором $Гр\#172$, как некоторое множество людей (каждый человек является как интеллектуальной системой, так и динамическим физическим объектом). Пусть $tp(14) = \text{нат}$. Так как $\text{нат} \rightarrow \text{нат}$, то в случае выполнения соотношения $\{[сущн]\} \vdash \{интс*дин.физ.об\}$ выражение *Колич-эле*($Гр\#172, 14$) допустимо.

Определение 3. Пусть S – произвольная сортовая система вида (2.5.1). Тогда элементарными составными типами будем называть цепочки из $Tr(S)$ вида $s_1 * s_2 * \dots * s_k$, где $k > 1$, для $\forall i = 1, \dots, k$ $s_i \in St$.

Пример 5. Цепочка $интс*дин.физ.об$ является элементарным составным типом для сортовой системы S_0 .

Определение 4. Пусть S – сортовая система вида (2.5.1). Тогда через $Elt(S)$ обозначим объединение множества сортов St с множеством всех элементарных составных типов. Элементы множества Elt будем называть элементарными типами.

Определение 5. Если S – сортовая система вида (2.5.1), $t \in Elt(S)$, то спектр типа t , обозначаемый через $Spr(t)$, в случае $t \in St$ является множеством $\{t\}$, а в случае $t = s_1 * s_2 * \dots * s_k$, где $k > 1$, для $\forall i = 1, \dots, k$ $s_i \in St$, является множеством $\{s_1, \dots, s_k\}$.

Пример 6. Для сортовой системы S_0 спектр $Spr(физ.об) = \{физ.об\}$, $Spr(интс*дин.физ.об) = \{интс, дин.физ.об\}$.

Определение 6. Пусть S – с.с. вида (2.5.1), $u \in St$, t – элементарный составной тип из $Tr(S)$. Тогда тип t называется уточнением сорта $u \Leftrightarrow$ когда спектр $Spr(t)$ включает такой сорт w , что $u \rightarrow w$ (т.е. $(u, w) \in Gen$).

Пример 7. Пусть $u = физ.об$, $t = интс*дин.физ.об$. Тогда спектр $Spr(t) = \{интс, дин.физ.об\}$. Поэтому из $физ.об \rightarrow дин.физ.об$ вытекает, что t – уточнение сорта u . Напомним, что в данной работе сорта считаются символами, т.е. неделимыми единицами.

Определение 7. Пусть $u \in St$, $t \in Tr(S)$, и t включает символ u . Тогда вхождение символа u называется свободным $\Leftrightarrow$ когда либо

$t = u$, либо это вхождение u в t не является вхождением в какую-либо подцепочку вида $s_1 * s_2 * \dots * s_k$, где $k > 1$, для $\forall i = 1, \dots, k$ $s_i \in St$, и существует такое m , $1 \leq m \leq k$, что $u = s_m$.

Пример 8. С функцией «Друзья» можно связать тип $t1 = \{(\text{интс} * \text{дин. физ. об.}, \{\text{интс} * \text{дин. физ. об.}\})\}$. Как первое, так и второе вхождения в цепочку $t1$ символа дин. физ. об. не являются свободными вхождениями. С функцией «Вес множества физических объектов» можно ассоциировать тип $t2 = \{(\{\text{физ. об.}\}, (\text{цел}, \text{кг}))\}$; вхождения символа физ. об. в $t2$ и в $t3 = \uparrow \text{физ. об.}$ (возможный тип понятия «физический объект») являются свободными.

Определение 8. Пусть S – с.с. вида (2.5.1), тогда $Tc(S) = \{t \in Tp(S) \mid (Spectp \cup Toptr) \mid t \text{ начинается с символа } \uparrow\}$, где $Spectp = \{[\uparrow \text{сущн.}], [\uparrow \text{пон.}], [\uparrow \text{об.}]\}$, $Toptr = \{[\text{сущн.}], [\text{пон.}], [\text{об.}]\}$; $Tob = Tp(S) \setminus (Spectp \cup Toptr \cup Tc(S))$.

Элементы $Tc(S)$ интерпретируются как типы понятий (кроме наиболее общего типа $[\text{сущн.}]$). Элементы $Tob(S)$ интерпретируются как типы объектов (сущности, не рассматриваемые как понятия).

Определение 9. Пусть S – с.с. вида (2.5.1). Тогда преобразования $tr_1, \dots, tr_6$, частично применимые к элементам из $Tp(S)$, задаются следующим образом:

1. Если $t \in Tp(S)$, t включает символ $[\text{сущн.}]$, то tr_1 и tr_2 применимы к t . Пусть $w1$ – результат замены в t произвольного вхождения символа $[\text{сущн.}]$ на символ $[\text{пон.}]$; $w2$ – результат замены в t произвольного вхождения символа $[\text{сущн.}]$ на символ $[\text{об.}]$. Тогда $w1$ и $w2$ – возможные результаты применения к цепочке t преобразований tr_1 и tr_2 соответственно.

2. Если $t \in Tp(S)$, t включает символ $[\text{пон.}]$, $u \in Tc(S)$, то tr_3 применимо к t , и результат замены произвольного вхождения символа $[\text{пон.}]$ на u является возможным результатом применения преобразования tr_3 к типу t .

3. Если $t \in Tp(S)$, t включает символ $[\text{об.}]$, $z \in Tob(S)$, то tr_4 применимо к t , и результат замены в t произвольного вхождения символа $[\text{об.}]$ на тип z является возможным результатом применения преобразования tr_4 к типу t .

4. Если $t \in Tp(S)$, t включает символ $s \in St$, $u \in St$, $(s, u) \in Gen$, то тип, получающийся из t заменой какого-либо свободного вхождения s на сорт u , является возможным результатом применения преобразования tr_5 к типу t .

5. Если $t \in Tr(S)$, $u \in St$, z – элементарный составной тип из $Tr(S)$, являющийся уточнением сорта u , w получается из t заменой произвольного свободного вхождения сорта u в цепочку t на цепочку z , то w – возможный результат применения преобразования tr_6 к цепочке t .

Пример 9. Если S_0 – построенная ранее с.с., $t1 = [об]$, $t2 = \text{про-стр.об}$, $w1 = \text{интс} * \text{дин.физ.об}$, $w2 = \text{дин.физ.об}$, то $w1$ и $w2$ – возможные результаты применения преобразования tr_4 и tr_5 к $t1$ и $t2$, соответственно. Если $t3 = \{\text{физ.об}\}$, $w3 = \{\text{интс} * \text{дин.физ.об}\}$, то $w3$ – возможный результат применения tr_6 к $t3$.

Определение 10. Пусть S – с.с. вида (2.5.1), $t, u \in Tr(S)$. Тогда тип u называется конкретизацией типа t , а тип t называется обобщением типа u (обозначается через $t \vdash u$), если либо $t = u$, либо найдутся такие $x_1, \dots, x_n \in Tr(S)$, где $n > 1$, что $x_1 = t$, $x_n = u$, и для $i = 1, \dots, n-1$ найдется такое $k[i] \in \{1, \dots, 6\}$, что преобразование $tr_{k[i]}$ применимо к x_i , и x_{i+1} является возможным результатом применения преобразования $tr_{k[i]}$ к x_i .

Пример 10. Для с.с. S_0 легко проверить, что $[сущи] \vdash [нон]$, $[сущи] \vdash [об]$, $[об] \vdash \text{интс}$, $\text{интс} \vdash \text{интс} * \text{физ.об}$, $\text{физ.об} \vdash \text{дин.физ.об}$, $[об] \vdash \{\text{интс}\}$, $\{\text{интс}\} \vdash \{\text{интс} * \text{дин.физ.об}\}$, $[об] \vdash (\text{вещ}, \text{вещ})$, $[об] \vdash \{(\text{вещ}, \text{вещ})\}$, $[нон] \vdash \uparrow \text{интс}$, $[нон] \vdash \uparrow \text{интс} * \text{дин.физ.об}$, $[нон] \vdash \uparrow \{\text{интс} * \text{дин.физ.об}\}$.

Утверждение 2.1. Пусть S – произвольная сортовая система. Тогда отношение конкретизации $\vdash$ на множестве типов $Tr(S)$ является частичным порядком.

Доказательство. Рефлексивность и транзитивность отношения $\vdash$ следуют непосредственно из определения. Антисимметричность вытекает из свойств преобразований $tr_1, \dots, tr_6$. В результате применения преобразования tr_1 или tr_2 количество вхождений символа $[сущи]$ уменьшается на 1. После применения преобразования tr_3 или tr_4 на 1 уменьшается количество вхождений символа $[нон]$ или символа $[об]$, соответственно. Если $t1, t2 \in Tr(S)$ и тип $t2$ получен из $t1$ в результате однократного применения преобразования tr_5 , то это означает, что найдутся такие $s, u \in St$, что $s \neq u$, $(s, u) \in \text{Gen}$, тип $t1$ включает символ s , и тип $t2$ получается заменой некоторого вхождения символа s в $t1$ на символ u . Из антисимметричности отношения Gen на St следует, что обратное преобразование $t2$ в $t1$ невозможно. Если тип $t2$ получен из типа $t1$ однократным применением преобразования tr_6 , то количество символов в $t2$ больше, чем количество символов в $t1$.

2.7. Концептуально-объектные системы

Предположим, что для описания какой-то предметной области мы выбрали некоторую сортовую систему S вида (2.5.1). Тогда на следующем шаге выберем некоторое множество X , состоящее из таких элементарных информационных единиц, с помощью которых мы будем описывать сообщения, команды и вопросы, относящиеся к рассматриваемой области; это множество X будем называть *первичным информационным универсумом*. Затем выберем V – некоторое счетное множество символов, называемых *переменными* и используемых в качестве меток разнообразных сущностей, в том числе в качестве меток СП текстов и фрагментов СП текстов.

Далее зададим отображение $tp: X \cup V \rightarrow Tp(S)$ из объединения $X \cup V$ в множество типов, порождаемых с.с. S , тогда каждая переменная и каждая сущность получают тип. На последнем шаге выделим некоторое подмножество F множества X так, что элементы F будут являться обозначениями функций, рассматриваемых в данной ПО. Тогда набор (X, V, tp, F) будет являться концептуально-объектной системой, согласованной с сортовой системой S .

Определение 1. Пусть S – сортовая система вида (2.5.1). Тогда произвольную упорядоченную четверку Ct вида

$$(X, V, tp, F) \quad (2.7.1)$$

назовем *концептуально-объектной системой (к.о.с.), согласованной с системой S* (или *к.о.с. для S*) у когда выполняются следующие условия:

- (1) X, V – счетные непересекающиеся множества символов; tp – отображение вида $X \cup V \rightarrow Tp(S)$;
- (2) $F \subset X$, для каждого $r \in F$ цепочка $tp(r)$ начинается с подцепочки ‘(’ и заканчивается подцепочкой ‘)’;
- (3) $St \subset X$, и для любого $s \in St$ $tp(s) = \uparrow s$;
- (4) $\{v \in V \mid tp(v) = [сущи]\}$ – счетно.

Множество X называется *первичным информационным универсумом*, элементы множеств V и F называются соответственно *переменными* и *функциональными символами*. Если элемент $d \in X \cup V$, $tp(d) = t$, то будем говорить, что t – тип элемента d .

Пример. Построим некоторую к.о.с. Ct_0 для с.с. S_0 . Пусть N – множество всех цепочек из цифр ‘0’, ‘1’, ..., ‘9’, таких, что если первый символ цепочки 0, то и вся цепочка – 0. Будем полагать, что символы *чел*, *химик*, *биолог*, *студ.гр*, *тур.гр*, *П.Сомов*, *А.Зубов*, *И.Семенов*,

Друзья, Колич, Меньше, Знает, Явл1 Сейчас, Раньше, Включить1, Элем являются соответственно обозначениями понятий «человек», «химик», «биолог», «студенческая группа», «туристическая группа», трех конкретных людей, функции «Друзья», функции «Количество элементов множества», отношения «Меньше» на множестве вещественных чисел, отношения «В памяти некоторой интеллектуальной системы X1 в момент времени X2 имеется концептуальное представление некоторого сообщения X3», отношения «Некоторый объект X1 характеризуется понятием X2» (пример реализации в тексте: «П.Сомов является химиком»), текущего момента времени, отношения «Раньше» на множестве моментов времени, отношения «Некоторая интеллектуальная система X1 включает сущность X2 в момент X3 в состав множества сущностей X4», отношения «Элемент множества», отношения «Подмножество». Символ *понятие* будем интерпретировать как информационную единицу, соответствующую словам «понятие» и «концепт».

Пусть $U1 = \{\text{чел, химик, биолог, студ.гр, тур.гр, П.Сомов, А.Зубов, И.Семенов, Друзья, Колич, Меньше, Знает, Явл1, сейчас, Раньше, Включить1, Элем, понятие}\}$.

Зададим отображение $t1$ из $U1$ в $Tr(S_0)$ следующей таблицей:

Таблица 2.1

Примеры соответствий между сущностями и типами

x	$t1(x)$
чел, химик, биолог	$\uparrow_{интс*дин.физ.об}$
студ.гр, тур.гр	$\uparrow\{интс*дин.физ.об\}$
П.Сомов, А.Зубов, И.Семенов	$интс*дин.физ.об$
Друзья	$\{(интс*дин.физ.об, \{интс*дин.физ.об\})\}$
Колич	$\{([суцн]), nat)\}$
Меньше	$\{(вещ, вещь)\}$
Знает	$\{(интс, мом, сообщ)\}$
Явл1	$\{([об], [пон])\}$
сейчас	$мом$
Раньше	$\{(мом, мом)\}$
Включить1	$\{(интс, [суцн], мом, \{[суцн]\})\}$
Элем	$\{([суцн], \{[суцн]\})\}$
понятие	$\uparrow[пон]$

Будем полагать, что $АО_«Салют»$, $АО_«Старт»$, $НПО_«Радуга»$ являются обозначениями организаций, $Поставщики$, $Персонал$, $Директор$ – обозначения функций «Множество всех поставщиков данной организации», «Множество всех сотрудников данной организации» и «Директор данной организации», соответственно. Пусть $U2 = \{АО_«Салют»\}$, $АО_«Старт»$, $НПО_«Радуга»$, $Поставщики$, $Персонал$, $Директор\}$, и отображение $t2$ из $U2$ в $Tr(S_0)$ задается следующими условиями:

$$t2(АО_«Салют») = t2(АО_«Старт») = t2(НПО_«Радуга») = \\ = орг*простр.об*интс,$$

$$t2(Поставщики) = \{(орг, \{орг\})\},$$

$$t2(Персонал) = \{(орг, \{интс*дин.физ.об\})\},$$

$$t2(Директор) = \{(орг, интс*дин.физ.об\})\}.$$

Пусть $Vx = \{x1, x2, \dots\}$, $Ve = \{e1, e2, \dots\}$, $Vp = \{P1, P2, \dots\}$, $Vset = \{S1, S2, \dots\}$, $V_0 = Vx \cup Ve \cup Vp \cup Vset$, $X_0 = St_0 \cup N \cup U1 \cup U2$, и отображение $tp_0: X_0 \cup V_0 \rightarrow Tr(S_0)$ задается следующими соотношениями:

$$d \in St_0 \Rightarrow tp_0(d) = \uparrow d; d \in Nat \Rightarrow tp_0(d) = nam;$$

$$d \in U1 \Rightarrow tp_0(d) = t1(d); d \in U2 \Rightarrow tp_0(d) = t2(d);$$

$$d \in Vx \Rightarrow tp_0(d) = [сущн]; d \in Ve \Rightarrow tp_0(d) = cum;$$

$$d \in Vp \Rightarrow tp_0(d) = сообщ, d \in Vset \Rightarrow tp_0(d) = \{\{сущн\}\}.$$

Пусть $F_0 = \{Друзья, Колич, Поставщики, Персонал, Директор\}$, $Ct_0 = (X_0, V_0, tp_0, F_0)$, тогда нетрудно проверить, что Ct_0 – концептуально-объектная система, согласованная с сортовой системой S_0 .

2.8. Системы кванторов и логических связей.

Концептуальные базисы

Предположим, что мы определили сортовую систему S вида (2.5.1) и концептуально-объектную систему Ct вида (2.7.1) для описания рассматриваемой предметной области. Тогда предлагается выделить в первичном информационном универсуме X два непересекающихся и конечных (следовательно, непустых) подмножества Int_1 и Int_2 следующим образом: выделим в St два сорта int_1 и int_2 и предположим, что для $m = 1, 2$ $Int_m = \{x \in X \mid tp(x) = int_m\}$. Элементы Int_1 соответствуют значениям выражений «каждый», «какой-то», «некоторый», «произвольный» и т.д. в случаях, когда эти выражения являются частями групп слов, и эти группы связаны с единственным числом. Элементы Int_2 интерпретируются как семантические единицы, соответствующие

выражениям «все», «несколько», «почти все», «многие» и т.д.; минимальное требование к Int_2 заключается в том, чтобы Int_2 содержало семантическую единицу, соответствующую слову «все». Пусть Int_1 содержит выделенный элемент ref , рассматриваемый как аналог слова «некоторый» в смысле «какой-то вполне определенный» (но, возможно, неизвестный). Если Ct – к.о.с. вида (2.7.1), $d \in X$, d обозначает понятие, и семантическое представление (СП) текста включает подцепочку вида $ref\ d$ (например, цепочку *нек человек*, где $ref = нек$, $d = человек$), тогда будем полагать, что эта подцепочка обозначает некоторую конкретную сущность (но не произвольную), которая характеризуется понятием d .

Кроме того, будем предполагать, что X содержит элементы ' $\equiv$ ', ' $\neg$ ', ' $\wedge$ ', ' $\vee$ ', понимаемые как связи «тождественно», «не», «и», «или», и элементы $\forall$ и $\exists$, понимаемые как квантор всеобщности и квантор существования. Наконец, будем считать, что множество St включает выделенные сорта eqv , neg , $binlog$, ext , интерпретируемые, соответственно, как типы (а) связи ' $\equiv$ ', (б) связи отрицания ' $\neg$ ' (в) связок ' $\wedge$ ', ' $\vee$ ' (конъюнкция и дизъюнкция), (г) квантора всеобщности и квантора существования.

Эти предположения в наглядной форме отражает рисунок 2.1. На этом рисунке [*сущн*] – это тип «сущность»; элементы *интс*, *дин.физ.об*, *нат*, *сит*, *сообщ* – сорта «интеллектуальная система», «динамический физический объект», «натуральное число», «ситуация», «смысл сообщения»; *чел*, *нек*, *произвол*, *определ*, *все*, *нескол* – информационные единицы, соответствующие словам «человек», «некоторый» («некоторая», «некоторое»), «произвольный» («любой»), «определенный», «все», «несколько»; *Колич* – обозначение функции «Количество элементов множества». Элемент *нек* интерпретируется как квантор референтности.

Определение 1. Пусть S – сортовая система вида (2.5.1), Ct – концептуально-объектная система вида (2.7.1) для S , $ref \in X$, различные элементы int_1 , int_2 , eqv , neg , $binlog$, ext – некоторые выделенные сорта из $St \setminus \{P\}$, и каждая пара их несравнима для отношения общности Gen и несравнима для отношения совместимости Tol . Тогда упорядоченная семерка Ql вида

$$(int_1, int_2, ref, eqv, neg, binlog, ext) \quad (2.8.1)$$

называется *системой кванторов и логических связок* (с. к. л. с.) для S и $Ct \Leftrightarrow$ когда выполнены следующие условия:

1. Для каждого $m = 1, 2$ множество $Int_m = \{x \in X \mid tp(x) = int_m\}$ конечно; $ref \in Int_1$; Int_1 и Int_2 не пересекаются.

2. $\{\equiv, \neg, \wedge, \vee, \forall, \exists\} \subset X$; кроме того, $tp(\equiv) = eqv$, $tp(\neg) = neg$, $tp(\wedge) = tp(\vee) = binlog$, $tp(\forall) = tp(\exists) = ext$.
3. Не найдется такого $d \in X \setminus (Int1 \cup Int2 \cup \{\equiv, \neg, \wedge, \vee, \forall, \exists\})$ и такого $s \in \{int_1, int_2, eqv, neg, binlog, ext\}$, что $tp(d)$ и s сравнимы для отношения общности Gen или сравнимы для отношения совместимости Tol .
4. Для каждого $u \in \{int_1, int_2, eqv, neg, binlog, ext\}$ сорт u и сорт P несравнимы для отношения общности Gen и несравнимы для отношения совместимости Tol .
5. Для каждого $u \in \{int_1, int_2, eqv, neg, binlog, ext\}$ сорт u и сорт P несравнимы для отношения общности Gen и несравнимы для отношения совместимости Tol .

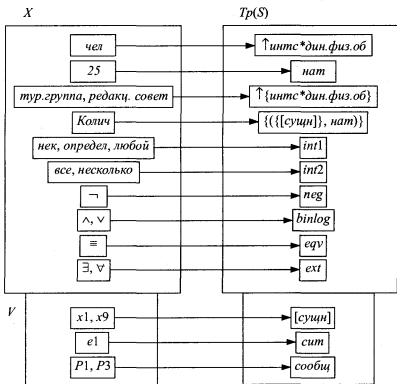


Рис. 2.1. Иллюстрация соответствия между элементами первичного информационного универсума X , переменными из множества V и их типами из множества $Tp(S)$, где S – сортовая система

Элементы Int_1 и Int_2 называются интенциональными кванторами, ref называется квантором референтности, $\forall$ и $\exists$ называются экстенциональными кванторами.

Пример 1. Пусть $S_0 = (St_0, сообш, Gen_0, Tol_0)$ – сортовая система, построенная в разделе 2.5, $Ct_0 = (X_0, V_0, tp_0, F_0)$ – концептуально-объектная система для S_0 , построенная в разделе 2.7, $Str = \{кв.инт1, кв.инт2, экв, не, бин, экст\}$, $Gen_1 = Gen_0 \cup \{(s, s) | s \in Str\}$, $St_1 = St_0 \cup Str$, $S_1 = (St_1, сообш, Gen_1, Tol_0)$. Тогда, очевидно, S_1 – сортовая система.

Определим теперь некоторую концептуально-объектную систему Ct_1 и систему кванторов и логических связок Ql_1 . Пусть $Z = \{нек, все, \equiv, \neg, \wedge, \vee, \forall, \exists\}$, $X_1 = X_0 \cup Str \cup Z$. Зададим отображение tp_1 из $X_1 \cup V_0$ в $Tr(S_1)$ следующим образом:

$$u \in Str \Rightarrow tp_1(u) = \uparrow u; d \in X_0 \Rightarrow tp_1(u) = tp_0(u);$$

$$tp_1(нек) = кв.инт1, tp_1(все) = кв.инт2, tp_1(\equiv) = экв, tp_1(\neg) = не,$$

$$tp_1(\wedge) = tp_1(\vee) = бин, tp_1(\forall) = tp_1(\exists) = экст.$$

Пусть $Ct_1 = (X_1, V_0, tp_1, F_0)$, $Ql_1 = (кв.инт1, кв.инт2, нек, экв, не, бин, экст)$. Тогда легко проверить, что Ct_1 – к.о.с. для S_1 , Ql_1 – с.к.л.с. для S_1 . В системе Ql_1 информационная единица *нек* интерпретируется как квантор референтности *ref* (т.е. как обозначение значения слов «некоторый», «некоторая», «некоторое»).

Определение 2. Упорядоченная тройка B вида

$$(S, Ct, Ql) \tag{2.8.2}$$

называется *концептуальным базисом* (к.б.) $\Leftrightarrow S$ – с.с., Ct – к.о.с. вида (2.7.1) для S , Ql – система кванторов и логических связок (с.к.л.с.) для S и Ct , и $(X \cup V) \cap \{', (' , ')', ':', '*', '<', '>', '&'\} = \emptyset$.

Обозначим через $S(B)$, $Ct(B)$, $Ql(B)$ компоненты произвольного к.б. вида (2.8.2). Каждый компонент h систем видов (2.5.1), (2.7.1), (2.8.1) будет обозначаться через $h(B)$. Например, сорт «смысл сообщения» будет обозначаться через $P(B)$, первичный информационный универсум X через $X(B)$, множество переменных V через $V(B)$, множество функциональных символов F через $F(B)$.

Каждый концептуальный базис будет интерпретироваться как формальное перечисление:

(а) первичных единиц, необходимых для построения семантических представлений (СП) ЕЯ-текстов, для описания знаний о реальности и для представления целей интеллектуальных систем;

(б) информации, связанной с этими единицами и необходимой для построения СП текстов, формирования фрагментов знаний и представления целей интеллектуальных систем.

Пример 2. Пусть St_1, Ct_1, Ql_1 – соответственно с.с., к.о.с., с.к.л.с., определенные выше. Тогда очевидно, что упорядоченная тройка $B_1 = (St_1, Ct_1, Ql_1)$ является концептуальным базисом и $St(B_1) = St_1, P(B_1) = сообщ, X(B_1) = X_1, V(B_1) = V_0$.

Из этого примера следует, что множество всех концептуальных базисов не является пустым, поскольку мы построили формальный объект B_1 , являющийся концептуальным базисом.

Введенные понятия дадут возможность в главе 3 для каждого к.б. B задать множество формул $Ls(B)$, удобных для описания содержания (т.е. структурированных значений) ЕЯ-текстов, представления знаний о мире и целей интеллектуальных систем. Множество формул $Ls(B)$ будет названо СК-языком (стандартным концептуальным языком), порождаемым базисом B .

2.9. Обсуждение разработанной математической модели для описания системы первичных единиц концептуального уровня

2.9.1. Особенности модели с математической точки зрения

По своей форме разработанная математическая модель для описания системы первичных единиц концептуального уровня, используемых многоагентной системой, является оригинальной. Рассмотрим отличительные черты построенной модели, представляющиеся наиболее важными как с математической точки зрения, так и с точки зрения использования модели при проектировании ЛП.

1. Конструктивно учитывается существование иерархии понятий: для этого на множестве сортов St задается частичный порядок Gen , называемый отношением общности.

2. Многие сущности, рассматриваемые в той или иной предметной области, могут быть охарактеризованы с разных точек зрения. Например, люди являются, с одной стороны, интеллектуальными системами (поскольку могут читать, решать задачи и т.д.), но, с другой стороны, являются физическими объектами, способными перемещаться в пространстве. Поэтому многие понятия как бы имеют «координаты» по разным «семантическим осям». Для учета этого важного явления в модель вводится бинарное отношение Tol (отношение совместимости, или толерантности) на множестве сортов. Накопленный опыт показал, что эта оригинальная черта модели является чрезвычайно важной для разработки алгоритмов семантико-синтаксического анализа текстов: дело

в том, что появление одного и того же слова в нескольких контекстах может объясняться реализацией в этих контекстах различных «семантических координат» данного слова.

3. Фраза «это понятие используется в физике и химии» (относящаяся, например, к понятию «молекула») для человека, владеющего русским языком, является очень простой. Между тем смысловая структура данной фразы не может быть адекватно отображена средствами основных известных подходов к формализации семантики ЕЯ. Причина заключается в том, что такие подходы не предлагают формального аналога информационной единицы, соответствующей слову «понятие». Модель, построенная выше, во-первых, рассматривает специальный базовый тип $[\uparrow\text{пон}]$, интерпретируемый как тип информационной единицы, соответствующей слову «понятие». Во-вторых, компонент концептуально-объектной системы Ct вида (2.7.1), обозначаемый через X и называемый первичным информационным универсумом, может включать символ, интерпретируемый как информационная единица, соответствующая слову «понятие» (см. пример в разделе 2.7). Данная черта модели важна для разработки ЛП, обрабатывающих научные и научно-технические тексты, а также ЛП прикладных интеллектуальных систем, извлекающих знания из энциклопедических словарей или пополняющих электронные энциклопедические словари.

4. Одной из наиболее важных отличительных черт построенной модели является оригинальное определение множества типов, порождаемого произвольной сортовой системой, где типы рассматриваются как формальные характеристики сущностей, относящихся к выбранной предметной области. В соответствии с этим определением, (а) типы объектов из предметной области по своей форме отличаются от типов понятий, квалифицирующих данные объекты, (б) типы объектов по своей форме отличаются от типов множеств, состоящих из таких объектов, (в) типы понятий, обозначающих объекты, по своей форме отличаются от типов понятий, квалифицирующих множества данных объектов (например, тип понятия «человек» отличается от типа понятий «ученый совет», «студенческая группа») и т.д.

5. Модель связывает типы и с именами функций. При этом определение множества типов позволяет разумным образом связать типы с целым рядом довольно нестандартных, но практически важных функций. В частности, к ним относятся функции, значениями которых являются: (а) множество понятий, поясняемых в энциклопедическом словаре, (б) множество понятий, входящих в определение данного понятия в данном словаре, (в) множество известных определений данного по-

нения, (г) количество элементов данного множества, (д) множество поставщиков данного предприятия, (е) множество сотрудников данной организации.

2.9.2. Сравнение модели с другими подходами к описанию первичных единиц концептуального уровня

Сравним построенную модель с подходами к описанию первичных единиц концептуального уровня, предлагаемыми логикой предикатов первого порядка, теорией представления дискурсов, теорией обобщенных кванторов, теорией концептуальных графов и эпизодической логикой.

В стандартной логике предикатов рассматриваются неструктурированные множества констант, функциональных символов и предикатных символов. В многосортных логиках предикатов множество констант разбито на непересекающиеся классы, каждый из которых характеризуется некоторым сортом. Разработанная выше модель предоставляет, в частности, следующие дополнительные возможности по сравнению с многосортными логиками предикатов: (1) благодаря введению отношения совместимости в качестве компонента сортовой системы, с первичной единицей концептуального уровня можно связать не только один, но и, во многих случаях, несколько сортов, как бы являющихся «координатами по ортогональным семантическим осям» сущностей, квалифицируемых или обозначаемых такими единицами; (2) «привязывание» типов к первичным информационным единицам означает, что множество таких единиц обладает развитой структурой; в частности, типы позволяют формально различать (а) типы объектов из предметной области и типы понятий, квалифицирующих данные объекты, (б) типы объектов и типы множеств, состоящих из таких объектов, (в) типы понятий, обозначающих объекты, и типы понятий, квалифицирующих множества данных объектов, (3) рассмотрение единиц концептуального уровня, соответствующих словам «некоторый», «определенный», «какой-нибудь», «все», «большинство», «несколько».

Кроме того, рамки построенной модели позволяют рассматривать функции, аргументами и/или значениями которых могут быть семантические представления (СП) высказываний и повествовательных текстов. Например, такая функция может ставить в соответствие каждому понятию, определяемому в энциклопедическом словаре, формулу – СП определения данного понятия. Между тем в логике предикатов первого порядка аргументами и значениями функций могут быть только

термы, но не формулы. Аналогичные ограничения должны выполняться и для атрибутов отношений, т.е. для аргументов предикатов.

Теорию представления дискурсов (ТПД) можно рассматривать как один из вариантов логики предикатов первого порядка, сочетающий использование формул и двумерных диаграмм для более наглядного представления информации. Поэтому перечисленные преимущества разработанной модели для описания системы первичных единиц концептуального уровня, используемых ЛП, относятся и к ТПД.

В теории обобщенных кванторов (ТОК), как и в построенной модели, рассматриваются единицы концептуального уровня, соответствующие словам «некоторый», «определенный», «все», «большинство», «несколько». Однако все остальные перечисленные преимущества модели по сравнению с логикой предикатов первого порядка являются одновременно и преимуществами по сравнению с подходом ТОК.

В отличие от логики предикатов первого порядка, нотация теории концептуальных графов (ТКГ) позволяет различать обозначения конкретных объектов (конкретных компьютеров, предприятий, городов и т.д.) и обозначения понятий, квалифицирующих эти объекты («компьютер», «предприятие», «город»). Остальные же перечисленные выше свойства модели являются преимуществами и по сравнению с ТКГ.

Наконец, указанные преимущества по сравнению с логикой предикатов первого порядка являются одновременно и преимуществами по сравнению с подходом к структурированию совокупности первичных единиц концептуального уровня, предлагаемым эпизодической логикой.

Очевидно, что если какая-либо модель предназначена как для описания системы первичных единиц концептуального уровня, используемых ЛП, так и для представления информации, связанной с этими единицами и определяющей возможности их соединения в правильные составные структуры, то подобная модель предлагает некоторый способ концептуальной структуризации рассматриваемых предметных областей. Поэтому на основании проведенного анализа можно прийти к заключению, что построенная выше модель предлагает более «тонкоячеистую» структуризацию предметных областей по сравнению с основными известными подходами к формализации семантики ЕЯ, значительно увеличивает «разрешающую способность» формального инструментария, предназначенного для исследования различных предметных областей.

В конце 1990-х – начале 2000-х гг. большую актуальность приобрели исследования по разработке онтологий различных предметных об-

ластей, т.е. по созданию формальных описаний систем понятий, относящихся к выбранной области, вместе с их определениями и фрагментами знаний, «привязанных» к понятиям. Первым шагом в каждом проекте такого рода является выбор некоторой начальной (другими словами, базовой) структуризации рассматриваемой предметной области или группы областей.

Представляется, что построенная в данной главе математическая модель для описания системы первичных единиц концептуального уровня, используемых многоагентной системой, и для представления информации, связанной с этими единицами, может найти применение в проектах разработки более совершенных онтологий в произвольных предметных областях, поскольку разработанная модель предлагает формальный инструментарий с наибольшей «разрешающей способностью» по сравнению с другими известными подходами к формализации семантики ЕЯ и, как следствие, к концептуальной структуризации предметных областей.

Глава 3

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ДЛЯ ОПИСАНИЯ СТРУКТУРИРОВАННЫХ ЗНАЧЕНИЙ ПРЕДЛОЖЕНИЙ И СВЯЗНЫХ ТЕКСТОВ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ

3.1. Постановка задачи

В разделе 2.1. была обоснована актуальность разработки широко применимых формальных языков для построения семантических представлений (СП) ЕЯ-текстов, выразительные возможности которых позволяют отображать многие особенности поверхностной структуры предложений и связных ЕЯ-текстов. В качестве первого шага на пути разработки определения такого класса формальных языков в главе 2 был определен класс формальных объектов, называемых концептуальными базисами.

В данной главе ставится задача разработки математической модели для описания структурированных значений предложений и связных естественно-языковых текстов.

По своей форме модель должна являться описанием такого соответствия между произвольным концептуальным базисом B и некоторым множеством формул $Forms(B)$, чтобы класс формальных языков $\{Forms(B), \text{ где } B - \text{к.б.}\}$ был удобен для построения СП фраз и связных ЕЯ-текстов, отражающих многие особенности поверхностной структуры текстов.

С целью выработки критериев для построения такой модели был проведен системный анализ структурных особенностей (а) текстов на русском, английском, немецком и французском языках, (б) ряда искусственных языков, используемых для построения семантических представлений текстов лингвистическими процессорами, (в) выражений искусственных языков представления знаний в прикладных интеллектуальных системах (в частности, терминологических языков представления знаний).

Проведенный анализ показал, что есть несколько важных аспектов формализации семантики ЕЯ, которые до недавнего времени недооце-

нивались или игнорировались большей частью исследователей. В частности, это относится к формальному исследованию смысловых структур (а) повествовательных текстов, включающих описания множеств; (б) дискурсов со ссылками на смысл предложений и более крупных частей текста; (в) фраз, где логические связки «и», «или» используются нетрадиционными способами и соединяют не фрагменты, выражающие высказывания, а описания объектов, множеств, понятий; (г) фраз с придаточными определительными и причастными оборотами; (д) фраз со словами «понятие», «термин».

Кроме того, несколько наиболее популярных подходов к математическому изучению семантики ЕЯ не принимают или недостаточно принимают во внимание роль знаний о мире в понимании ЕЯ и, следовательно, не изучают проблем формального описания фрагментов знаний (определения понятий и т.д.). Например, это относится к весьма популярной теории представления дискурсов (Kamp, 1981; Kamp, Reyle, 1993, 1996; van Eijck, Kamp, 1996).

Надо добавить, что тексты имеют авторов, могут быть опубликованы теми или другими источниками, могут вводиться с того или другого терминалов и т.д. Информация об этих внешних связях текстов может быть важна для их смысловой интерпретации. Поэтому целесообразно рассматривать текст как некий структурированный объект, обладающий поверхностной структурой T , множеством значений S (в большинстве случаев S состоит из одного значения), соответствующим T , и некоторыми значениями $V_1, \dots, V_N$, обозначающими автора (авторов) T , дату написания (или коррекции) T , указывающими новую информацию в T и т.д. Но наиболее популярные подходы к математическому исследованию ЕЯ не предусматривают формальных средств для представления текстов как структурированных объектов подобного рода.

На основании проведенного системного исследования поставим задачу построения такой модели, чтобы ее формальные средства позволяли нам следующее:

(Свойство 1): Строить обозначения структурированных значений (СЗ) как фраз, выражающих высказывания, так и повествовательных текстов; такие обозначения обычно называют семантическими представлениями (СП) ЕЯ-выражений.

(Свойство 2): Строить и различать формальными средствами обозначения СЗ повествовательных текстов, СЗ целей (выраженных неопределенными формами глаголов с зависимыми словами, таких как «окончить с отличием МГУ, подготовить и защитить кандидатскую диссертацию по биохимии») и СЗ вопросов.

(Свойство 3): Строить и различать обозначения единиц, соответствующих (а) объектам, ситуациям, процессам в реальном мире и (б) понятиям, квалифицирующим (характеризующим) эти объекты, ситуации, процессы.

(Свойство 4): Строить и различать обозначения: (3.1) объектов и множеств объектов; (3.2) понятий и множеств понятий; (3.3) СП текстов и множеств СП текстов.

(Свойство 5): Различать формальным образом понятия, квалифицирующие объекты, и понятия, квалифицирующие множества объектов тех же видов.

(Свойство 6): Строить составные обозначения понятий, т.е. строить формулы, отражающие поверхностно-семантическую структуру ЕЯ-выражений, подобных выражению «человек, окончивший МГУ имени М.В. Ломоносова и являющийся биологом или химиком».

(Свойство 7): Строить объяснения более общих понятий с помощью менее общих; в частности, строить цепочки вида ($a = Des(b)$), где a обозначает некоторое понятие, которое необходимо объяснить, а $Des(b)$ обозначает описание некоторой конкретизации известного понятия b .

(Свойство 8): Строить обозначения упорядоченных n -местных наборов различных сущностей, где $n > 1$.

(Свойство 9): Строить (9.1) формальные аналоги составных обозначений множеств («эта группа, состоящая из 12 туристов, являющихся химиками или биологами» и т.п.), (9.2) обозначения множеств упорядоченных наборов сущностей, (9.3) обозначения множеств, состоящих из множеств, и т.д.

(Свойство 10): Описывать теоретико-множественные отношения и операции над множествами.

(Свойство 11): Строить обозначения СЗ фраз, содержащих, в частности:

(11.1) слова «произвольный», «некоторый», «все», «каждый» и т.д.;

(11.2) выражения, полученные применением связок «и», «или» к обозначениям (11.2а) предметов, событий; (11.2б) понятий; (11.2в) множеств;

(11.3) выражения, где связка «не» стоит непосредственно перед обозначением предмета, события и т.д.;

(11.4) косвенную речь;

(11.5) причастные обороты и придаточные определительные предложения;

(11.6) слова «понятие», «термин».

(Свойство 12): Строить обозначения СЗ дискурсов со ссылками на упомянутые объекты.

(Свойство 13): Указывать явно в СП дискурсов причинно-следственные и временные отношения между описываемыми ситуациями (событиями).

(Свойство 14): Описывать СЗ дискурсов со ссылками на смысл фраз и более крупных фрагментов рассматриваемых текстов.

(Свойство 15): Выражать суждения о тождественности двух сущностей.

(Свойство 16): Строить формальные аналоги формул логики предикатов первого порядка с кванторами существования и/или всеобщности.

(Свойство 17): Рассматривать нетрадиционные функции (и другие нетрадиционные отношения) с аргументами и/или значениями, являющимися: (17.1) множествами предметов, ситуаций (событий); (17.2) множествами понятий; (17.3) множествами СП текстов.

(Свойство 18): Строить концептуальные представления текстов как информационные объекты, отражающие не только смысл, но и значения внешних характеристик текста: авторов, дату, области применения результатов и т.д.

Эта постановка задачи отражена в публикациях (Фомичев, 1981a, 1981b, 1983, 1988; 2002b, в; Fomichov, 1984; Fomichov, 1992, 1996a, б, 2002b).

3.2. Краткая характеристика предлагаемого решения поставленной задачи

3.2.1. Краткая характеристика новых правил построения формул

В данной главе произвольному концептуальному базису B будут поставлены в соответствие три множества формул $Ls = Ls(B)$, $Ts = Ts(B)$, $Ys = Ys(B)$ (l -формулы, t -формулы, y -формулы). Объединение этих множеств будет обозначено через $Forms(B)$. Множество $Ls(B)$ будет названо *стандартным концептуальным языком (СК-языком) в базисе B* . Концептуальный базис B оказывается возможным определить таким образом, что цепочки языка $Ls = Ls(B)$ будет удобно использовать для описания структурированных значений (другими словами, смысловых структур) ЕЯ-текстов, представления знаний о мире и представления целей интеллектуальных систем. Другими словами, цепочки из языка $Ls = Ls(B)$ окажется удобным использовать для построения семантических представлений (СП) текстов на естественном языке. Формулы из первого класса, т.е. l -формулы, будут называться также K -цепочками.

Каждая формула из множества $Ts(B)$ представима в виде $d \& t$, где $d \in Ls(B)$, t – тип из $Tps(B)$. Формулы из множества $Ys(B)$ являются выражениями вида $a_1 \& \dots \& a_n \& d$, где $a_1, \dots, a_n, d \in Ls(B)$, n имеет разные значения для разных d , и цепочка d строится из $a_1, \dots, a_n$ как из элементарных информационных единиц (некоторые из них могут быть немного преобразованы) однократным применением некоторого правила построения.

В данной работе предлагается оригинальная схема подхода к определению трех классов выводимых формул; эта схема заключается в следующем. Будут сформулированы некоторые высказывания $P[0], \dots, P[10]$; они будут интерпретироваться как правила построения семантических представлений (СП) ЕЯ-текстов из элементов первичного информационного универсума $X(B)$, переменных из $V(B)$ и нескольких специальных символов при условии, что B является концептуальным базисом для рассматриваемой области.

Каждое из этих правил фактически задает некоторую операцию на множестве всевозможных наборов, компоненты которых являются СП простых или составных выражений естественного языка (ЕЯ). Всего 10 операций достаточно для построения формул, отображающих смысл (или структурированные значения) сколь угодно сложных ЕЯ-текстов. Поэтому можно сказать, что система этих правил задает некоторую полную систему квазилингвистических концептуальных операций.

Классы формул Ls , Ts , Ys для произвольного к.б. B определяются совместной индукцией правилами $P[0], P[1], \dots, P[10]$. Для любого к.б. B правило $P[0]$ задает начальный запас формул.

Определение 1. Обозначим через $P[0]$ высказывание «Если $d \in X(B) \cup V(B)$, $t \in Tp(S(B))$, $tp = tp(B)$, $tp(d) = t$, то $d \in L(B)$, и цепочка вида $d \& t$ входит в $T^0(B)$ ».

Пусть B – произвольный к.б., $L(B)$ и $T^0(B)$ — наименьшие множества, задаваемые утверждением $P[0]$, $Lnr_0(B) = L(B)$ (обозначение « Lnr » расшифровывается как « L нумерованное»). Тогда, очевидно, $Lnr_0(B) = X(B) \cup V(B)$, $T^0(B) = \{b \mid b = d \& t, d \in X(B) \cup V(B), t \in Tp(S(B)), t = tp(d)\}$.

Таким образом, в соответствии с правилом $P[0]$ информация о типах элементов первичного информационного универсума $X(B)$ и переменных из $V(B)$ отображается в структуре формул из множества $T^0(B)$.

Пример 1. Пусть S_1 – с.с., построенная в примере из раздела 2.8, $B_1 = (S_1, Ct_1, Ql_1)$ – к.б., определенный в разделе 2.8, $B = B_1$. Тогда легко увидеть, что выполняются следующие соотношения:

чел, П.Сомов, НПО_«Радуга», Друзья $\in Lnr_0(B)$,

Персонал, Поставщики $\in Lnr_0(B)$; чел & $\uparrow_{\text{интс}} * \text{дин.физ.об} \in T^0(B)$,

П.Сомов & $\text{интс} * \text{дин.физ.об} \in T^0(B)$;

НПО_«Радуга» & $\text{орг} * \text{простр.об} * \text{интс} \in T^0(B)$;

Друзья & $\{(\text{интс} * \text{дин.физ.об}, \{\text{интс} * \text{дин.физ.об}\})\} \in T^0(B)$,

Персонал & $\{(\text{орг}, \{\text{интс} * \text{дин.физ.об}\})\} \in T^0(B)$,

Поставщики & $\{(\text{орг}, \{\text{орг}\})\} \in T^0(B)$.

Правило $P[1]$ предназначено для присоединения информационных единиц, соответствующих словам «некоторый», «каждый», «какой-нибудь», «все», «несколько», «большинство» (такие информационные единицы в данной работе называются интенциональными кванторами) к простым или составным обозначениям понятий. Поэтому правило $P[1]$ позволяет строить формальные аналоги выражений: «некоторый человек», «все люди», «большинство людей», «некоторый человек ростом 175 см», «все тридцатилетние люди», «все города Европы». Примерами l -формул (К-цепочек) для $P[1]$, как последнего примененного правила, являются цепочки

$\text{нек чел, все чел} * (\text{Возраст}, 30/\text{год}), \text{все город} * (\text{Регион}, \text{Европа})$.

Правило $P[2]$ предназначено для построения цепочек вида $f(a_1, \dots, a_n)$, где f – обозначение функции, $n \geq 1$, $a_1, \dots, a_n$ – l -формулы, построенные с применением каких-то правил из списка $P[0], P[1], \dots, P[10]$. Например, после применения правила на последнем шаге вывода можно получить цепочки

$\text{Города}(\text{Европа}), \text{Колич-элемент}(\text{Города}(\text{Европа}))$.

Правило $P[3]$ позволяет строить цепочки вида $(a_1 \equiv a_2)$, где a_1, a_2 – l -формулы, полученные при помощи любых правил из $P[0], \dots, P[10]$, и a_1, a_2 обозначают сущности, являющиеся однородными в некотором смысле. Примеры К-цепочек для $P[3]$ как последнего примененного правила:

$(y_1 \equiv \text{нек город} * (\text{Название}, \text{'Саратов'}))$,

$(\text{Директор}(\text{АО_«Салют»}) \equiv \text{П.Сомов})$.

Правило $P[4]$ позволяет строить К-цепочки вида $r(a_1, \dots, a_n)$, где r – n -арное отношение, $n \geq 1$, $a_1, \dots, a_n$ – К-цепочки, полученные при помощи некоторых правил из $P[0], \dots, P[10]$. Примеры К-цепочек для $P[4]$: $\text{Принадлежит}(\text{Намюр}, \text{Города}(\text{Бельгия}))$, $\text{Подмножество}(\text{Города}(\text{Бельгия}), \text{Города}(\text{Европа}))$.

Правило $P[5]$ предназначено для построения К-цепочек вида $d : v$, где d – К-цепочка, не включающая v ; v – переменная, и выполнены

некоторые условия. При помощи правила $P[5]$ можно помечать переменными в семантических представлениях текстов на естественном языке: а) описания различных сущностей, встречающихся в тексте (физических объектов, событий, понятий и др.); б) семантические представления предложений или более крупных фрагментов текста, на которые имеется ссылка в любой части текста. Примерами К-цепочек для правила $P[5]$, примененного на последнем шаге вывода, являются выражения

все чел : Z1, Меньше(Возраст(П.Сомов), 30/год) : P1.

Это правило дает возможность строить семантические представления текстов таким образом, чтобы они отражали референтную (ссылочную) структуру текстов. Демонстрирующие это утверждение примеры приведены ниже.

Правило $P[6]$ позволяет строить К-цепочки вида $\neg d$, где d – К-цепочка, удовлетворяющая ряду условий. Примеры К-цепочек для $P[6]$:

—биолог, —Принадлеж(Бонн, Города(Бельгия)).

Здесь $\neg$ обозначает связку «не».

При помощи правила $P[7]$ можно строить К-цепочки вида $(a_1 \wedge \dots \wedge a_n)$ или $(a_1 \vee \dots \vee a_n)$, где $n > 1$, $a_1, \dots, a_n$ – К-цепочки, обозначающие однородные в некотором смысле сущности. В частности, $a_1, \dots, a_n$ могут быть семантическими представлениями высказываний, описаниями физических объектов, описаниями множеств, состоящих из объектов одной природы, описаниями понятий. Следующие цепочки являются примерами К-цепочек (или l -формул) для $P[7]$:

(Финляндия $\vee$ Норвегия $\vee$ Швеция),

(Принадлеж((Намюр $\wedge$ Гент), Города(Бельгия)) $\wedge$ —Принадлеж(Бонн, Города((Финляндия $\vee$ Норвегия $\vee$ Швеция))))).

Назначение правила $P[8]$ состоит в том, что оно позволяет строить, в частности, К-цепочки вида $c^*(r_1, b_1), \dots, (r_n, b_n)$, где c – информационная единица из первичного универсума X , обозначающая понятие, для $i = 1, \dots, n$ r_i – название функции с одним аргументом или бинарного отношения, b_i обозначает возможное значение r_i для объектов, характеризующихся понятием c . Например, если выбрать соответствующим образом первичные информационные единицы, то после применения на последнем шаге вывода правила $P[8]$ можно получить К-цепочки

город(Страна, Россия), поворот*(Направление, левое).*

Правило $P[9]$ дает возможность строить, в частности, К-цепочки вида $\forall v(des)D$ и $\exists v(des)D$, где $\forall$ – квантор всеобщности, $\exists$ – квантор

существования, *des* обозначает понятие («человек», «город», «целое число» и др.) или составное понятие («целое число, большее 200» и др.). *D* можно интерпретировать как семантическое представление высказывания с переменной *v* о любой сущности, характеризуемой понятием *des*. Примеры К-цепочек для *P*[9] как правила, примененного на заключительном шаге построения формулы:

$\exists x1(\text{нат.число}) \exists x2(\text{нат.число}) \text{Меньше}(x1, x2),$

$\exists y(\text{страна}^*(\text{Регион, Европа}))\text{Больше}(\text{Колич}(\text{Города}(y)), 50).$

Правило *P*[10] позволяет строить, в частности, К-цепочки вида $\langle a_1, \dots, a_n \rangle$, где $n > 1$, $a_1, \dots, a_n$ – К-цепочки. Цепочки, получаемые с использованием правила *P*[10] на последнем шаге вывода, интерпретируются как обозначения *n*-местных упорядоченных наборов. Компонентами такого набора могут быть не только обозначения чисел, объектов, но и семантические представления выражений, множеств, понятий и др. Используя правила *P*[10] и *P*[4] на последних шагах вывода, можно построить цепочку

$\text{Учиться}1(\langle \text{Агент}1, \text{нек чел}^*(\text{Имя, 'Игорь'}) \rangle, \langle \text{Учеб.заведение, ГУ-ВШЭ} \rangle, \langle \text{Начал. момент, 2004} \rangle),$

где *Агент1*, *Учеб.заведение*, *Начал. момент* – обозначения тематических ролей, т.е. обозначения отношений между значением глагола «учиться» и значениями зависящих от него в предложениях групп слов.

3.2.2. Схема определения трех классов формул, порождаемых концептуальными базисами

Рассмотрим более детально предлагаемую оригинальную схему подхода к определению трех классов выводимых формул.

Определение 2. Если *B* – произвольный концептуальный базис, то пусть

(а) $D(B) = X(B) \cup V(B) \cup \{', ', '(', ')', ':', '*', '<', '>'\},$

(б) $Ds(B) = D(B) \cup \{ '&' \},$

(в) $D^+(B)$ и $Ds^+(B)$ – множества всех непустых конечных последовательностей элементов из $D(B)$ и $Ds(B)$ соответственно.

Если $1 \leq i \leq 10$, то для любого к.б. *B* и для $k = 1, \dots, i$ утверждения *P*[0], ..., *P*[*i*] определяют совместной индукцией некоторые множества формул $Lnr_i(B) \subset D^+(B)$, $T^0(B)$, $Tnr_i^1(B)$, ..., $Tnr_i^i(B)$, $Ynr_i^1(B)$, ..., $Ynr_i^i(B) \subset Ds^+(B)$. Множество $Lnr_i(B)$ рассматривается как главный подкласс формул, порождаемых правилами *P*[0], ..., *P*[*i*]. Формулы из этого множества предназначены для описания содержания (смысловых структур) ЕЯ-текстов.

Если $1 \leq k \leq i$, то множество $Tnr_k^i(B)$ состоит из цепочек вида $b \& t$, где $b \in Lnr_i(B)$, $t \in Tp(S(B))$, и b понимается как результат применения правила $P[k]$ к некоторым более простым формулам на последнем шаге вывода. Надо добавить, что при построении b из элементов $X(B)$ и $V(B)$ могут использоваться любые правила $P[0], \dots, P[k], \dots, P[i]$; эти правила можно применять произвольно много раз. Если базис B выбран для описания некоторой области, то b можно понимать как СП текста или фрагмент СП текста, относящегося к данной области. В этом случае t можно рассматривать как описание вида сущностей, характеризующихся этим СП или фрагментом СП. Кроме того, t может квалифицировать b как СП повествовательного текста. Номер i интерпретируется в этих обозначениях как максимальный номер правила из списка $P[0], P[1], \dots, P[10]$, которое мы используем для того, чтобы определить множества формул.

Таким образом, как будет показано ниже, $Lnr_4(B), \dots, Lnr_{10}(B)$ включают формулы *Элем(П.Сомов, Друзья(И.Семенов))*, *Элем(АО_«Салют», Поставщики(НПО_«Радуга»))*, и $Tnr_4^4(B), \dots, Tnr_4^{10}(B)$ включают формулы *Элем(П.Сомов, Друзья(И.Семенов)) & сообщ*, *Элем(АО_«Салют», Поставщики(НПО_«Радуга»)) & сообщ*, где *сообщ* – выделенный сорт $P(B_1)$ «смысл сообщения».

Каждая цепочка $c \in Ynr_k^i(B)$, где $1 \leq k \leq i$, может быть представлена в виде $c = a_1 \& a_2 \& \dots \& a_m \& b$, где $a_1, \dots, a_m, b \in Lnr_i(B)$. Кроме того, найдется такое $t \in Tp(S(B))$, что цепочка $b \& t$ принадлежит $Tnr_i^k(B)$. Цепочки $a_1, \dots, a_m$ строятся с помощью любых правил из списка $P[0], \dots, P[10]$, а цепочка b построена из «блоков» $a_1, \dots, a_m$ (некоторые из них могут быть немного изменены) применением только один раз правила $P[k]$. Возможное количество «блоков» $a_1, \dots, a_m$ зависит от k . Таким образом, множество $Ynr_k^i(B)$ фиксирует результат применения правила $P[k]$ один раз. Ниже мы увидим, что множества $Ynr_4^4(B_1), \dots, Ynr_4^{10}(B_1)$ включают формулы

Элем & П.Сомов & Друзья((И.Семенов) &

& Элем(П.Сомов, Друзья((И.Семенов)),

Элем & АО_«Салют» & Поставщики(НПО_«Радуга») &

& Элем(АО_«Салют», Поставщики(НПО_«Радуга»)).

Пусть для $i = 1, \dots, 10$ $T_i(B) = T^0(B) \cup Tnr_i^1(B) \cup \dots \cup Tnr_i^i(B)$; $Y_i(B) = Ynr_i^1(B) \cup \dots \cup Ynr_i^i(B)$; $Form_i(B) = Lnr_i(B) \cup T_i(B) \cup Y_i(B)$.

Будем интерпретировать $Form_i(B)$ как множество формул, порожаемых к.б. B . Это множество представляет собой объединение трех классов формул, главным из которых является $Lnr_i(B)$. Формулы из этих трех классов будем называть соответственно l -формулами, t -формула-

ми и Y -формулами. Класс t -формул необходим для того, чтобы связать тип из $Tr(S(B))$ с каждым $b \in Lnr_i(B)$, где $i = 1, \dots, 10$. Для $i = 0, \dots, 9$, $Lnr_i(B) \subseteq Lnr_{i+1}(B)$. Множество $Lnr_{10}(B)$ называется стандартным концептуальным языком (стандартным К-языком, СК-языком) в стационарном базисе B и обозначается через $Ls(B)$. Поэтому l -формулы будут часто называться К-цепочками.

Множество $T_{10}(B)$ обозначается через $Ts(B)$. Для любого к.б. B и для любой формулы A из множества $Ts(B)$ существуют такой тип $t \in Tr(S(B))$ и такая формула $C \in Ls(B)$, что $A = C \ \& \ t$. Для построения СП текстов будут использоваться только формулы из $Ls(B)$ и $Ts(B)$, т.е. l -формулы и t -формулы. Y -формулы рассматриваются как вспомогательные и нужны для того, чтобы сформулировать некоторые полезные свойства множеств $Ls(B)$ и $Ts(B)$.

3.3. Использование интенциональных кванторов в формулах

В разделе 2.8 было введено понятие *интенционального квантора*. Этот термин используется для обозначения информационных единиц (другими словами, семантических единиц), соответствующих, в частности, словам и выражениям «каждый», «некоторый», «произвольный», «какой-нибудь», «определенный», «все», «несколько», «большинство», «почти все». Совокупность интенциональных кванторов делится на два подкласса, обозначаемые через Int_1 и Int_2 . Это осуществляется следующим образом. Компонентом каждого концептуального базиса B вида (2.8.2) является система кванторов и логических связок (с.к.л.с.) Ql вида (2.8.1). Компонентами с.к.л.с. Ql вида являются, в частности, два выделенных сорта int_1 и int_2 . Это дает возможность для $m = 1, 2$ определить Int_m как $\{x \in X \mid tp(x) = int_m\}$, где первичный информационный универсум X является одним из компонентов концептуально-объектной системы $Ct(B)$ вида (2.7.1).

Элементы множества Int_1 соответствуют значениям выражений «каждый», «какой-то», «некоторый», «произвольный» и т.д. в случаях, когда эти выражения являются частями групп слов, и эти группы связаны с единственным числом. Элементы множества Int_2 интерпретируются как семантические единицы, соответствующие выражениям «все», «несколько», «почти все», «многие» и т.д. ; минимальное требование к Int_2 заключается в том чтобы Int_2 содержало семантическую единицу, соответствующую слову «все».

Правило $P[1]$ позволяет нам присоединять интенциональные кванторы к простым или составным обозначениям понятий. В результате

применения этого правила, во-первых, строятся: (а) l -формулы вида $Int.qr Conc.expr$, где $Int.qr$ – интенсиональный квантор из $Int(B)$, а $Conc.expr$ – простое или составное обозначение понятия. Во-вторых, строятся t -формулы вида $Int.qr Conc.expr \& t$, где t – тип из множества $Tr(S(B))$.

Например, можно выбрать к.б. B так, что в этом базисе с помощью правил $P[0]$ и $P[1]$ можно будет построить l -формулы

нек город, нек город(Назв, «Чита»),*
каждый город, каждый человек(Квалиф., студент),*
все город, все город(Страна, Россия)*

и t -формулы

нек. город & простр. об, все город & {простр. об},
каждый человек(Квалиф, студент) & интс*дин.физ.об.*

Определение 1. Если B — произвольный к.б., то для $m = 1, 2$
 $Int_m(B) = \{q \in X(B) \mid tp(q) = int_m(B)\}$, $Int(B) = Int_1(B) \cup Int_2(B)$,
 $Tconc(B) = \{t \in Tr(S(B)) \mid t \text{ начинается с символа } \uparrow\} \cup Spectr$,
 где $Spectr = \{[\uparrow сущн], [\uparrow пон], [\uparrow об]\}$.

Напомним, что в разделе 2.6 выражения $[\uparrow сущн]$, $[\uparrow пон]$, $[\uparrow об]$ интерпретируются как такие символы (т.е. неделимые единицы), которые являются информационными единицами, соответствующими словам «сущность», «понятие» (или «концепт») и «объект». В данной работе термин «сущность» является наиболее общим. Объектами называются все те сущности, которые не являются понятиями.

Используя правила $P[0]$ и $P[1]$, мы можем строить l -формулы вида $q d$, где $q \in Int(B)$, $d \in X(B)$, $tp(d) \in Tconc(B)$. Так, рассматривая к.б. B_1 , определенный в разделе 2.8, мы можем построить l -формулы

нек чел, нек тур.гр, нек понятие, все чел, все тур.гр, все понятие.

Можно также строить более сложные цепочки вида $q descr$, где q – интенсиональный квантор, $descr$ – составное обозначение понятия. Для построения цепочки $descr$ используются правила $P[0]$, $P[1]$, правило $P[8]$ (см. раздел 3.6) и, возможно, некоторые другие правила. Например, для к.б. B_1 будет возможно построить l -формулы *нек тур.гр*(Колич, 12)*, *все тур.гр*(Колич, 12)*. Эти формулы понимаются как семантические представления (СП) выражений «некоторая туристическая группа из 12 человек» и «все тургруппы из 12 человек».

Переход от l -формулы s , обозначающей понятие, к l -формуле $q s$, где q – интенсиональный квантор, описывается с помощью специальной функции h .

Определение 2. Пусть B – произвольный к.б., $S = S(B)$, $Tr(S)$ – множество типов, порождаемых торговой системой S . Тогда отображение $h: \{1, 2\} \times Tr(S) \rightarrow Tr(S)$ задается следующим образом:
 (а) если $u \in Tr(S)$ и цепочка $\uparrow u$ входит в $Tr(S)$, то $h(1, \uparrow u) = u$, $h(2, u) = \{u\}$;
 (б) $h(1, [\uparrow_{сущн}]) = [сущн]$, $h(1, [-пон]) = [пон]$, $h(1, [\uparrow_{об}]) = [об]$,
 $h(2, [\uparrow_{сущн}]) = \{[сущн]\}$, $h(2, [\uparrow_{пон}]) = \{[пон]\}$, $h(2, [\uparrow_{об}]) = \{[об]\}$.

С точки зрения построения СП текстов, отображение h описывает преобразование типов как следующие переходы:

(а) от понятия «человек», «туристическая группа» к СП выражений «некоторый человек», «каждый человек», «любой человек», «некоторая туристическая группа», «любая туристическая группа» и т.д. (в случае, когда первый аргумент h равен 1) и к СП выражений «все люди», «все туристические группы» и т.д. (в случае, когда первый аргумент h равен 2);

(б) от выражений «сущность», «понятие», «объект» к СП выражений «некоторая сущность», «произвольная сущность», «некоторое понятие», «произвольное понятие», «некоторый объект», «произвольный объект» и т.д. (если первый аргумент h равен 1) и к СП выражений «все сущности», «все понятия», «все объекты» и т.д. (если первый аргумент h равен 2).

Определение 3. Через $P[1]$ обозначим высказывание

«Пусть $c \in L(B) \setminus V(B)$, $u \in Tconc(B)$, $k \in \{0, 8\}$, и цепочка $c \& u$ входит в $T^k(B)$. Пусть $m \in \{1, 2\}$, $q \in Int_m$, $t = h(m, u)$, и b – цепочка вида qc . Тогда $b \in L(B)$, цепочка вида $b \& t$ входит в $T^1(B)$, и цепочка вида $q \& a \& b$ входит в $Y^1(B)$ ».

Пример 1. Пусть B – к.б. B_1 , построенный в разделе 2.8; $L(B)$, $T^0(B)$, $T^1(B)$, $Y^1(B)$ – наименьшие множества, совместно определяемые высказываниями $P[0]$ и $P[1]$. Тогда легко убедиться в справедливости следующих соотношений:

$чел \in L(B) \setminus V(B)$, $чел \& \uparrow_{интс*дин.физ.об} \in T^0(B)$, $нек \in Int_1(B)$,
 $h(1, \uparrow_{интс*дин.физ.об}) = интс*дин.физ.об \Rightarrow$
 $нек чел \in L(B)$, $нек чел \& интс*дин.физ.об \in T^1(B)$, $нек \& чел \&$
 $\& нек чел \in Y^1(B)$;

$все \in Int_2(B)$, $h(2, \uparrow_{интс*дин.физ.об}) = \{интс*дин.физ.об\} \Rightarrow$
 $все чел \in L(B)$, $все чел \& \{интс*дин.физ.об\} \in T^1(B)$, $все \& чел \&$
 $\& все чел \in Y^1(B)$;

$тур.гр \in L(B) \setminus V(B)$, $тур.гр \& \uparrow \{интс*дин.физ.об\} \in T^0(B)$,

$$\begin{aligned}
& h(1, \uparrow\{\text{интс} * \text{дин. физ. об}\}) = \{\text{интс} * \text{дин. физ. об}\}, \\
& h(2, \uparrow\{\text{интс} * \text{дин. физ. об}\}) = \{\{\text{интс} * \text{дин. физ. об}\}\} \Rightarrow \\
& \text{нек тур.гр.}, \text{все тур.гр.} \in L(B), \text{нек тур.гр.} \& \{\text{интс}\}, \text{все тур.гр.} \& \\
& \& \{\{\text{интс}\}\} \in T^1(B), \\
& \text{нек} \& \text{тур.гр.} \& \text{нек тур.гр.} \in Y^1(B); \text{все} \& \text{тур.гр.} \& \text{все тур.гр.} \in Y^1(B); \\
& \text{понятие} \in L(B) \setminus V(B), \text{понятие} \& [\uparrow \text{пон}] \in T^0(B), \\
& h(1, [\uparrow \text{пон}]) = [\text{пон}], h(2, [\uparrow \text{пон}]) = \{[\text{пон}]\} \Rightarrow \\
& \text{нек понятие}, \text{все понятие} \in L(B), \\
& \text{нек понятие} \& [\text{пон}], \text{все понятие} \& \{[\text{пон}]\} \in T^1(B), \\
& \text{нек} \& \text{понятие} \& \text{нек понятие} \in Y^1(B), \\
& \text{все} \& \text{понятие} \& \text{все понятие} \in Y^1(B).
\end{aligned}$$

Комментарий к правилу P[1]. Фрагмент правила P[1] «Пусть $c \in L(B) \setminus V(B)$, $u \in T_{\text{conc}}(B)$, $k \in \{0, 8\}$, и цепочка $c \& u$ входит в $T^k(B)$ » означает, что u – тип понятия (так как либо начинается с символа ‘ $\uparrow$ ’, либо является одним из символов $[\uparrow \text{сущн}]$, $[\uparrow \text{пон}]$, $[\uparrow \text{об}]$), при $k = 0$ c – простое обозначение понятия ($c \in X(B)$); при $k = 8$ c – составное обозначение понятия. Такие составные обозначения понятий будут строиться с помощью правила P[8]; примерами таких выражений являются выражения *человек*(Область.деят, биология)*, *город*(Страна, Россия)*.

Правило P[1] будет очень часто использоваться при построении СП текстов, так как оно нужно для построения семантических образов выражений с существительными. Например, пусть $T1 =$ «Откуда поступил двухтонный алюминиевый контейнер?». Тогда в результате выполнения первого шага построения СП $T1$ можно получить выражение

$E1 = \text{нек контейнер}1 * (\text{Вес}, 2/\text{Тонна}) (\text{Материал}, \text{алюминий})$,
а после выполнения второго шага – заключительное выражение

$$\begin{aligned}
E2 = & \text{Вопрос}(x1, \text{Ситуация}(s1, \text{поступление}2 * \\
& * (\text{Место}1, x1) (\text{Объект}1, \text{нек контейнер}1 * \\
& * (\text{Вес}, 2/\text{Тонна}) (\text{Материал}, \text{алюминий})))
\end{aligned}$$

Более подробно такого рода шаги при построении СП текстов рассматриваются ниже в данной главе и в главе 4.

Часто используемые обозначения. Рассмотрим обозначения, которые в дальнейшем будут часто использоваться в примерах. Для $k = 1, \dots, 10$ правило P[k] утверждает, что некоторая формула b входит в $L(B)$, некоторая формула $b \& t$ принадлежит $T^k(B)$, где $t \in Tp(S(B))$, и некоторая формула z принадлежит $Y^k(B)$. Если $1 \leq i \leq 10$, B – произ-

вольный к.б., то правила $P[0], P[1], \dots, P[i]$ определяют совместной индукцией множества формул $L(B), T^0(B), T^1(B), \dots, T^i(B), Y^1(B), \dots, Y^i(B)$.

Обозначим эти множества через $Lnr_i(B), T^0(B), Tnr_i^1(B), \dots, Tnr_i^i(B), Ynr_i^1(B), \dots, Ynr_i^i(B)$; семейство, состоящее из всех этих множеств, обозначим через $Globset_i(B)$.

Пусть $n \geq 1, Z_1, \dots, Z_n \in Globset_i(B), w_1 \in Z_1, \dots, w_n \in Z_n$. Тогда, если эти соотношения для формул $w_1, \dots, w_n$ являются следствием применения некоторых правил $P[l_1], \dots, P[l_m]$, где $m \geq 1$, то обозначим этот факт выражением вида $B(l_1, \dots, l_m) \Rightarrow w_1 \in Z_1, \dots, w_n \in Z_n$. Последовательность $l_1, \dots, l_m$ может содержать повторяющиеся номера. В выражениях такого рода будет часто пропускаться символ B в обозначениях множеств $Z_1, \dots, Z_n$; кроме того, будут использоваться выражения $w_1, w_2 \in Z_1, w_3, w_4, w_5 \in Z_2$ и т.д. Используя эти обозначения, некоторые соотношения, полученные в Примере 1, можно представить следующим образом:

$$B_1(0,1) \Rightarrow \text{все чел, все тур.гр} \in Lnr_1,$$

$$\begin{aligned} \text{все чел} \& \{\text{интс}^* \text{дин. физ. об}\}, \text{все тур.гр} \& \{\{\text{интс}^* \text{дин. физ. об}\}\} \in Tnr_1^1, \\ \text{все} \& \text{чел} \& \text{все чел} \in Ynr_1^1, \text{все} \& \text{тур.гр} \& \text{все тур.гр} \in Ynr_1^1. \end{aligned}$$

Выражение $B_1(0, 1) \Rightarrow \text{все чел} \in Lnr_1, \text{все чел} \& \{\text{интс}^* \text{дин. физ. об}\} \in Tnr_1^1$ равносильно выражению

$$B_1(0,1) \Rightarrow \text{все чел} \in Lnr_1(B_1), \text{все чел} \& \{\text{интс}^* \text{дин. физ. об}\} \in Tnr_1^1(B).$$

3.4. Использование реляционных символов и разметка формул

3.4.1. Правила для применения реляционных символов

Правило $P[2]$ позволяет нам, в частности, строить К-цепочки вида $f(a_1, \dots, a_n)$, где f обозначает функцию с n аргументами $a_1, \dots, a_n$. Правило $P[3]$ предназначено для построения К-цепочек вида $(a_1 \equiv a_2)$, где a_1 и a_2 обозначают сущности, характеризующиеся типами, сравнимыми друг с другом для отношения конкретизации $\vdash$. Используя последовательно $P[2]$ и $P[3]$, мы сможем строить К-цепочки вида $(f(a_1, \dots, a_n) \equiv b)$, где b – значение f для аргументов $a_1, \dots, a_n$.

Отметим, что для с.с. S множество главных типов $Mtp(S) = Tp(S) \setminus \{\uparrow \text{сущн}, \uparrow \text{об}, \uparrow \text{пон}\}$ (см. раздел 2.6).

Определение 1. Пусть B – произвольный к.б., $S = S(B)$. Тогда:

- (а) $R_1(B) = \{d \in X(B) \mid \text{найдется такой тип } t \in Mtp(S), \text{ что } t \text{ начинается с символа ' и } tp(d) - \text{цепочка вида } \{t\}\}$;
- (б) для произвольного $n > 1, R_n(B) = \{d \in X(B) \mid \text{найдутся такие } t_1, \dots, t_n \in Mtp(S), \text{ что } tp(d) - \text{цепочка вида } \{(t_1, \dots, t_n)\}\}$;
- (в) для произвольного $n > 1, F_n(B) = F(B) \cap R_{n+1}(B)$.

Если $n \geq 1$, то элементы множества $R_n(B)$ будем называть n -арными реляционными символами, а элементы $F_n(B)$ будем дополнительно называть n -арными функциональными символами.

Легко показать, что для произвольного к.б. B и произвольных $k, m > 1$ из $k \neq m$ следует $R_k(B) \cap R_m(B) = \emptyset$.

Определение 2. Обозначим через $P[2]$ высказывание «Пусть $n \geq 1$, $f \in F_n(B)$, $tp = tp(B)$, $u_1, \dots, u_m, t \in Mtp(S(B))$, $tp(f) = \{(u_1, \dots, u_m, t)\}$; для $j = 1, \dots, n$, $0 \leq k \leq i$, $z_j \in Mtp(S(B))$, $a_j \in L(B)$, цепочка $a_j \& z_j$ принадлежит $T_j^k(B)$; если a_j не входит в $V(B)$, то $u_j \vdash z_j$ (т.е. z_j является конкретизацией типа u_j); если $a_j \in V(B)$, то u_j и z_j сравнимы для отношения конкретизации $\vdash$. Пусть b – цепочка вида $f(a_1, \dots, a_n)$. Тогда $b \in L(B)$, $b \& t \in T^2(B)$, $f \& a_1 \& \dots \& a_n \& b \in Y^2(B)$ ».

Перед тем, как сформулировать следующее утверждение, следует напомнить, что символ '=' является элементом первичного информационного универсума $X(B)$ для произвольного к.б. B .

Определение 3. Обозначим через $P[3]$ высказывание «Пусть $a_1, a_2 \in L(B)$, $u_1, u_2 \in Mtp(S(B))$, типы u_1 и u_2 сравнимы для отношения конкретизации $\vdash$. Пусть для $m = 1, 2$, $0 \leq k[m] \leq i$, $a_m \& u_m \in T^{k[m]}(B)$; P – сорт «смысл сообщения» для к.б. B , b – цепочка ($a_1 \equiv a_2$). Тогда $b \in L(B)$, $b \& P \in T^3(B)$, и цепочка $a_1 \& \equiv \& a_2 \& b$ входит в множество $Y^3(B)$ ».

В правилах $P[2]$ и $P[3]$ символ i обозначает неизвестное число, такое что $2 \leq i \leq 10$. Дело в том, что правила $P[0]$ – $P[3]$ и последующие правила будут использоваться вместе с объединяющим их определением, которое начинается с такой фразы: «Пусть B – произвольный к.б., $1 \leq i \leq 10$ ». Число i будет интерпретироваться как максимальный номер правила, которое будет использоваться для построения формулы. Например, если $i = 3$, то мы можем использовать правила с номерами 0, 1, 2, 3, но не можем использовать правила с номерами 4–10. Параметр i в правилах вывода позволяет нам после введения очередного правила с номером $i + 1$ определить формальный язык $Lnr_{i+1}(B)$ и исследовать выразительные возможности этого языка.

Пример 1. Пусть B_1 – к.б., построенный в разделе 2.8; $i = 3$;

$b_1 = \text{Поставщики(НПО_«Радуга»),}$

$b_2 = \text{Колич (Поставщики(НПО_«Радуга»),)}$

$b_3 = (\text{Колич (Поставщики(НПО_«Радуга»),}) \equiv 12),$

$b_4 = \text{Колич(все понятие),}$

$b_5 = \text{Колич(все химик)}$, $b_6 = (\text{все химик} \equiv x1)$, $b_7 = \text{Колич}(x1)$.

Тогда легко убедиться в справедливости следующих соотношений (принимая во внимание обозначения, введенные в конце предыдущего раздела):

$$B_1 \Rightarrow \text{Поставщики} \& \{(\text{орг}, \{\text{орг}\})\} \in T^0,$$

$$\text{НПО_«Радуга»} \& \text{орг} * \text{простр.об} * \text{интс} \in T^0,$$

$$\text{Колич} \& \{(\{\text{сущн}\}, \text{нат})\} \in T^0;$$

$$B_1(1, 2) \Rightarrow b_1 \& \{\text{орг}\} \in \text{Tnr}_3^2;$$

$$B_1(0, 2, 2) \Rightarrow b_2 \in \text{Lnr}_3, b_2 \& \text{нат} \in \text{Tnr}_3^2, \text{Колич} \& b_1 \& b_2 \in \text{Ynr}_3^2;$$

$$B_1(0, 2, 2, 0, 3) \Rightarrow b_3 \in \text{Lnr}_3, b_3 \& \text{сообщ} \in \text{Tnr}_3^3;$$

$$B_1(0, 2, 2) \Rightarrow b_4, b_5 \in \text{Lnr}_3, b_4 \& \text{нат}, b_5 \& \text{нат} \in \text{Tnr}_3^2;$$

$$B_1(0, 1, 3) \Rightarrow b_6 \in \text{Lnr}_3, b_6 \& \text{сообщ} \in \text{Tnr}_3^3;$$

$$x1 \in V(B_1), tp(x1) = \{\{\text{сущн}\}\}, B_1(0, 2) \Rightarrow b_7 \in \text{Lnr}_3, b_7 \& \text{нат} \in \text{Tnr}_3^2.$$

Определение 4. Обозначим через $P[4]$ высказывание «Пусть $n \geq 1$, $r \in R_n(B) \setminus F(B)$, $u_1, \dots, u_n \in \text{Mtp}(S(B))$, $tp = tp(B)$, $tp(r)$ – цепочка $\{(u_1, \dots, u_n)\}$ при $n > 1$ или $\{u_1\}$ при $n = 1$; для $j = 1, \dots, n$, $0 \leq k[j] \leq i$, $z_j \in \text{Mtp}(S(B))$, $a_j \in L(B)$, цепочка $a_j \& z_j$ принадлежит $T_j^{k[j]}(B)$; если $a_j \notin V(B)$, то $u_j \vdash z_j$; если $a_j \in V(B)$, то u_j и z_j сравнимы для отношения конкретизации $\vdash$. Пусть b – цепочка вида $r(a_1, \dots, a_n)$, $P = P(B)$ – сорт «смысл сообщения» для B . Тогда $b \in L(B)$, $b \& P \in T^4(B)$, и $r \& a_1 \& \dots \& a_n \& b \in Y^4(B)$ ».

Пример 2. Пусть $B1$ – к.б., рассмотренный в разделе 2.8; $i = 4$;

$$b_8 = \text{Меньше}(10000, \text{Колич(все химик)}),$$

$$b_9 = \text{Меньше}(5000, \text{Колич(все понятие)}),$$

$$b_{10} = \text{Элем(биолог, все понятие)},$$

$$b_{11} = \text{Элем(АО_«Старт», Поставщики(НПО_«Радуга»))},$$

$$b_{12} = (\text{П.Сомов} \equiv \text{Директор(АО_«Старт»)}),$$

$$b_{13} = \text{Знает(П.Сомов, (Колич(Поставщики(НПО_«Радуга»))} \equiv 12)),$$

$$b_{14} = \text{Знает(И.Семенов, (П.Сомов} \equiv \text{Директор(АО_«Старт»))},$$

$$b_{15} = \text{Меньше}(10000, \text{Колич}(x1)).$$

Принимая во внимание определение к.о.с. $St(B_1)$ (см. пример в разделе 2.7) и применяя правила $P[0], \dots, P[4]$, мы получаем следующие соотношения:

$$B_1(0, 1, 2, 3, 4) \Rightarrow b_8, \dots, b_{15} \in \text{Lnr}_4, \text{ для } m = 8, \dots, 15 \text{ } b_m \& \text{сообщ} \in \text{Tnr}_4^4,$$

$$\text{Меньше} \& 1000 \& \text{Колич(все химик)} \& \text{Меньше}(10000, \text{Колич(все химик)}) \in \text{Ynr}_4^4.$$

3.4.2. Правило, позволяющее пометать формулы

Основное назначение правила $P[5]$ заключается в том, чтобы с помощью переменных из множества $V(B)$, где B – рассматриваемый к.б., пометать в семантических представлениях (СП) ЕЯ-текстов: (а) описания различных сущностей, упомянутых в тексте (физических объектов, событий, понятий и т.д.), (б) фрагменты, являющиеся семантическими представлениями предложений и более крупных частей текстов, на которые имеются ссылки в любой части текста. С помощью этого правила (в сочетании с другими правилами) окажется возможным отражать референтную (ссылочную) структуру дискурсов, в которых имеются, в частности, выражения (а) этот прибор, на этом заводе, в этом городе, ему, о нем, (б) данный метод, это распоряжение, эту команду, этот вопрос, об этом, про это.

Определение. Обозначим через $P[5]$ высказывание «Пусть $a \in L(B) \setminus V(B)$, $0 \leq k \leq i$, $k \neq 5$, $t \in Mtp(S(B))$, $a \ \& \ t \in T^k(B)$; $v \in V(B)$, $z \in Mtp(S(B))$, $v \ \& \ z \in T^0(B)$, $z \vdash t$, v не является подцепочкой цепочки a . Пусть b – цепочка вида $a : v$. Тогда $b \in L(B)$, $b \ \& \ t \in T^5(B)$, $a \ \& \ v \ \& \ b \in Y^5(B)$ ».

Условие « $k \neq 5$ » вводится для того, чтобы правило $P[5]$ нельзя было применять произвольное число раз подряд; в противном случае мы могли бы получать формулы с «избыточной разметкой», т. е. выражения вида $a : v_1 : v_2 \dots v_n$, где $n > 1$, и $v_1, v_2, \dots, v_n \in V(B)$.

Пример 3. Рассмотрим, как и прежде, к.б. B_1 , построенный в разделе 2.8. Пусть $i = 5$, $a_1 = b_3 = (\text{Колич}(\text{Поставщики}(\text{НПО_«Радуга»})) \equiv 12)$, $k_1 = 3$, $t_1 = \text{сообщ} = P(B_1)$. Тогда, очевидно, $a_1 \ \& \ t_1 \in \text{Trn}_5^3(B_1)$.

Предположим, что $v_1 = P1$, $z_1 = \text{сообщ} = P(B_1)$. Тогда, в соответствии с определением базиса B_1 , $v_1 \ \& \ z_1 \in T^0(B_1)$; кроме того, $z_1 \vdash t_1$ (так как $z_1 = t_1$), и v_1 не является подстрокой a_1 .

Пусть $b_{16} = (\text{Колич}(\text{Поставщики}(\text{НПО_«Радуга»})) \equiv 12) : P1$. Тогда, в соответствии с правилом $P[5]$, $b_{16} \in \text{Lnr}_5(B_1)$, $b_{16} \ \& \ \text{сообщ} \in \text{Trn}_5^5(B_1)$, $a_1 \ \& \ v_1 \ \& \ b_{16} \in \text{Ynr}_5^5(B_1)$.

Пусть $b_{17} = \text{Поставщики}(\text{НПО_«Радуга»}) : x3$. Тогда легко видеть, что

$$B_1(0, 2, 5) \Rightarrow b_{17} \in \text{Lnr}_5, b_{17} \ \& \ \{op_2\} \in \text{Trn}_5^5,$$

$$\text{Поставщики}(\text{НПО_«Радуга»}) \ \& \ x_3 \ \& \ b_{17} \in \text{Ynr}_5^5.$$

В выражении b_{16} переменная $P1$ пометает СП фразы $\Pi_1 = \text{«У НПО «Радуга» имеется 12 поставщиков»}$. Поэтому, если это выражение является частью длинной цепочки, то справа от вхождения b_{16} в такую

цепочку можно использовать переменную $P1$ для повторного представления смысла указанной фразы $P1$ вместо значительно более длинного СП фразы $P1$. В цепочке b_{17} переменная x_3 является меткой множества, состоящего из всех поставщиков научно-производственного объединения «Радуга».

3.5. Использование логических связок «не», «и», «или»

Правила $P[6]$ и $P[7]$ в сочетании с другими правилами позволяют по сравнению с языком логики предикатов более полно моделировать (на уровне семантических представлений текстов) способы использования связок «не», «и», «или» в предложениях на русском, английском и многих других языках. В частности, конструктивно учитывается существование фраз вида «Этот препарат выпускается не в Польше», «Профессор Сухов работает не в МГУ», «Этот патент внедрен в Австрии, Венгрии, Нидерландах и Великобритании». С этой целью, во-первых, допускается присоединение связки $\neg$ («не») не только к семантическим представлениям (СП) фраз, обозначающих высказывания, но и к обозначениям предметов, событий, понятий. Во-вторых, разрешается соединять связками $\wedge$ («конъюнкция», т.е. логическое «и») и $\vee$ («дизъюнкция», т.е. логическое «или») не только СП высказываний, но и обозначения предметов, ситуаций, понятий, множеств предметов и т.д.

Правило $P[6]$ предназначено для построения l -формул вида $\neg a$, где $a \in Lnr_i(B)$.

Определение 1. Обозначим через $P[6]$ высказывание «Пусть $a \in L(B)$, $t \in Mtp(S(B))$, $0 \leq k \leq i$, $k \notin \{2, 5, 10\}$, $a \& t \in T^6(B)$, b – цепочка вида $\neg a$. Тогда $b \in L(B)$, $b \& t \in T^6(B)$, цепочка вида $\neg \& a \& b$ входит в множество $Y^6(B)$ ».

Замечание 1. Условие $k \notin \{2, 5, 10\}$ означает следующее: если l -формула a выведена каким-либо образом с помощью правил $P[0]$, $P[1]$, ..., $P[i]$, то правило $P[2]$ или $P[5]$ или $P[10]$ не может быть применено на последнем шаге вывода. Таким образом, условие $k \neq 2$ означает, что не разрешается строить l -формулы вида $\neg f(d_1, \dots, d_m)$, где $f \in F(B)$. Условие $k \neq 5$ вводится для того, чтобы по любому выражению вида $\neg a : v$, где $v \in V(B)$, можно было однозначно найти то правило, которое применялось последним; этим правилом будет $P[5]$. Условие $k \neq 10$ запрещает строить выражения вида $\neg \langle C_1, \dots, C_m \rangle$, т.е. запрещает присоединять связку «не» к обозначениям упорядоченных наборов.

Пример 1. Пусть B_1 – к.б., построенный в разделе 2.8, $i = 6$. Тогда легко видеть, что выполняются следующие соотношения:

$$B_1(0, 6) \Rightarrow \neg \text{биолог} \in Lnr_6, \neg \text{биолог} \ \& \ \uparrow_{\text{интс}^* \text{дин. физ. об}} \in Tnr_6^6;$$

$$B_1(0, 6, 4, 4) \Rightarrow \text{Знает}(\text{П. Сомов, Сейчас},$$

$$\text{Явл1}(\text{И. Семенов}, \neg \text{биолог})) \in Tnr_6^4;$$

$$B_1(0, 4, 4, 6) \Rightarrow \neg \text{Знает}(\text{П. Сомов, Сейчас}, \text{Явл1}(\text{И. Семенов},$$

$$\text{химик})) \in Tnr_6^6.$$

Правило $P[7]$ позволяет строить l -формулы вида $(a_1 \wedge a_2 \wedge \dots \wedge a_n)$ и вида $(a_1 \vee a_2 \vee \dots \vee a_n)$. Например, формулы $(\text{химик} \vee \text{биолог})$, $(\text{математик} \wedge \text{художник})$, $(\text{Имя}(x1, \text{'Сергей'}) \wedge \text{Фам}(x1, \text{'Жаворонков'}) \wedge \text{Квалиф}(x1, \text{химик}))$.

Определение 2. Обозначим через $P[7]$ высказывание «Пусть $n > 1$, $t \in Mtp(S(B))$, для $m = 1, \dots, n$, $0 \leq k[m] \leq i$, $a_m \ \& \ t \in T^{k[m]}(B)$, $s \in \{ \wedge, \vee \}$, b – цепочка вида $(a_1 \ s \ a_2 \ s \ \dots \ s \ a_n)$. Тогда $b \in L(B)$, $b \ \& \ t \in T^7(B)$, $s \ \& \ a_1 \ \& \ \dots \ \& \ a_n \ \& \ b \in Y^7(B)$ ».

Замечание 2. Данное правило требует, чтобы все выражения, соединенные за один шаг логической связкой, имели один и тот же тип. Так как t не обязательно является сортом «смысл сообщения», то по правилу $P[7]$ можно соединять логическими связками не только семантические представления высказываний, но и обозначения различных объектов, простые и составные обозначения понятий, простые и составные обозначения целей интеллектуальных систем.

Пример 2. Пусть B_1 – к.б., построенный в в разделе 2.8, $i = 7$,

$$b_1 = (A.Зубов \wedge \text{И. Семенов});$$

$$b_2 = (\text{химик} \vee \text{биолог});$$

$$b_3 = ((\text{Колич}(\text{Друзья}(A.Зубов))) \equiv 3) : P1 \wedge \text{Знает}(\text{П. Сомов, Сейчас},$$

$$P1) \wedge \neg \text{Знает}(\text{П. Сомов, Сейчас}, \text{Явл}(A.Зубов, (\text{химик} \vee \text{биолог}))));$$

$$b_4 = \text{Элем}((AO_ \text{«Салют»} \wedge AO_ \text{«Старт»}),$$

$$\text{Поставщики}(\text{НПО_«Радуга»})).$$

Тогда легко показать, что

$$B_1(0, 7) \Rightarrow b_1, b_2 \in Lnr_7, b_1 \ \& \ \uparrow_{\text{интс}^* \text{дин. физ. об}} \in Tnr_7^7,$$

$$b_2 \ \& \ \uparrow_{\text{интс}^* \text{дин. физ. об}} \in Tnr_7^7;$$

$$B_1(0, 2, 2, 3, 5, 4, 7, 4, 4, 6, 7) \Rightarrow b_3 \in Lnr_7,$$

$$b_3 \ \& \ \text{сообщ} \in Tnr_7^7;$$

$$B_1(0, 7, 0, 2, 4) \Rightarrow b_4 \in Lnr_7, b_4 \ \& \ \text{сообщ} \in Tnr_7^4.$$

3.6. Построение составных обозначений понятий и объектов

3.6.1. Правило для построения составных обозначений понятий

Рассмотрим правило $P[8]$, предназначенное для построения составных обозначений понятий. С помощью этого правила строятся l -формулы вида $a^*(r_1, d_1) \dots (r_n, d_n)$ и t -формулы вида $a^*(r_1, d_1) \dots (r_n, d_n) \ \& \ t$, где a – элемент первичного информационного универсума $X(B)$ и интерпретируется как простое обозначение понятия, $n \geq 1$, для $i = 1, \dots, n$, $r_i \in R_2(B)$, d_i – обозначение некоторой сущности. Например, выбирая подходящий концептуальный базис, можно построить l -формулы

*город**(Страна, Россия), *учебник**(Область, биология),

*понятие**(Имя.пон, «молекула»),

*туристич.группа**(Колич, 12)(Качеств-состав, (химик $\wedge$ биолог))

и t -формулу *город**(Страна, Россия) $\& \uparrow$ простр.об.

Используя правило $P[8]$ вместе с правилом $P[1]$ и другими правилами, можно будет строить составные обозначения объектов и множеств объектов в виде $q \ des$, где q – интенциональный квантор (см. раздел 2.8), des – составное обозначение понятия, построенное с помощью правила $P[8]$ на последнем шаге вывода. В частности, формулы

*некотор учебник**(Область, биология),

*все город**(Страна, Россия),

*некотор чел**(Возраст, 18/год),

*некотор понятие**(Имя.пон, «молекула»),

*некотор туристич.группа**(Колич, 12)

(Качеств-состав, (химик $\wedge$ биолог)).

Определение 1. Для произвольного концептуального базиса B

$$Tconc(B) = \{t \in Tp(S(B)) \mid t \text{ начинается с } \uparrow\} \cup Spectr,$$

где $Spectr = \{[\uparrow сущи], [\uparrow пон], [\uparrow об]\}$.

Поясним обозначения, используемые ниже в определении правила $P[8]$. Каждый такой элемент s первичного информационного универсума $X(B)$, что $tp(s) \in Tconc(B)$, будем интерпретировать как обозначение понятия. Множество $R_2(B)$ состоит из бинарных реляционных символов (некоторые из них могут соответствовать функциям с одним аргументом); $F(B)$ – множество функциональных символов. Элемент $ref = ref(B)$ из $X(B)$ называется *квантором референтности* (см. раздел 2.8) и интерпретируется как информационная единица (другими словами, семантическая единица), соответствующая значению слова «некоторый» в выражениях в единственном числе («некоторая книга»,

«некоторая страна» и т.д.); в рассматриваемых примерах $ref(B)$ – это символ *нек*; $P(B)$ – обозначение сорта «смысл сообщения» к.б. B .

Определение 2. Обозначим через $P[8]$ высказывание «Пусть $a \in \in X(B)$, $tp = tp(B)$, $t \in Tconc(B)$, $t = tp(a)$, $P = P(B)$, $ref = ref(B)$. Пусть $n \geq 1$, $\forall m = 1, \dots, n$ $r_m \in R_2(B)$, c_m – цепочка вида $ref a$, $d_m \in L(B)$, h_m – цепочка вида $(r_m(c_m) \equiv d_m)$ в случае $r_m \in R_2(B) \cap \cap F(B)$, и h_m – цепочка вида $r_m(c_m, d_m)$ в случае $r_m \in R_2(B) \setminus F(B)$; если $r_m \in F(B)$, то $h_m \& P \in T^3(B)$; если $r_m \notin F(B)$, то $h_m \& P \in \in T^4(B)$. Пусть b – цепочка вида $a^*(r_1, d_1) \dots (r_n, d_n)$. Тогда $b \in \in L(B)$, $b \& t \in T^8(B)$, $a \& h_1 \& \dots \& h_n \& b \in Y^8(B)$ ».

Пример 1. Предположим, что $B1$ – к.б., определенный в разделе 2.8, $i = 8$. Тогда рассмотрим возможный путь построения формулы b_1 (задаваемой ниже), соответствующей понятию «туристическая группа, состоящая из 12 человек». При этом будем использовать обозначения, введенные в конце раздела 3.3.

Пусть $a = тур.гр$, $t = tp(a) = \uparrow\{интс*дин.физ.об\}$, $P = сообщ$, $ref = нек$, $n = 1$, $r_1 = Колич$, $c_1 = ref a = нек тур.гр$, $d_1 = 12$, $h1 = (r(c) \equiv \equiv d) = (Колич(нек тур.гр) \equiv 12)$.

Тогда $B_1(0, 1, 2, 3) \Rightarrow h_1 \in Lnr_3(B_1)$, $h_1 \& сообщ \in Tnr_3^8(B_1)$.

Пусть $b_1 = a^*(r_1, d_1) = тур.гр^*(Колич, 12)$. Тогда из $P[8]$ следует, что $b_1 \in Lnr_8(B_1)$, $b_1 \& \uparrow\{инс*дин.физ.об\} \in Tnr_8^8(B_1)$.

3.6.2. Построение составных обозначений объектов

Правило $P[8]$ можно использовать для построения составных обозначений разных предметов, ситуаций и множеств предметов или ситуаций. Для построения составного обозначения предмета после правила $P[8]$ применяется правило $P[1]$ и строится выражение вида $ref a^*(r_1, d_1) \dots (r_n, d_n)$. Например, таким образом можно построить выражения

нек город (Страна, Россия) (Колич.жит, 350000), нек чел(Фам, «Сомов»)(Имя, «Петр»).*

Пример 2. Пусть $b_2 = нек биолог^*(Элем, нек тур.гр^*(Колич, 12))$. Тогда $B_1(0, 1, 2, 3, 8, 1, 1, 4, 8, 1) \Rightarrow b_2 \in Lnr_8(B_1)$, $b_2 \& интс*дин.физ.об \in Tnr_81(B1)$.

Цепочку b_2 будем интерпретировать как составное обозначение какого-то (вполне определенного) человека, являющегося биологом и входящего в состав тургруппы из 12 человек. Подцепочку $нек тур.гр^*(Колич, 12)$ будем интерпретировать как обозначение какой-то конкретной (в контексте ситуации общения) тургруппы, состоящей из 12 человек.

Пример 3. Исходя из предположений Примера 1, рассмотрим путь построения возможного СП фразы «А.Зубов включил И.Семенова в туристическую группу, состоящую из 12 человек». Пусть переменная x_1 обозначает момент времени, и $b_3 = (Включил(А.Зубов, И.Семенов, x_1, нек тур.гр*(Колич, 12)) \wedge Раньше(x_1, Сейчас))$.

Тогда $B_1(0, 1, 2, 3, 8, 1, 4, 4, 7) \Rightarrow b \in Lnr_8(B_1), b_3 \& сообщ \in Tnr_8^7(B_1)$.

Пример 4. Пусть $T1 = \langle \text{П.Сомов знает, что И.Семенов является директором фирмы, персонал которой включает 38 человек} \rangle$. Это предложение включает составное обозначение фирмы с персоналом из 38 человек. Пусть $b_4 = \text{Знает}(\text{П.Сомов, Сейчас, } ((\text{И.Семенов} \equiv \text{Директор}(\text{нек фирма}*(\text{Описание, P1}): x_1)) \wedge (P1 \equiv (\text{Колич}(\text{Персонал}(x_1)) \equiv 38))))$.

Проследим путь вывода формулы b_4 с помощью правил $P[0], P[1], \dots, P[8]$.

Пусть $i = 8, B_1$ – к.б., построенный в разделе 2.8. Рассмотрим новый к.б. B_2 , отличающийся от базиса B_1 тем, что

$$X(B_2) = X(B_1) \cup \{\text{фирма, Описание}\},$$

$$tp(\text{фирма}) = \uparrow орг*простр.об*интс, tp(\text{Описание}) = \{([ob], P)\}.$$

Если $a = \text{фирма}, t = орг*простр.об*интс$, то, очевидно, $a \& t \in T^0(B_2)$. По определению к.б. B_2 , множество переменных $V(B_2)$ включает такую переменную $P1$, что $tp(P1) = P(B_2) = сообщ$.

Пусть $n = 1, r_1 = \text{Описание}, c_1 = ref a = \text{нек фирма}, h_1 = r1(c1, d1) = \text{Описание}(\text{нек фирма, P1}), b_5 = a*(r_1, d_1) = \text{фирма}*(\text{Описание, P1})$.

Тогда из правила $P[8]$ и правил $P[0], P[1], P[4]$ следует, что

$$b_5 \in Lnr_i(B_2), b_1 \& \uparrow орг*простр.об*интс \in Tnr_i^8(B_2).$$

Пусть $b_6 = ref b1 = \text{нек фирма}*(\text{Описание, P1})$. Тогда

$$B_2(0, 1, 4, 8, 1) \Rightarrow b_6 \in Lnr_i(B), b_6 \& орг*простр.об*интс \in Tnr_i^1(B_2).$$

Пусть $b_7 = b_6: x_1 = \text{нек фирма}*(\text{Описание, P1}): x_1$. Тогда

$$B_2(0, 1, 4, 8, 1, 5) \Rightarrow b_7 \in Lnr_i, b_7 \& орг*простр.об*интс \in Tnr_i^5.$$

Пусть $b_8 = (P1 \equiv (\text{Колич}(\text{Персонал}(x_1)) \equiv 38))$. Тогда

$$B_2(0, 2, 2, 3, 3) \Rightarrow b_8 \in Lnr_i, b_8 \& сообщ \in Tnr_i^3.$$

Пусть $b_9 = (\text{И.Семенов} \equiv \text{Директор}(b_7))$, тогда

$$B_2(0, 1, 4, 8, 1, 5, 2, 3) \Rightarrow b_9 \in Lnr_i, b_9 \& сообщ \in Tnr_i^3.$$

Пусть $b_{10} = \text{Знает}(\text{П.Сомов, Сейчас, } (b_9 \wedge b_8))$. Тогда

$$B_2(0, 1, 4, 8, 1, 5, 2, 2, 3, 7, 4) \Rightarrow b_{10} \in Lnr_i, b_{10} \& сообщ \in Tnr_i^4.$$

Легко видеть, что b_{10} совпадает с b_4 . Значит, мы построили вывод формулы b_4 . Рассмотренный метод построения СП текста $T1$ является весьма общим; этот метод может использоваться в самых разнообразных случаях для построения СП предложений со сложными причастными оборотами и придаточными определительными предложениями.

3.7. Использование в формулах кванторов существования и всеобщности. Построение обозначений упорядоченных наборов

3.7.1. Применение кванторов существования и всеобщности

Правило $P[9]$ позволяет строить формулы с кванторами $\exists$, $\forall$, похожие на формулы логики предикатов первого порядка. Отличие, в частности, заключается в том, что явным образом ограничивается область действия кванторов, и переменные могут обозначать не только предметы, числа, но и множества различных сущностей. С помощью $P[9]$ можно построить l -формулы вида $Q \ v \ (concept) \ A \ (v)$, где $Q \in \{\exists, \forall\}$, $v \in V(B)$ – переменная; *concept* – это простое обозначение понятия (в этом случае *concept* – такой элемент первичного информационного универсума $X(B)$, что $tp(concept)$ начинается с символа $\uparrow$) или составное обозначение понятия; например, *concept* = страна*(Место, Европа), $A(v)$ – формула, включающая v и интерпретируемая как СП высказывания.

Пример 1. Выбирая подходящий к.б. B , с помощью $P[9]$ и нескольких других правил можно построить СП предложения «В каждой стране Европы есть город с количеством жителей, превышающим 30 тысяч человек» следующим образом:

$$\forall x1(страна^*(Место, Европа)) \exists x2(город) (Место(x2, x1) \wedge \wedge Меньше(30000, Колич.элемент(Жители(x2))))).$$

Определение 1. Через $P[9]$ обозначим высказывание «Пусть $Q \in \{\exists, \forall\}$, $A \in L(B)$, $P = P(B)$, $k \in \{3, 4, 6, 7, 9\}$, $A \ \& \ P \in T^k(B)$, $v \in V(B)$, $tp(v) = [сущи]$ – базовый тип «сущность», A включает символ v , $m \in \{0, 8\}$, *concept* $\in L(B) \setminus V(B)$, $u \in Tconc(B)$, где $Tconc(B)$ – множество всех типов из $Tr(S(B))$, начинающихся с символа $\uparrow$; цепочка вида *concept* $\& \ u$ входит в множество $T^m(B)$, цепочка A не включает подцепочек видов $:v, \forall v, \exists v$, и цепочка A не имеет окончания вида $:z$, где z – произвольная переменная из $V(B)$. Пусть $b = Q \ v \ (concept) \ A$. Тогда $b \in L(B)$, $b \ \& \ P \in T^9(B)$, $Q \ \& \ v \ \& \ concept \ \& \ A \ \& \ b \in Y^9(B)$ ».

Условие « $k \in \{3, 4, 6, 7, 9\}$ » означает, что перед использованием правила $P[9]$ (т.е. перед присоединением к формуле квантора существования или всеобщности) должно применяться одно из правил $P[3]$, $P[4]$, $P[6]$, $P[7]$ или $P[9]$.

Пример 2. Пусть $i = 9$ и существует такой к.б. B , что выполняются следующие соотношения:

город, страна, Европа, Место, Колич.элемент., Жители, Меньше $\in X(B)$,
простр.об., нат.чис., дин.физ.об., интс $\in St(B)$,
 $tp(\text{город}) = tp(\text{страна}) = \uparrow \text{простр.об.}$, $tp(\text{Европа}) = \text{простр.об.}$,
 $tp(\text{Место}) = \{(\text{простр.об.}, \text{простр.об.})\}$,
 $tp(\text{Колич.элемент.}) = \{(\{сущн\}), \text{нат.чис.})\}$,
 $tp(\text{Жители}) = \{(\text{простр.об.}, \{\text{дин.физ.об.} * \text{интс}\})\}$,
 $tp(\text{Меньше}) = \{(\text{нат.чис.}, \text{нат.чис.})\}$, *Колич.элемент., Жители* $\in F(B)$,
 $x_1, x_2 \in V(B)$, $tp(x_1) = tp(x_2) = [сущн]$, $30000 \in X(B)$,
 $tp(30000) = \text{нат.чис.}$, *сообщ* $= P(B)$.

Перечисленные информационные единицы интерпретируются следующим образом: *простр.об.*, *нат.чис.*, *дин.физ.об.*, *интс* – сорта «пространственный объект», «натуральное число», «динамический физический объект», «интеллектуальная система»; *город*, *страна* – обозначения одноименных понятий; *Европа* – обозначение региона Европа; *Место* – обозначение бинарного отношения, связывающего пространственные объекты; *Колич.элемент.* – обозначение функции «Количество элементов множества»; *Жители* – обозначение функции, ставящей в соответствие населенному пункту множество всех его жителей; *Меньше* – обозначение бинарного отношения «меньше» на множестве натуральных чисел.

Пусть $Q_1 = \exists, v_1 = x_2$, $concept_1 = \text{город}$, $A_1 = (\text{Место}(x_2, x_1) \wedge \wedge \text{Меньше}(30000, \text{Колич.элемент.}(\text{Жители}(x_2))))$, $b_1 = Q_1 v_1 (concept_1) A_1$. Тогда нетрудно проверить, что выполняется следующее соотношение: $B(0, 2, 2, 3, 4, 4, 7, 9) \Rightarrow b_1 \in Lnr_i$, $b_1 \& \text{сообщ} \in Tnr_i^9$. Пусть $q_2 = \forall, v_2 = x_1$, $concept_2 = \text{страна} * (\text{Место}, \text{Европа})$, $A_2 = b_1$, $b_2 = Q_2 v_2 (concept_2) A_2$. Тогда $B(0, 2, 2, 3, 4, 4, 7, 9, 0, 1, 4, 8, 9) \Rightarrow b_2 \in Lnr_{i=9}$, $b_2 \& \text{сообщ} \in Tnr_i^9$, $Q_2 \& v_2 \& concept_2 \& b_1 \& b_2 \in Y_i^9(B)$.

3.7.2. Построение обозначений упорядоченных наборов

Правило $P[10]$ предназначено для построения l -формул вида $\langle a_1, \dots, a_n \rangle$, где $n > 1$ и $a_1, \dots, a_n$ – обозначения некоторых сущностей. Такие формулы будут интерпретироваться как обозначения упорядоченных наборов.

Определение 2. Через $P[10]$ обозначим высказывание: «Пусть $n > 1$, для $m = 1, \dots, n$ выполняются соотношения $a_m \in L(B)$, $u_m \in Tr(S(B))$, $0 \leq k[m] \leq 10$, $a_m \& u_m \in T^{k[m]}(B)$. Пусть t – цепочка вида $(u_1, u_2, \dots, u_n)$, b – цепочка вида $\langle a_1, \dots, a_n \rangle$. Тогда $b \in L(B)$, $b \& t \in T^{10}(B)$, $a_1 \& a_2 \& \dots \& a_n \& b \in Y^{10}(B)$ ».

Пример 3. Пусть B_1 – к.б., построенный в разделе 2.8, $i = 10$, b_3 – цепочка ($\text{Элем}(x3, S1) \equiv ((x3 \equiv \langle \text{нек вещь} : x1, \text{нек вещь} : x2 \rangle) \wedge \wedge (\text{Меньше}(x1, x2) \vee (x1 \equiv x2)))$). Тогда $S1$ можно интерпретировать как обозначение отношения « $\leq$ » на множестве вещественных чисел. То есть S_1 обозначает множество всех таких пар $(x1, x2)$, что $x1, x2$ – вещественные числа, и $x1 \leq x2$. Легко проверить, что

$B(0, 4, 1, 5, 1, 5, 4, 10, 3, 7, 7, 3) \Rightarrow b_3 \in \text{Lnr}_i(B_1), b_3 \& \text{сообщ} \in \text{Tnr}_i^{10}(B_1)$.

3.7.3. Сводная таблица правил $P[0]$ – $P[10]$

Суммарный объем определений правил $P[0]$ – $P[10]$ и поясняющих их примеров довольно велик. При построении семантических представлений (СП) не только связных текстов, но и большинства отдельных предложений обычно используется значительная часть этих правил, причем в самых разнообразных комбинациях (анализу возможных применений этих правил посвящена глава 4). В связи с этим представляется целесообразным дать сжатую, недетализированную характеристику каждого правила из списка $P[0]$ – $P[10]$ в приводимой ниже сводной таблице. Эта таблица облегчит анализ использования правил $P[0]$ – $P[10]$ при построении СП ЕЯ-текстов и при формировании фрагментов знаний о мире в примерах этой и последующих глав.

Таблица 3.1

Краткая характеристика правил $P[1]$ – $P[10]$

Правило	Результаты применения
$P[0]$	Начальный запас формул, определяемый первичным информационным универсумом $X(B)$, множеством переменных $V(B)$ и отображением tp , задающим типы элементов из этих множеств
$P[1]$	l -формулы вида $q \ a$ или вида $q \ a^*(r_1, d_1) \dots (r_n, d_n)$, где q – интенсивный квантор, a – простое обозначение понятия, $1 \leq n$, $r_1 \dots r_n$ – характеристики сущностей
$P[2]$	l -формула вида $f(a_1, \dots, a_n)$, $1 \leq n$, где f – имя функции; t – формулы вида $f(a_1, \dots, a_n) \& t$, где t – тип значения функции f для аргументов $a_1, \dots, a_n$
$P[3]$	l -формулы вида $(a_1 \equiv a_2)$ и t -формулы вида $(a_1 \equiv a_2) \& P$, где P – сорт «смысл сообщения»
$P[4]$	l -формулы вида $r(a_1, \dots, a_n)$ и t -формулы вида $r(a_1, \dots, a_n) \& P$, где $n \geq 1$, r – n -арный реляционный символ

Правило	Результаты применения
$P[5]$	l -формула вида $form : v$, где v – метка формулы $form$
$P[6]$	по l -формуле $form$ строится l -формула $\neg form$ (отрицание)
$P[7]$	по логической связке $s \in \{ \wedge, \vee \}$ и l -формулам $a_1, \dots, a_n$ строится l -формула $(a_1 s a_2 s \dots a_n)$, где $n > 1$
$P[8]$	по простому обозначению понятия $conc$, характеристикам объектов $r_1, \dots, r_n$ ($n \geq 1$), l -формулам $d_1 \dots d_n$ строится l -формула $conc^*(r_1, d_1) \dots (r_n, d_n)$ и t -формула $conc^*(r_1, d_1) \dots (r_n, d_n) : t$, где t начинается с символа $\uparrow$. Такие формулы интерпретируются как составные обозначения понятия. Пример: $страна^*(Место, Европа) \& \uparrow простр. объект$
$P[9]$	строются l -формулы вида $Q \vee (conc)A$ и t -формулы $Q \vee (conc)A \& P$, где $Q \in \{ \exists, \forall \}$, $v \in V(B)$, $conc$ – простое или составное обозначение понятия, A – l -формула, обозначающая высказывание, P – сорт «смысл сообщения»
$P[10]$	Для построения l -формул вида $\langle a_1, \dots, a_n \rangle$, где $n > 1$, интерпретируемых как обозначения упорядоченных наборов

3.8. СК-языки. Математическое исследование их свойств

Определение 1. Пусть B – произвольный концептуальный базис, тогда:

(а) $D(B) = X(B) \cup V(B) \cup \{ ', (' , ') ', ': ', '* ', '< ', '> ' \}$;

(б) $Ds(B) = D(B) \cup \{ '&' \}$;

(в) $D^+(B)$ и $Ds^+(B)$ – множества всех непустых конечных последовательностей элементов из $D(B)$ и $Ds(B)$ соответственно.

Таким образом, $Ds(B) = X(B) \cup V(B) \cup \{ ', (' , ') ', ': ', '* ', '< ', '> ', '&' \}$; каждое из множеств $D(B)$, $Ds(B)$ включает, в частности, символ «запятая».

Определение 2. Пусть B – произвольный к.б., $1 \leq i \leq 10$, и множества цепочек $L(B) \subset D^+(B)$; $T^0(B), T^1(B), \dots, T^i(B), Y^1(B), \dots, Y^i(B) \subset Ds^+(B)$ являются наименьшими множествами, совместно задаваемыми правилами $P[0] \dots P[i]$. Тогда обозначим эти множества соответственно через $Lnr_i(B)$, $T^0(B)$, $Tnr_i^1(B)$, $\dots$, $Tnr_i^i(B)$, $Ynr_i^1(B)$, $\dots$, $Ynr_i^i(B)$ и обозначим семейство (т.е. множество), со-

стоящее из всех этих множеств, через $\text{Globset}_i(B)$. Кроме того, пусть

$$T_i(B) = T_0(B) \cup \text{Inr}_i^1(B) \cup \dots \cup \text{Inr}_i^i(B), \quad (3.8.1)$$

$$Y_i(B) = \text{Inr}_i^1(B) \cup \dots \cup \text{Inr}_i^i(B), \quad (3.8.2)$$

$$\text{Form}_i(B) = \text{Inr}_i(B) \cup T_i(B) \cup Y_i(B). \quad (3.8.3)$$

Определение 3 (итоговое). Если B – произвольный к.б., то

$$Ls(B) = \text{Inr}_{10}(B), \quad (3.8.4)$$

$$Ts(B) = T_{10}(B), \quad (3.8.5)$$

$$Ys(B) = Y_{10}(B), \quad (3.8.6)$$

$$\text{Forms}(B) = \text{Form}_{10}(B), \quad (3.8.7)$$

$$Ks(B) = (B, Rls), \quad (3.8.8)$$

где $Rls = \{P[0], P[1], \dots, P[10]\}$.

Упорядоченная пара $Ks(B)$ называется K -исчислением (концептуальным исчислением) в базисе B ; элементы множества $\text{Forms}(B)$ называются формулами, выводимыми в базисе B . Формулы из $Ls(B)$, $Ts(B)$ и $Ys(B)$ называются соответственно l -формулами, t -формулами, y -формулами. Множество l -формул $Ls(B)$ называется стандартным концептуальным языком (стандартным K -языком, СК-языком) в базисе B .

Утверждение 3.1. Если B – произвольный к.б., то (а) множество $\text{Inr}_0(B)$ не является пустым; (б) если $1 \leq i \leq 10$, то $\text{Inr}_{i-1}(B) \subseteq \text{Inr}_i(B)$.

Доказательство.

(а) Для любого к.б. B первичный информационный универсум $X(B)$ включает непустое множество сортов $St(B)$, причем $\forall s \in St(B)$, $tp(s) = \uparrow s$. Тогда при $i = 0$ из правила $P[0]$ следует, что $\text{Inr}_0(B)$ включает $X(B)$ и, как следствие, включает $St(B)$. Поэтому множество $\text{Inr}_0(B)$ непусто.

(б) Структура правил $P[1]–P[10]$ показывает, что добавление нового правила может либо не изменить множество выводимых формул, либо его расширить. В частности, если $i = 2$, то в случае, когда множество функциональных символов $F(B)$ пусто, $\text{Inr}_{i-1}(B) = \text{Inr}_i(B)$.

Утверждение 3.2. Если B – произвольный к.б., то множества $Ls(B)$, $Ts(B)$, $Ys(B)$ не являются пустыми.

Доказательство.

Из Утверждения 3.1 вытекает, что $\text{Inr}_0(B) \subset Ls(B)$. Поэтому $Ls(B)$ непусто. Из определения концептуально-объектной системы следует, что существуют такие различные переменные $v_1, v_2 \in V(B)$, что $tp(v_1) = tp(v_2) = [сущн]$. Тогда из правила $P[3]$ получаем, что цепочка $(v_1 \equiv v_2) \ \& \ P$ входит в множество $Ts(B)$, и $v_1 \ \& \ v_2 \ \& \ (v_1 \equiv v_2) \in Ys(B)$, где $P = P(B)$ – сорт «смысл сообщения». Поэтому множества $Ts(B)$ и $Ys(B)$ не являются пустыми.

Утверждение 3.3. Если B – произвольный концептуальный базис, то:

(а) Если $\tau \in Ts(B)$, то τ – цепочка вида $\alpha \& t$, где $\alpha \in Ls(B)$, $t \in Tr(S(B))$, и такое представление, зависящее от τ , единственно для каждой цепочки τ ;

(б) Если $\gamma \in Ys(B)$, то найдутся такое $n > 1$ и такие цепочки $\alpha_1, \alpha_2, \dots, \alpha_n, \beta \in Ls(B)$, что γ – цепочка вида $\alpha_1 \& \alpha_2 \& \dots \& \alpha_n \& \beta$; кроме того, такое представление, зависящее от γ , единственно для любой цепочки $\gamma \in Ys(B)$.

Доказательство. Справедливость этого предложения непосредственно следует из определения концептуального базиса и из структуры правил $P[0]$ – $P[10]$.

Утверждение 3.4. Пусть B – произвольный к.б., $d \in X(B) \cup V(B)$. Тогда не найдутся такие k, n , где $1 \leq k \leq 10, n > 1$, и такие $\alpha_1, \alpha_2, \dots, \alpha_n \in Ls(B)$, что

$$\alpha_1 \& \alpha_2 \& \dots \& \alpha_n \& d \in Ynr_{10}^k(B). \quad (3.8.9)$$

Интерпретация. Смысл утверждения в том, что для каждого элемента d , входящего в первичный информационный универсум $X(B)$ или являющегося переменной из $V(B)$, нельзя получить этот элемент d с помощью каких-либо операций, задаваемых правилами $P[1]$ – $P[10]$.

Доказательство (от противного).

Предположим, что существуют такие к.б. $B, d \in X(B) \cup V(B)$, натуральное число k , где $1 \leq k \leq 10, n > 1, \alpha_1, \alpha_2, \dots, \alpha_n \in Ls(B)$, что справедливо соотношение (3.8.9). Для любого такого m , что $1 \leq m \leq 10$, $Ynr_{10}^m(B)$ включает цепочку $\alpha'_1 \& \alpha'_2 \& \dots \& \alpha'_n \& d'$, где d' не включает символ $\&$, только в том случае, когда цепочка d' построена из цепочек $\alpha'_1, \alpha'_2, \dots, \alpha'_n$ применением правила $P[m]$ на последнем шаге вывода. Но тогда из структуры правил $P[1]$ – $P[10]$ непосредственно следует, что цепочка d' должна содержать по крайней мере два символа. Но так как $d \in X(B) \cup V(B)$, то элемент d рассматривается как символ и поэтому имеет длину 0. Мы получили противоречие из нашего предположения, что доказывает Утверждение 3.4.

Утверждение 3.5. Пусть B – произвольный концептуальный базис, $z \in Ls(B) \setminus (X(B) \cup V(B))$. Тогда существует один и только один такой набор $(k, n, y_1, y_2, \dots, y_n)$, где $1 \leq k \leq 10, n > 1, y_1, y_2, \dots, y_n \in Ls(B)$, что $y_1 \& y_2 \& \dots \& y_n \& z \in Ynr_1^k(B)$.

Интерпретация. Если l -формула z не входит в $(X(B) \cup V(B))$, то тогда найдутся единственное правило $P[k]$, где $1 \leq k \leq 10$, и единственный такой набор l -формул $y_1, y_2, \dots, y_n$, что цепочка z построена из «блоков» $y_1, y_2, \dots, y_n$ применением ровно один раз правила $P[k]$.

Справедливости Утверждения 3.5 вытекает из двух лемм, рассматриваемых ниже. Для того, чтобы сформулировать эти леммы, потребуется следующее определение.

Определение 4. Пусть B – произвольный концептуальный базис, $n \geq 1$, для $i = 1, \dots, n$ $c_i \in D(B)$, $s = c_1 \dots c_n$, $1 \leq k \leq 10$. Тогда через $lt_1(s, k)$ и $lt_2(s, k)$ обозначим количество вхождений символа '(' и символа '<', соответственно, в подцепочку $c_1 \dots c_k$ цепочки $s = c_1 \dots c_n$. Через $rt_1(s, k)$ и $rt_2(s, k)$ обозначим количество вхождений символа ')' и символа '>' в подцепочку $c_1 \dots c_k$ цепочки s . Если в подцепочку $c_1 \dots c_k$ не входит символ '(' или символ '<', то соответственно $lt_1(s, k) = 0$, $lt_2(s, k) = 0$, $rt_1(s, k) = 0$, $rt_2(s, k) = 0$.

Лемма 1. Пусть B – произвольный к.б., $y \in Ls(B)$, $n \geq 1$, для $i = 1, \dots, n$ $c_i \in D(B)$, $y = c_1 \dots c_n$. Тогда:

(а) при $n > 1$ для каждого $k = 1, \dots, n - 1$ и каждого $m = 1, 2$ $lt_m(y, k) \geq rt_m(y, k)$;

(б) $lt_m(y, k) = rt_m(y, k)$.

Лемма 2. Пусть B – произвольный к.б., $y \in Ls(B)$, $n > 1$, $y = c_1 \dots c_n$, где для $i = 1, \dots, n$ $c_i \in D(B)$, цепочка y включает запятую или какой-либо из символов $\equiv, \wedge, \vee$, и k – такое произвольное натуральное число, что $1 < k < n$. Тогда:

(а) если c_k – один из символов $\equiv, \wedge, \vee$, то $lt_1(y, k) > rt_1(y, k) \geq 0$;

(б) если c_k – запятая, то выполняется по крайней мере одно из соотношений $lt_1(y, k) > rt_1(y, k) \geq 0$, $lt_2(y, k) > rt_2(y, k) \geq 0$.

Доказательства Леммы 1, Леммы 2 и Утверждения 3.5 изложены в Приложении к данной книге.

Определение 5. Пусть B – произвольный концептуальный базис, $z \in Ls(B) \setminus (X(B) \cup V(B))$, и существует такой набор $(k, n, y_0, \dots, y_n)$, где $1 \leq k \leq 10$, $n > 1$, и $y_1, \dots, y_n \in Ls(B)$, что $y_1 \& \dots \& y_n \& z \in Ynr_{10}^k(B)$. Тогда упорядоченный набор вида $(k, n, y_1, \dots, y_n)$ будем называть *формообразующим набором* цепочки z .

С учетом этого определения Утверждение 3.5 говорит о том, что для произвольного к.б. B каждая цепочка $z \in Ls(B) \setminus (X(B) \cup V(B))$ имеет единственный формообразующий набор.

Утверждение 3.6. Пусть B – произвольный концептуальный базис, $z \in Ls(B)$. Тогда существует один и только один такой тип $t \in Tp(S(B))$, что $z \& t \in Ts(B)$.

Доказательство.

Рассмотрим два возможных случая.

Случай 1.

Пусть B – произвольный к.б., $z \in X(B) \cup V(B)$, $t \in Tp(S(B))$, $tp(z) = t$. Тогда из правила $P[0]$ следует, что $z \& t \in Ts(B)$.

Предположим, что w – такой тип из $Tp(S(B))$, что $z \& w \in Ts(B)$. Анализ правил $P[0]–P[10]$ показывает, что такое соотношение может вытекать только из правила $P[0]$. Но тогда w однозначно определяется данным правилом, поэтому w совпадает с t .

Случай 2.

Пусть B – произвольный концептуальный базис, $z \in Ls(B) \setminus (X(B) \cup V(B))$. В силу Утверждения 3.5 найдутся такое натуральное k , где $1 \leq k \leq 10$, $n \geq 1$, и такие $y_0, y_1, \dots, y_n \in Ls(B)$, что цепочка z построена из этих элементов в результате одного применения правила $P[k]$. Поэтому найдется такой тип $t \in Tp(S(B))$, что $z \& t \in Tnr_{10}^k(B)$, и, следовательно, $z \& t \in Ts(B)$.

Из Утверждения 3.5 вытекает, что z определяет однозначным образом такие $k, n, y_0, y_1, \dots, y_n$. Но тогда набор $(k, n, y_0, y_1, \dots, y_n)$ однозначно определяет такой тип u , что $z \& u \in Tnr_{10}^k(B)$. Поэтому u совпадает с t .

Таким образом, Утверждение 3.6 говорит о том, что каждой цепочке СК-языка $Ls(B)$, где B – произвольный к.б., можно поставить в соответствие единственный тип t из $Tp(S(B))$.

Определение 6. Пусть B – произвольный к.б., $z \in Ls(B)$. Тогда типом l -формулы z называется такой элемент $t \in Tp(S(B))$, обозначаемый через $tpl(z)$, что $z \& t \in Ts(B)$.

3.9. Обсуждение построенной математической модели

Анализ литературы показывает, что разработанная модель для описания структурированных значений (СЗ) ЕЯ-текстов обладает целым рядом отличий от известных формальных и «полуформальных» подходов к описанию смысловой структуры текстов на естественном языке.

Представляется заслуживающим внимания обсуждение вопроса не только о содержании разработанной модели (т.е. о характере предлагаемых моделью операций на СЗ текстов), но и о форме новой модели. Поиск адекватной формы модели был сопряжен с рядом трудностей. Во-первых, на определенном этапе исследования стало ясно, что модель должна представлять собою индуктивное определение, задающее одно вспомогательное и десять основных правил построения формул. Однако получавшееся индуктивное определение было слишком слож-

ными для восприятия, поскольку требовало одновременного понимания всех одиннадцати способов построения формул трех видов с учетом взаимодействия этих способов.

Поэтому в данной работе предлагается шаг за шагом задавать расширяющиеся множества формул $Forms_k$, где $k = 0, 1, \dots, 10$, определяемые совместной индукцией некоторыми правилами $P[0], P[1], \dots, P[k]$. Форма каждого из правил $P[k]$, где $k = 0, 1, \dots, 10$, такова, что правило $P[k]$ может использоваться без изменения в определениях множеств $Forms_k, Forms_{k+1}, \dots, Forms_{10}$. Такой подход позволяет детально проиллюстрировать эффект введения нового правила $P[k]$, где $k = 1, \dots, 10$, с точки зрения расширения множества определенных на предыдущем шаге формул $Forms_{k-1}$.

Во-вторых, в логике предикатов отдельно определяются сначала множество термов, а затем множество формул, причем составные термы интерпретируются как обозначения различных объектов. Но мы знаем, что в русском, английском и многих других языках для построения составных обозначений объектов могут, в частности, использоваться причастные обороты и придаточные определительные предложения. Как следствие, структура таких составных обозначений объектов различных видов может быть ничуть не проще, чем структура фраз, выражающих высказывания. Поэтому в данной работе предложено в одном определении задавать множество формул, часть из которых может интерпретироваться как составные обозначения объектов (т.е. как термы).

В-третьих, подход к концептуальному структурированию предметных областей, предложенный в данной работе, существенно отличается от подхода логики предикатов. В логике предикатов исходят из того, что есть некоторые объекты, функции, заданные на объектах, и высказывания об объектах. Множество объектов не структурируется в традиционной логике предикатов и делится на классы (сорты) в многосортных логиках. Функции не могут быть определены на высказываниях (кроме функции, задающей истинностное значение высказывания), и их значениями не могут быть высказывания.

Между тем на высказываниях могут быть определены различные функции, например, *Автор (Авторы)* и *Дата*. С другой стороны, естественным было бы рассматривать и функции, значениями которых являются высказывания. Например, если сущность является понятием, то значением функции *Определение* может быть формула, поясняющая смысл понятия.

В данной работе предложен новый подход к концептуальному структурированию предметных областей. Образно говоря, этот подход обла-

дает значительную большей «разрешающей способностью» по сравнению с подходом логики предикатов. Это увеличение разрешающей способности структурирования действительности в данной работе обеспечивается:

(1) формальным различением (а) обозначений объектов и обозначений понятий, характеризующих эти объекты, (б) обозначений сущностей (предметов, событий, понятий) и обозначений множеств, состоящих из этих сущностей;

(2) наличием средств формального представления упорядоченных наборов и множеств, состоящих из упорядоченных наборов (т.е. n -арных отношений);

(3) возможностью рассматривать функции, аргументами и/или значениями которых могут быть семантические представления ЕЯ-текстов, множества объектов, множества понятий.

В свою очередь, для реализации перечисленных выше идей потребовалось задавать совместной индукцией три класса выводимых формул, а затем каждой формуле одного из этих классов (множества $Ls(B)$, где B – концептуальный базис) поставить в соответствие цепочку, называемую типом данной формулы и интерпретируемую как концептуальную характеристику сущности, обозначаемой рассматриваемой формулой.

Таким образом, предложенный метод пошагового ввода правил построения формул с последующим соединением этих правил с помощью итогового индуктивного определения является оригинальным и может рассматриваться как один из научных результатов данной работы. Этот метод обогащает дискретную математику (в первую очередь – теорию формальных языков) и может использоваться для создания новых исчислений, формулы которых предназначены для описания структуры сложных линейных объектов.

В последние три десятилетия класс языков логики предикатов первого порядка (ЛППП) являлся стандартом, с которым сравнивались предлагавшиеся новые подходы к формальному представлению содержания ЕЯ-текстов. Чаще всего, такие новые подходы рассматривались их авторами как расширения языка ЛППП. Учитывая это, можно выделить следующие преимущества класса стандартных К-языков по сравнению с классом языков ЛППП (при этом будет использоваться нумерация свойств ЕЯ-текстов, введенная в разделе 3.1):

(Свойство 3) возможность строить и формально различать обозначения единиц, соответствующих (а) объектам, ситуациям, процессам в реальном мире и (б) понятиям, квалифицирующим (характеризующим) эти объекты, ситуации, процессы;

(Свойство 4.1) возможность строить и различать обозначения объектов и множеств объектов;

(Свойство 5) возможность различать формальным образом понятия, квалифицирующие объекты, и понятия, квалифицирующие множества объектов тех же видов;

(Свойство 6) возможность строить составные обозначения понятий, т.е. строить формулы, отражающие поверхностно-семантическую структуру ЕЯ-выражений, подобных выражению «человек, окончивший МГУ имени М.В. Ломоносова и являющийся биологом или химиком»;

(Свойство 8) возможность строить обозначения упорядоченных n -местных наборов различных сущностей, где $n > 1$;

(Свойство 9) возможность строить (9.1) формальные аналоги составных обозначений множеств, (9.2) обозначения множеств упорядоченных наборов сущностей, (9.3) обозначения множеств, состоящих из множеств;

(Свойство 11) возможность моделировать смысловую структуру фраз, содержащих, в частности: (11.2) выражения, полученные применением связок «и», «или» к обозначениям (11.2а) предметов, событий; (11.2б) понятий; (11.3) выражения, где связка «не» стоит непосредственно перед обозначением предмета, события, понятия и т.д.; (11.4) косвенную речь; (11.5) причастные обороты и придаточные определительные предложения;

(Свойство 14) возможность моделировать смысловую структуру дискурсов со ссылками на смысл фраз и более крупных фрагментов рассматриваемых текстов содержит семантическое представление дискурса «У А.Зубова есть три друга. П.Сомов это знает»;

(Свойство 17) возможность рассматривать нетрадиционные функции (и другие нетрадиционные отношения) с аргументами и/или значениями, являющимися множествами предметов, ситуаций.

Глава 4

ИССЛЕДОВАНИЕ ВЫРАЗИТЕЛЬНЫХ ВОЗМОЖНОСТЕЙ СК-ЯЗЫКОВ

Проведем дополнительный анализ выразительных возможностей стандартных концептуальных языков (СК-языков) по сравнению с анализом возможностей математического описания структурированных значений ЕЯ-текстов, выполненным в предыдущих разделах. Набор примеров, рассмотренных выше, недостаточно полно демонстрирует реальную мощность построенной модели. Поэтому рассмотрим ряд дополнительных примеров, иллюстрирующих некоторые важные возможности СК-языков.

Если цепочка *Expr* некоторого СК-языка является семантическим представлением выражения *T* на естественном языке, то такую цепочку *Expr* будем называть возможным К-представлением (КП) выражения *T*.

4.1. Удобный способ описания событий

Ключевую роль в формировании предложений играют глаголы и лексические единицы, являющиеся производными от глаголов – причастия, деепричастия и отглагольные существительные, потому что они выражают разнообразные отношения между объектами рассматриваемой предметной области.

Тематической ролью (концептуальным падежом, семантическим падежом, глубинным падежом, семантической ролью) в компьютерной лингвистике называется смысловое отношение между значением глагольной формы и значением зависящей от нее в предложении группы слов (или отдельного слова).

В таких разных языках, как русский, английский, немецкий и французский языки, можно наблюдать следующую закономерность: в предложениях с одним и тем же глаголом, обозначающим событие, явно реализуется разное количество тематических ролей, связанных со зна-

чением данного глагола. Например, пусть $T1 = \langle \text{Профессор Новиков прилетел вчера} \rangle$ и $T2 = \langle \text{Профессор Новиков прилетел вчера из Праги} \rangle$. Тогда в предложении $T1$ явно реализуются тематическая роль, которой можно дать название Агент1 (Агент действия), а также тематическая роль Время. При этом в предложении $T2$ явно реализуются тематические роли Агент1, Время и Место1 (отношение, связывающее событие перемещения в пространстве и исходный пространственный объект).

Рассмотрим столь же гибкий способ построения СП сообщений о событиях. Для этого потребуется сделать определенное предположение о свойствах рассматриваемого концептуального базиса (к.б.) B .

Предположение 1. Множество сортов $St(B)$ включает выделенный сорт sit («ситуация»); множество переменных $V(B)$ включает счетное подмножество $V_{sit} = \{e1, e2, e3, \dots\}$, такое, что для каждого $v \in V_{sit}$ $tp(v) = sit$; первичный информационный универсум $X(B)$ включает бинарный реляционный символ *Ситуация*, такой, что $tp(\text{Ситуация})$ – цепочка $\{(sit, \uparrow sit)\}$.

Смысл выражения $\uparrow sit$ в правой части соотношения $tp(\text{Ситуация}) = \{(sit, \uparrow sit)\}$ заключается в том, что мы сможем строить выражения вида *Ситуация* ($e_k, concept$), где e_k – переменная, обозначающая конкретное событие (продажа, покупка, отлет), *concept* – простое или составное понятие, являющееся семантической характеристикой события.

Уточним, что связь между меткой ситуации и видом ситуации будет осуществляться с помощью формул вида *Ситуация* ($v, conc^* (r_1, d_1) \dots (r_n, d_n)$), где v – переменная типа *sit*, $conc \in X(B)$, *conc* интерпретируется как понятие, характеризующее ситуацию, $n \geq 1$, для $i = 1, \dots, n$ r_i – характеристика ситуации, d_i – значение характеристики.

Пример. Пусть
 $Expr1 = \exists e1(sit) (\text{Ситуация}(e1, \text{прилет}^*(\text{Время}, x1)(\text{Агент1}, \text{нек чел}^*(\text{Квалиф}, \text{профессор})(\text{Фамилия}, \langle \text{Новиков} \rangle) : x2)) \wedge \text{Раньше}(x1, \text{Сейчас}));$
 $Expr2 = \exists e1(sit) (\text{Ситуация}(e1, \text{прилет}^*(\text{Время}, x1)(\text{Агент1}, \text{нек чел}^*(\text{Квалиф}, \text{профессор})(\text{Фамилия}, \langle \text{Новиков} \rangle) : x2)(\text{Место1}, \text{нек город}^*(\text{Название}, \langle \text{Прага} \rangle) : x3)) \wedge \text{Раньше}(x1, \text{Сейчас})).$

Тогда легко видеть, что можно построить такой к.б. B , что для него будет справедливо Предположение 1 и выполнено соотношение

$$B(0, 1, 2, 3, 4, 5, 7, 8, 9) \Rightarrow Expr1, Expr2 \in Ls(B),$$

$$Expr1 \ \& \ \text{сообщ} \in Ts(B), Expr2 \ \& \ \text{сообщ} \in Ts(B),$$

где $\text{сообщ} = P(B)$ – выделенный сорт «смысл сообщения» базиса B .

Данный способ описания сообщений будет многократно использован в этом и последующих разделах. При этом чаще всего, по соображениям компактности, будет опускаться квантор существования при переменной, обозначающей событие. Например, вместо формулы *Expr1* будет рассматриваться формула *Expr3* вида

$$(Ситуация(e1, \text{прилет}*(Время, x1)(Агент1, \text{нек чел}^* \\ *(Квалиф, профессор)(Фамилия, «Новиков») : x2)) \wedge \\ \wedge \text{Раньше}(x1, \text{сейчас})).$$

4.2. Формализация предположений о структуре семантических представлений множеств

Сообщения, вопросы, команды могут включать обозначения множеств. Для обеспечения единства подхода в разных ситуациях (при рассмотрении сообщений, команд, вопросов) к построению СП описаний множеств целесообразно ввести ряд дополнительных предположений об используемых концептуальных базах.

В связи с тем, что обозначения множеств в текстах часто включают количественные числительные или обозначения натуральных чисел («два алюминиевых контейнера» и т.п.), будем полагать, что для рассматриваемого к.б. *В* справедливо

Предположение 2. Множество сортов $St(B)$ включает выделенный сорт *nat* («натуральное число»), первичный информационный универсум $X(B)$ включает подмножество цепочек Nt , такое, что $Nt = \{ \{d_1 \dots d_k\} \mid k \geq 1, \text{ для } i = 1, \dots, k \ d_i - \text{символ из множества } \{ '0', '1', '2', '3', '4', '5', '6', '7', '8', '9' \}, \text{ и из } d_1 = '0' \text{ следует, что } k = 1 \} \}$. При этом для каждого $z \in Nt$ $tp(z) = nat$.

Потребуем также, чтобы первичный информационный универсум $X(B)$ включал выделенные элементы *множ*, *Колич*, *Кач-состав*, *Предм-состав*, интерпретируемые следующим образом: *множ* – это обозначение понятия «конечное множество», *Колич* – имя одноместной функции «Количество элементов множества», *Кач-состав* – имя бинарного отношения «Качественный состав множества», *Предм-состав* – имя бинарного отношения «Предметный состав множества».

Предположение 3. Первичный информационный универсум $X(B)$ включает элементы *множ*, *Колич*, *Кач-состав*, *Предм-состав*, такие, что $tp(\text{множ}) = \uparrow \{ [сущн] \}$, $tp(\text{Колич}) = \{ (\{ [сущн] \}, nat) \}$, $tp(\text{Кач-состав}) = \{ (\{ [сущн] \}, [пон]) \}$, $tp(\text{Предм-состав}) = \{ (\{ [сущн] \}, [об]) \}$, где $[сущн]$, $[об]$, $[пон]$ – базовые типы «сущность», «объект», «понятие» (см. раздел 1.6).

Рассмотрим назначение выделенных элементов универсума $X(B)$, упоминаемых в Предположении 3.

Используя элементы *множ*, *Колич* и произвольную цепочку *numb* из Nt , мы сможем построить СП выражения «некоторое множество, содержащее *numb* элементов» в виде *нек множ*(Колич, numb)*, где *нек* = *ref(B)* – квантор референтности рассматриваемого концептуального базиса B .

Назначение бинарного реляционного символа *Кач-состав* заключается в следующем. Пусть v – переменная, обозначающая некоторое множество, и *conc* – простое или составное обозначение понятия. Тогда выражение *Кач-состав* (v , *conc*) обозначает высказывание «Каждый элемент множества v квалифицируется понятием *conc*», и это высказывание может быть истинным или ложным. Примерами выражений этого вида являются *Кач-состав*($S1$, *контейнер1*), *Кач-состав*($S2$, *статья1*), *Кач-состав*($S3$, *контейнер1*(Материал, алюминий)*), *Кач-состав*($S4$, *статья1*(Область1, биология)*).

С другой стороны, символ *Кач-состав* будет применяться и при построении составных обозначений множеств вида *ref множ*(Кач-состав, conc) : v*, где *ref* – квантор референтности, *conc* – простое или составное обозначение понятия, v – переменная. В частности, концептуальный базис B можно выбрать так, чтобы язык $Ls(B)$ включал выражения

$$\begin{aligned} & \text{нек множ}^*(\text{Кач-состав, контейнер1}) : S1, \text{нек множ}^* \\ & *(\text{Кач-состав, статья1}) : S2, \text{нек множ}^*(\text{Кач-состав, контейнер1}^* \\ & *(\text{Материал, алюминий})) : S3, \text{нек множ}^*(\text{Кач-состав, статья1}^* \\ & *(\text{Область1, биология})) : S4. \end{aligned}$$

Фрагмент текста, обозначающий множество, может представлять собою явное перечисление элементов множества. Таким, в частности, является текст «Два заказчика, АО «Радуга» и ТОО «Зенит», не оплатили сентябрьские поставки».

Бинарный реляционный символ *Предм-состав* предназначен, в частности, для построения выражений вида

$$\text{Предм-состав} (v, (x_1 \wedge x_2 \wedge \dots \wedge x_n)),$$

где $v, x_1, x_2, \dots, x_n$ – переменные, причем v обозначает множество, а $x_1, \dots, x_n$ – это обозначения всех элементов, входящих в состав множества v .

Например, будем считать, что выражение

$$\begin{aligned} & (\text{Предм-состав} (y1, (x_1 \wedge x_2)) \wedge \text{Явл} (x1, \text{АО}) \wedge \text{Явл} (x2, \text{ТОО}) \wedge \\ & \wedge \text{Имя} (x1, \text{«Радуга»}) \wedge \text{Имя} (x2, \text{«Зенит»})) \end{aligned}$$

является СП высказывания «Множество y_1 состоит из АО «Радуга» и ТОО «Зенит», причем первая организация обозначается через x_1 , а вторая – через x_2 .

В то же время мы должны иметь возможность (если это необходимо) строить формулы вида

$$ref \text{ множ}^*(Предм-состав, (x_1 \wedge x_2 \wedge \dots \wedge x_n)) : y_1,$$

где ref – квантор референтности, и y_1, x_1, x_2 – переменные типа [сущн] (базовый тип «сущность»). Например, мы должны располагать возможностью построения выражения $нек \text{ множ}^*(Предм-состав, (x_1 \wedge x_2)) : y_1$.

Пример. Рассмотрим выражения $Вр1 = \langle \text{«3 контейнера с керамикой из Индии»} \rangle$ и $Вр2 = \langle \text{«Партия керамики, состоящая из коробок с номерами 3217, 3218, 3219»} \rangle$. Тогда можно построить такой к.б. B , для которого будут выполнены Предположения 2, 3, и $Ls(B)$ включает формулы

$$(1) \text{ нек множ.}^*(Колич, 3) (Кач-состав, Контейнер1^*$$

$$^*(Содерж1, \text{ нек множ}^*(Кач-состав, изделие^*$$

$$^*(Вид, керамика) (Страна, Индия))));$$

$$(2) (\text{нек партия}2^*(Колич, 3)(Предм-состав, (\text{нек коробка}1^*$$

$$^*(Номер, 3217) : x_1 \wedge \text{нек коробка}1^*(Номер, 3218) : x_2 \wedge \text{нек коробка}1^* \\ ^*(Номер, 3219) : x_3)) : S1.$$

Построенные формулы будем интерпретировать как возможные КП выражений $Вр1$ и $Вр2$; здесь x_1, x_2, x_3 – метки коробок, $S1$ – метка партии.

4.3. Построение семантических представлений вопросов с ролевыми вопросительными словами

Среди всех вопросительных местоимений и наречий можно выделить подмножество, включающее, в частности, слова «кто», «что», «кому», «чем», «когда», «откуда». Чтобы сформулировать свойство, выполняющееся для каждого элемента этого подмножества, введем обозначение nil для пустого предлога. Если в каком-либо вопросе некоторое вопросительное местоимение $qswd$ употреблено без предлога, то условимся говорить, что этому местоимению $qswd$ в данном предложении соответствует пустой предлог.

Для каждого местоимения $qswd$ из рассматриваемого подмножества найдется предлог $prep$ (возможно, он не является единственным), что паре $(prep, qswd)$ соответствует некоторая тематическая роль $role$. Например, парам $(nil, кому)$, $(для, кого)$, $(от, кого)$, могут соответствовать

тематические роли *Адресат*, *Адресат*, *Источник* 1. Таким образом, разным парам (*nil*, *кому*), (*для*, *кого*) соответствует одна тематическая роль *Адресат*, поскольку в равной степени правильными являются фразы «Кому прислана книга?» и «Для кого прислана книга?».

Местоимения и наречия, входящие в указанное подмножество, будем называть *ролевыми вопросительными словами*.

Предположение 4. Первичный информационный универсум $X(B)$ концептуального базиса B включает символ *Вопрос*, и $tp(\text{Вопрос}) = \{([сущн], P)\}$, где $tp = tp(B)$ – отображение, задающее тип информационной единицы, $[сущн]$ – базовый тип «сущность», $P = P(B)$ – выделенный сорт «смысл сообщения».

Пусть для к.б. B выполнено Предположение 4. Тогда СП вопроса с n ролевыми вопросительными словами можно представить в виде $\text{Вопрос}(v_1, A)$ при $n = 1$ и в виде $\text{Вопрос}((v_1 \wedge \dots \wedge v_n), A)$ при $n > 1$, где A – формула, зависящая от переменных $v_1, \dots, v_n$ и отображающая содержание высказывания (т.е. являющаяся семантическим представлением высказывания).

Пример 1. Пусть $B_1 = \text{«Откуда поступил трехтонный алюминиевый контейнер?»}$, Expr_1 – цепочка вида $\text{Вопрос}(x_1, (\text{Ситуация}(e_1, \text{поступление}_2^*(\text{Объект}_1, \text{нек контейнер}^*(\text{Вес}, 3/\text{тонна})(\text{Материал}, \text{алюминий}) : x_2)(\text{Место}_1, x_1)(\text{Время}, t_1)) \wedge \text{Раньше}(t_1, \text{сейчас})))$. Тогда нетрудно построить такой к.б. B , что для B выполняются Предположение 1 и Предположение 4, $P(B) = \text{сообщ}$, и $B(0, 1, 2, 3, 4, 5, 7, 8) \Rightarrow \text{Expr}_1 \in Ls(B)$, $\text{Expr}_1 \in Ts(B)$.

Цепочка Expr_1 является возможным КП вопроса B_1 . В этой цепочке символы x_1, x_2, e_1, t_1 являются переменными, поступление_2 – информационная единица (другими словами, семантическая единица), соответствующая существительному «поступление» и передающая значение «перемещение некоторого физического объекта на пространственный объект» (в отличие от значения «поступление абитуриента в учебное заведение»).

Пример 2. Пусть $B_2 = \text{«Откуда и когда поступил трехтонный алюминиевый контейнер?»}$. Тогда КП вопроса B_2 может являться выражением

$\text{Вопрос}((x_1 \wedge t_1), (\text{Явл}_1(e_1, \text{поступление}_2^*(\text{Объект}_1, \text{нек контейнер}^*(\text{Вес}, 3/\text{тонна})(\text{Материал}, \text{алюминий}) : x_2)(\text{Место}_1, x_1)(\text{Время}, t_1)) \wedge \text{Раньше}(t_1, \text{сейчас})))$.

4.4. Семантические представления вопросов о количестве предметов и о количестве событий

Предположение 5. Первичный информационный универсум $X(B)$ вида (S, Ct, Ql) , где Ql – система кванторов и логических связей вида $(ref, int1, int2, eq, neg, binlog, ext)$, включает элементы *произв*, *все*, *Элем*, такие, что $tp(произв) = int1$, $tp(все) = int2$, $tp(Элем) = \{([сущн], [сущн])\}$.

Элементы *произв*, *все*, *Элем* интерпретируются как информационные единицы «произвольный» («каждый»), «все» и «Элемент множества» (имя отношения «Быть элементом множества»).

Следует заметить, что $int1$ и $int2$ – это выделенные элементы множества сортов $St(B)$. По определению (см. раздел 2.8), элементы $int1$ и $int2$ являются типами интенциональных кванторов соответственно первого и второго видов.

Пример 1. Пусть $B1 =$ «Сколько экземпляров книг А.П.Сомова имеется в библиотеке?». Тогда можно определить такой к.б. B , что для B выполняются Предположение 4 и Предположение 5, и цепочка

Вопрос($x1$, ($x1 \equiv Колич(все\ экземпляр1(Информ-объект, произв\ книга^*$
 $^*(Автор, нек\ чел^*(Инициалы, 'А.П.')(Фамилия,$
 $'Сомов') : x2) : x3)(Место-хранения, нек\ библиотека : x4))))$)

входит в $Ls(B)$. Поэтому данное выражение является возможным К-представлением вопроса $B1$.

Пример 2. Если $B2 =$ «Сколько человек участвовало в создании статистического сборника?», то возможным К-представлением $B2$ является выражение

Вопрос($x1$, (($x1 \equiv Колич(все\ чел^*(Элем, S1))) \wedge Описание(произв\ чел^*$
 $^*(Элем, s1) : y1, (Ситуация(e1, участие1^*(Агент1, y1)(Время, x2)$
 $(Вид-деятельности, создание1^*(Продукт1, нек\ сборник1^*$
 $^*(Область1, статистика)))) \wedge Раньше(x2, \#Сейчас\#))))$).

Пример 3. Пусть $B3 =$ «Сколько книг поступило в январе этого года в библиотеку № 18?». Тогда возможным К-представлением $B3$ является формула

Вопрос($x1$, (($x1 \equiv Колич(все\ книга^*(Элем, S1))) \wedge Описание(произв\ книга^*(Элем, S1) : y1, (Ситуация(e1, поступление2^*(Объект1, y1)$
 $(Время, <01, текущий-год>)(Место2, нек\ библиотека^*$
 $^*(Номер, 18) : x2))))$).

Пример 4. Вопрос В1 = «Сколько раз Иван Михайлович Семенов летал в Мексику?» может иметь следующее возможное КП:

Вопрос(x_1 , ($x_1 \equiv$ Колич (все полет $*(Агент1$, нек чел $*(Имя$, 'Иван') (Отчество, 'Михайлович'))(Фамилия, 'Семенов') : x_2)(Место2, нек страна $*(Название$, 'Мексика') : x_3)(Время, произв момент $*(Раньше$, #сейчас#)))).

4.5. Семантические представления вопросов с формами вопросительно-относительного местоимения «какой»

Метод, предложенный выше для построения К-представлений вопросов с ролевыми вопросительными словами, можно использовать и для построения КП вопросов с различными формами местоимения «какой».

Пример 1. Пусть В1 = «Какое издательство опубликовало роман «Ветры Африки»?». Тогда КП вопроса В1 может являться цепочкой

Вопрос(x_1 , (Ситуация(e_1 , опубликование $*(Время$, x_2) (Агент2, нек издательство : x_1) (Объект3, нек роман $*(Название$, 'Ветры Африки') : x_3)) $\wedge$ Раньше(x_2 , #сейчас#))).

Пример 2. Пусть В2 = «С какими зарубежными издательствами сотрудничает писатель Игорь Сомов?». Тогда КП В2 может являться формулой

Вопрос (S_1 , (Кач-состав (S_1 , издательство $*(Вид-географич$, зарубежное)) $\wedge$ Описание(произв издательство $*(Элем$, S_1) : y_1 , Ситуация(e_1 , сотрудничество $*(Агент1$, нек чел $*(Профессия$, писатель)(Имя, 'Игорь')(Фамилия, 'Сомов') : x_1)(Организация1, y_1)(Время, #сейчас#)))).

4.6. Построение семантических представлений вопросов с ответом «Да»/«Нет»

В лингвистике вопросы с ответом «Да» или «Нет» называются вопросами общеудостоверительного актуально-синтаксического типа. Такие вопросы задаются для того, чтобы в целом удостовериться в правильности имеющейся у спрашивающего информации (Воробьева, Панюшева, Толстой, 1975).

Оказывается, что предложенную выше форму отображения смысла вопросов с ролевыми вопросительными словами можно использовать

и для построения СП вопросов с ответом «Да»/«Нет». Для этого каждый такой вопрос будем интерпретировать как просьбу указать истинное значение некоторого высказывания. Например, вопрос $B1 = \langle \text{«Является ли Гент городом Бельгии»} \rangle$ можно интерпретировать как просьбу найти истинностное значение высказывания «Гент является одним из городов Бельгии». Для реализации этой идеи введем

Предположение 6. $St(B)$ включает выделенный сорт *лог*, называемый «логическая величина»; $X(B)$ включает различные элементы *ист*, *ложь*, причем $tp(ист) = tp(ложь) = лог$, $F(B)$ включает одноместный функциональный символ *Ист-знач*, такой, что $tp(Ист-знач) = \{(P, лог)\}$, где $P = P(B)$ – выделенный сорт «смысл сообщения».

Пример 1. Если для рассматриваемого концептуального базиса B выполняется указанное предположение, то КП вопроса $B1 = \langle \text{«Является ли Гент городом Бельгии?»} \rangle$ может являться формулой

$Вопрос(x1, (x1 \equiv Ист-знач (Элем(нек город*(Назв, «Гент»): x2,$
 $Города(нек страна*(Назв, «Бельгия»): x3))))).$

В этой формуле символ *Города* интерпретируется как имя одноместной функции, ставящий в соответствие стране множество всех городов этой страны, *нек* – квантор референтности $ref(B)$; $x1, x2, x3 \in V(B)$.

Пример 2. Пусть $B2 = \langle \text{«Проходила ли в Азии международная научная конференция “COLING”?»} \rangle$. Тогда К-представлением вопроса $B2$ может являться формула

$Вопрос(x1, (x1 \equiv Ист-знач ((Ситуация(e1, прохождение2*(Событие,$
 $нек конф*(Вид1, междун)(Вид2, научн)(Название, 'COLING'): x2)$
 $(Место, нек континент*(Название, 'Азия'): x3)(Время, x4)) \wedge$
 $\wedge Раньше(x4, \#сейчас\#))))).$

4.7. Отображение смысловой структуры команд

Будем использовать две основные идеи. Во-первых, когда мы говорим о команде (или о приказе, распоряжении и т.д.), то всегда подразумеваем, что имеется одна интеллектуальная система, формирующая команду (обозначается выражением $\#Оператор\#$), и другая интеллектуальная система (или же конечное множество интеллектуальных систем), которая должна выполнить команду (обозначается выражением $\#Исполнитель\#$). Во-вторых, глагол в повелительном наклонении или неопределенную форму глагола будем заменять соответствующим отглагольным существительным.

Предположение 7. Множество сортов $St(B)$ включает выделенные элементы *интс* (сорт «интеллектуальная система»), *мом* (сорт «момент времени»); первичный информационный универсум $X(B)$ включает элементы *Команда*, *#Оператор#*, *#Исполнитель#*, *#сейчас#*, такие что $tr(Команда) = \{(интс, интс, мом, \uparrow_{сит})\}$, $tr(\#Оператор\#) = tr(\#Исполнитель\#) = интс$, $tr(\#сейчас\#) = мом$.

Пример. Пусть $K1 = \langle \text{«Доставь ящик с деталями на склад № 3»}, где K1 - команда, отданная оператором гибкой производственной системы интеллектуальному транспортному роботу. Тогда базис B можно определить так, чтобы выполнялись Предположение 1, Предположение 7 и соотношение$

$$B(0, 1, 2, 3, 4, 5, 8) \Rightarrow Команда(\#Оператор\#, \#Исполнитель\#, \#сейчас\#, доставка1 * (Объект1, нек ящик * (Содерж1, нек множ * (Кач-состав, деталь)) : x1)(Место2, нек склад * (Номер, 3) : x2)) \in Ls(B).$$

4.8. Представление теоретико-множественных отношений и операций на множествах

Пример 1. Пусть $T1a = \langle \text{«Намюр – один из городов Бельгии»}. Тогда рассмотрим текст T1б = \langle \text{«Намюр входит в множество всех городов Бельгии»}.$

Пусть $E1 = Элем(нек город * (Назв, 'Намюр') : x1, Города(нек страна * (Назв, 'Бельгия') : x2))$. Тогда $\exists$ такой к.б. $B1$, что $B1(0, 1, 2, 3, 8, 1, 5, 0, 1, 2, 3, 8, 1, 5, 2, 4) \Rightarrow E1 \in Ls(B1)$. При построении нужно предполагать, что

$$\begin{aligned} &Элем, страна, город, нек \in X(B), \\ &tr(Элем) = \{([суцн], \{[суцн]\})\}, tr(страна) = tr(город) = \uparrow_{простр.об}; \\ &Города \in F1(B1), tr(Города) = \{(простр. об, \{простр. об\})\}, \\ &нек = ref(B) - \text{квантор референтности базиса } B. \end{aligned}$$

Пример 2. Пусть $T2a = \langle \text{«Включи контейнер № 4318 в партию, отправляемую в Тамбов»}. Преобразуем T2a в T2б = \langle \text{«Некоторый оператор распорядился включить контейнер № 4318 в некоторую партию, отправляемую в город Тамбов»}. Тогда построим КП текстов T2a и T2б в виде Команда(\#Оператор\#, \#Исполнитель\#, \#сейчас\#, включение1 * (Объект1, нек контейнер * (Номер, 4318) : x1) (Целевое.множество, нек партия2 * (Место-назн, нек город * (Название, 'Тамбов') : x2) : S1)),$

где $S1$ – метка партии продукции. Аналогично можно представить распоряжения о разделении множества объектов на несколько частей и об объединении нескольких множеств в одно, например, при перегрузке деталей из нескольких ящиков в один.

4.9. Представление смысла фраз с придаточными предложениями цели и с косвенной речью

Пример 1. Пусть $T1 =$ «Сергей поступил в ГУ-ВШЭ, чтобы получить специальность “Бизнес-информатика”», и

$Sr1 = (Ситуация(e1, поступление1 * (Агент, нек чел *$

$*(Имя, «Сергей») : x1)(Уч.заведение, нек вуз *$

$(Название, 'ГУ-ВШЭ') : x2)(Время, t1)(Цель, получение1 *$

$*(Квалификация, нек специальность *(Название,$

$'Бизнес-информатика') : x3)) \wedge Раньше(t1, \#сейчас\#))$.

Тогда $\exists$ такой к.б. B , что $B(0, 1, 2, 3, 4, 5, 7, 8) \Rightarrow Sr1 \in Ls(B)$, $Sr1 \& Сообщ \in Ts(B)$.

Пример 2. Пусть $T2 =$ «Директор сказал, что на февраль запланирована реорганизация фирмы» и

$Sr2 = (Ситуация(e1, устное-сообщение *$

$*(Агент1, Директор(нек организация: x1))(Время, t1)(Содержание1,$

$Планируется(нек реорганизация *(Объект2, нек фирма: x1),$

$Ближайший-месяц(февраль, t1)))) \wedge Раньше(t1, \#сейчас\#))$.

Тогда легко построить такой к.б. B , что выполняются соотношения $B(0, 2) \Rightarrow Ближайший-месяц(февраль, t1) \in Ls(B)$, $Ближайший-месяц(февраль, t1) \& \text{врем.интервал} \in Ts(B)$; $B(0, 1, 2, 4, 5, 7, 8) \Rightarrow Sr2 \in Ls(B)$, $Sr2 \& \text{сообщ} \in Ts(B)$.

Глава 5

ИССЛЕДОВАНИЕ ВОЗМОЖНОСТЕЙ ПРИМЕНЕНИЯ АППАРАТА СК-ЯЗЫКОВ К ПОСТРОЕНИЮ ОНТОЛОГИЙ

5.1. Онтологии и их значение для глобальных информационных сетей

Работа прикладной интеллектуальной системы существенным образом зависит от ее базы знаний. Еще с 1970-х гг. развиваются исследования по разработке все более совершенных языков представления знаний (ЯПЗ) в интеллектуальных системах. Важный класс ЯПЗ составляют терминологические языки представления знаний. В отличие от языка логики предикатов, в терминологических ЯПЗ есть специальные единицы, являющиеся обозначениями понятий, и есть средства построения из таких единиц составных обозначений понятий.

Например, во второй половине 1980-х и начале 1990-х гг. в Германии по заказу фирмы IBM был осуществлен проект LILOG (LInguistics & LOGic), реализованный институтом представления знаний (Штутгарт) совместно с несколькими университетами. В рамках этого проекта были разработаны новые средства для представления знаний о мире и для проектирования ЕЯ-диалоговых систем (Pletat, von Luck, 1990; Pletat, 1991; Herzog, Rollinger, 1991).

Терминологические языки представления знаний в англоязычной литературе называют также KL-ONE-like languages (языками, похожими на язык KL-ONE), потому что первым терминологическим языком представления знаний был язык KL-ONE, разработанный в конце 1970-х – начале 1980-х гг. (Brachman, Schmolze, 1985). Одним из потомков языка KL-ONE стал язык L_{LILOG} , разработанный в проекте LILOG.

Развитие исследований по разработке терминологических ЯПЗ привело в 1990-е гг. к появлению нового значения понятия «онтология». Согласно Большому энциклопедическому словарю под редакцией А.М. Прохорова, изданному в 2000-м г., в философии онтологи-

ей называется учение о бытии (в отличие от гносеологии – учения о познании), в котором исследуются всеобщие основы, принципы бытия, его структура и закономерности.

В работах по информатике онтология понимается как спецификация (т.е. описание) концептуализации (Gruber, 1993; Guarino, 1998; FIPA, 1998b). Термин «концептуализация» используется для указания способа, которым интеллектуальная система структурирует знания о мире, восприятие мира. Спецификация концептуализации дает значения терминам из словаря, используемого интеллектуальной системой для обработки знаний и взаимодействия с другими интеллектуальными системами.

На протяжении последнего десятилетия можно было наблюдать постоянный рост интереса исследователей к построению и изучению онтологий. Причина этого заключается в том, что ученые и разработчики компьютерных систем стали заинтересованы в повторном использовании или/и разделении (совместном использовании) знаний системами. Например, в нашей стране опубликованы, в частности, работы (Гаврилова, Хорошевский, 2000; Гаврилова, 2001; Нариньяни, 2001, 2002) и обстоятельный обзор (Смирнов, Пашкин, Шилов, Левашова, 2002а, 2002б), посвященные созданию и применению онтологий.

Созданию и использованию онтологий для разработки системы автоматизированного контроля смысловой полноты технической документации, описывающей поведение оператора летного экипажа и бортовой аппаратуры в различных полетных режимах, посвящены работы (Добров, Лукашевич и др., 2004; Лукашевич, 2004).

В основе исследования лежит частичный семантико-синтаксический анализ документации технических систем. Сущность анализа заключается в выделении понятий, называемых в тексте или ассоциированных с упоминаемыми в тексте объектами и действиями. Построенная модель знаний названа АвиаОнтологией.

Компьютерные системы используют различные понятия для описания предметных областей. Эти различия создают трудности для применения знаний одной системы в другой системе. Предположим, что мы построим онтологии, которые могут служить основой разработки баз знаний многих систем. В этом случае различные системы смогут применять общую терминологию, а это облегчит разделение и неоднократное использование знаний.

Примерно с начала 1990-х гг. исследователями многих стран ведется поиск эффективных формальных подходов к построению онтологий и средств программной реализации онтологий. В 1990-е и в начале

2000-х гг. наибольшую известность получили компьютерные онтологии CYC (Lenat, 1995; CYC, 2001), LOOM (Loom, 2001), OIL (Fensel et al., 2000; Horrocks, 2000), DAML (DAML, 2001).

В 1990-е гг. исследования, направленные на создание терминологических ЯПЗ и применение их к построению онтологий, привели к возникновению нового научного направления в области математической теории прикладных интеллектуальных систем – дескриптивной логики. Общим для различных вариантов логик, разработанных в рамках данного направления, является то, что важный подкласс рассматриваемых правильно построенных формул образуют простые и составные обозначения понятий.

В связи с развертыванием широкомасштабного проекта Семантической Всемирной Паутины (Semantic Web) роль онтологий еще более возросла. Исследования по терминологическим ЯПЗ и дескриптивным логикам, опыт разработки в 1990-е гг. компьютерных онтологий, а также разработка (в рамках подготовки проекта Семантической Паутины) языка описания метаданных об информационных ресурсах RDF (Resource Description Framework) позволили специалистам Западной Европы и США создать языковую систему для построения онтологий DAML + OIL (Horrocks, van Harmelen, Patel-Schneider, 2001) и затем преобразовать ее в систему OWL (McGuinness, van Harmelen, 2004).

Языковая система OWL (Web Ontology Language) сегодня рассматривается специалистами как важная часть теоретического фундамента Семантической Паутины. Языки DAML + OIL и OWL основываются на идеях логики предикатов первого порядка и идеях, реализованных в дескриптивных логиках.

Несмотря на интенсивность исследования проблемы, выразительные возможности разработанных формальных языков для построения онтологий являются довольно ограниченными. В частности, это относится к построению семантических представлений определений понятий, в которых либо встречается слово «понятие», либо упоминаются множества или назначения объектов или цели интеллектуальных систем, либо содержатся ссылки на смысл фраз и более крупных частей дискурса.

5.2. Анализ некоторых возможностей представления знаний о предметных областях средствами СК-языков

Авторы многих публикаций по информатике отмечают, что перспективный путь автоматизации формирования баз знаний заключается в разработке и использовании лингвистических процессоров для из-

влечения знаний из накопленных во всех областях текстов на естественном языке – монографий, статей, научных и технических отчетов, юридических документов и т.д. (см., в частности, Арлазаров, Журавлев и др., 1998).

Поэтому необходимы значительно более мощные (по сравнению с имеющимися) формальные средства для построения СП (а) ЕЯ-определений понятий и (б) предложений и дискурсов на ЕЯ, выражающих знания о предметной области.

В этой связи представляется целесообразным указать некоторые наиболее важные выразительные возможности СК-языков с точки зрения построения СП определений понятий и формального отображения знаний о предметных областях.

5.2.1. Представление содержания фрагментов знаний о мире

Пример 1. Пусть $T1$ = «Понятие “молекула” используется в физике, химии, биологии». Можно определить такой к.б. B , что множество сортов $St(B)$ включает элемент *область1* и первичный информационный универсум $X(B)$ включает элементы *область1*, *цепочка*, *понятие*, «молекула», *Использ*, *Имя-понятия*, *физика*, *нек.*, *химия*, *биология*, причем типы этих элементов задаются соотношениями

$$\begin{aligned} tr(\text{понятие}) &= [\uparrow \text{пон}], \quad tr(\text{«молекула»}) = \text{цепочка}, \\ tr(\text{физика}) &= tr(\text{химия}) = tr(\text{биология}) = \text{область1}, \\ tr(\text{Использ}) &= \{([\text{пон}], \text{область1})\}, \\ tr(\text{Имя-понятия}) &= \{([\text{пон}], \text{цепочка})\}. \end{aligned}$$

Пусть *нек* – квантор референтности концептуального базиса B , *Использ* и *Имя-понятия* – бинарные реляционные символы, не являющиеся именами функций, и

$$\begin{aligned} s_1 &= \text{Имя-понятия}(\text{нек понятие, «молекула»}); \\ s_2 &= \text{понятие}^*(\text{Имя-понятия, «молекула»}); \\ s_3 &= \text{Использ}(\text{нек понятие}^*(\text{Имя-понятия, «молекула»}), \\ &\quad (\text{физика} \wedge \text{химия} \wedge \text{биология})). \end{aligned}$$

Тогда $B(0, 1, 4) \Rightarrow s_1 \in Ls(B)$; $B(0, 1, 4, 8) \Rightarrow s_2 \in Ls(B)$; $B(0, 1, 4, 8, 1, 0, 7, 4) \Rightarrow s_3 \in Ls(B)$. Построенная формула s_3 является возможным КП для определения $T1$.

Пример 2. Пусть $T2$ = «Тинейджер – это человек в возрасте от 12 до 19 лет»; s – цепочка $((\text{тинейджер} \equiv \text{человек}^*(\text{Возраст}, x1)) \wedge \neg \text{Меньше}(x1, 12/\text{год}) \wedge \neg \text{Больше}(x1, 19/\text{год}))$. Тогда s – возможное КП для $T2$.

Пример 3. В работе (Nebel, Peltason, 1991) сформулировано определение: «Малое и среднее предприятие (*sme*) – это компания с числом служащих не более 50». Это определение может иметь, в частности, следующие К-представления:

$$\begin{aligned} & \text{Определение}(sme, \forall x1(\text{компания1})(\text{Явл1}(x1, sme) \equiv \\ & \equiv \neg \text{Больше}(\text{Колич}(\text{Персонал}(x1)), 50))), \\ & ((sme \equiv \text{компания1} * (\text{Описание}, P1)) \wedge (P1 \equiv \forall x1(\text{компания1}) \\ & (\text{Явл1}(x1, sme) \equiv \neg \text{Больше}(\text{Колич}(\text{Персонал}(x1)), 50)))). \end{aligned}$$

5.2.2. Объектно-ориентированные представления фрагментов знаний

Используя стандартные К-языки, мы можем строить сложные описания объектов и множеств объектов. Например, мы можем построить следующее К-представление описания международного журнала «Informatica»:

$$\begin{aligned} & \text{нек межд-науч-журнал} * (\text{Название}, 'Informatica') (\text{Страна}, \text{Словения}) \\ & (\text{Город}, \text{Любляна}) (\text{Области}, (\text{иск-интеллект} \wedge \text{когнитивная-наука} \wedge \\ & \wedge \text{базы-данных})) : k225, \end{aligned}$$

где *k225* – метка модуля знаний с данными об этом журнале.

Постановка задачи, изложенная в разделе 3.1, предусматривает возможность строить с помощью новых формальных средств концептуальные представления текстов как информационные объекты, отражающие не только смысл, но и значения внешних характеристик текста (метаданные): авторов, дату, области применения изложенных результатов и т.д.

Пример. Используя идею построения К-представлений разнообразных объектов, проиллюстрированную на примере модуля знаний с данными о журнале «Informatica», мы можем построить модуль знаний, содержащий теорему Пифагора и указывающий ее автора и предметную область. Например, подобный модуль может быть следующим выражением некоторого СК-языка:

$$\begin{aligned} & \text{нек информ-объект} * (\text{Вид}, \text{теорема}) (\text{Область}, \text{геометрия}) (\text{Автор}, \\ & \text{Пифагор}) (\text{Содержание}, \forall x1(\text{геом}) \forall x2(\text{геом}) \forall x3(\text{геом}) \forall x4(\text{геом}) \\ & \text{Если-то} ((\text{Явл}(x1, \text{прямоугольн}) \wedge \text{Гипотенуза}(x1, x2) \wedge \text{Катет}((x3 \wedge x4), \\ & x1)), (\text{Квадрат}(\text{Длина}(x2)) \equiv \text{Сумма}(\text{Квадрат}(\text{Длина}(x3)), \\ & \text{Квадрат}(\text{Длина}(x4))))) : k81. \end{aligned}$$

5.3. Моделирование выразительных механизмов основных языков дескриптивной логики

Покажем (примеры 1–3), что СК-языки позволяют моделировать выразительные механизмы основных языков дескриптивной логики (другими словами, терминологических языков представления знаний).

Пример 1. Пусть $T1 = \text{«Тинейджер – это человек в возрасте от 13 до 19 лет»}$. Тогда на языке L_{LILOG} содержание текста $T1$ может быть представлено выражением

$$teenager = person \cap \text{with-feature age in } [13...19].$$

Первым возможным К-представлением (КП) текста $T1$ является выражение

$$((\text{тинейджер} \equiv \text{человек} * (\text{Возраст}, x)) \wedge \neg \text{Меньше}(x, <13, \text{год}>) \wedge \neg \text{Больше}(x, <19, \text{год}>)).$$

Вторым возможным КП текста $T1$ является формула

$$((\text{тинейджер} \equiv \text{человек} * (\text{Возраст}, x)) \wedge \text{Диапазон}(x, \text{год}, 13, 19)).$$

Третье возможное КП текста $T1$:

$$\begin{aligned} &\text{Определение } (\text{тинейджер}, x, (\text{Явл}(x, \text{тинейджер}) \equiv \\ &\equiv (\text{Явл}(x, \text{человек}) \wedge \text{Диапазон}(\text{Возраст}(x), \text{год}, 13, 19))))). \end{aligned}$$

Пример 2. В языке L_{LILOG} семантическое представление текста $T2 = \text{«Порше 911 – это автомобиль с двумя дверями типа кабрио»}$ может выглядеть следующим образом (Pletat, 1991):

$$\text{Constant Porsche-911 : and (car, doors : \{2\}, body: \{cabrio\}).}$$

Возможным К-представлением текста $T2$ является выражение

$$(\text{Porsche-911} \equiv \text{car} * (\text{Doors-number}, 2) (\text{Body-type}, \text{cabrio})).$$

Пример 3. На языке L_{LILOG} информация о том, что различают типы корпуса автомобилей *cabrio*, *coupe*, *hatch-back*, *sedan*, формально представляется выражением

$$\text{Sort body-type;}$$

$$\text{Atoms cabrio, coupe, hatchback, sedan.}$$

Эту же информацию можно представить с помощью следующего выражения некоторого СК-языка:

$$\text{Kinds (body_type, (cabrio } \vee \text{ coupe } \vee \text{ hatch-back } \vee \text{ sedan)).}$$

Мы видим, что логические связки «и», «или» позволяют соединять не только семантические представления высказываний, но и обозначения понятий.

5.4. Построение сложноорганизованных фрагментов знаний о мире

Анализ показывает, что выразительные возможности языка L_{LILOG} и других разработанных терминологических ЯПЗ являются довольно ограниченными. Это касается построения составных обозначений понятий и целей интеллектуальных систем, описания множеств, отображения содержания ЕЯ-текстов со ссылками на смысл фраз и более крупных частей дискурса. В связи с этим рассмотрим некоторые возможности использования СК-языков в подобных случаях.

Пример 1. Пусть *Def1* – определение «A flock (английский язык) – это большое количество птиц или млекопитающих (например, овец или коз), собирающихся вместе с определенной целью, такой, как питание, миграция или оборона». Тогда *Def1* может иметь следующее К-представление *Expr1*:

Определение1 (*flock*, англ-яз, динамич-группа*(*Кач-состав*, (*птица* $\vee$ *млекопитающее**(*Примеры*, (*овца* $\vee$ *коза*))))), *S1*, (*Оценка*(*Колич-элемент*(*S1*), *большое*) $\wedge$ *Цель-формирования* (*S1*, *Нек намерение**(*Примеры*, (*питание* $\vee$ *миграция* $\vee$ *оборона*))))).

Анализ этой формулы позволяет сделать вывод о том, что при построении СП ЕЯ-текстов, являющихся определениями понятий, удобно использовать: (1) обозначение 5-арного отношения *Определение1*, (2) составные обозначения понятий (в примере использованы выражения *млекопитающее**(*Примеры*, (*овца* $\vee$ *коза*)) и *динамич-группа**(*Кач-состав*, (*птица* $\vee$ *млекопитающее**(*Примеры*, (*овца* $\vee$ *коза*))))), (3) имена функций, аргументами и/или значениями которых могут быть множества (в примере использовано имя одноместной функции *Колич-элемент*, значением которой является количество элементов множества), (4) составные обозначения намерений, целей (в примере – выражение *нек намерение**(*Примеры*, (*питание* $\vee$ *миграция* $\vee$ *оборона*))).

Пример 2. В предыдущем примере рассматривалось определение *Def1* = «A flock (английский язык) – это большое количество птиц или млекопитающих (например, овец или коз), собирающихся вместе с определенной целью, такой, как питание, миграция или оборона» и была построена цепочка *Expr1*, являющаяся СП этого определения.

Определение *Def1* взято из определенной книги, опубликованной в определенном году определенным издательством. СК-языки позволяют строить СП определений и других фрагментов знаний в объектно-ориентированной форме, отражая их внешние связи. Например, объек-

тно-ориентированное СП определения *Def1* может являться выражением

нек информ-объект(Вид, определ)(Содержание1, Expr1)*

(Источник1, нек словарь(Название, 'Longman Dictionary of Scientific Usage')(Издательство, (Longman-Group-Limited/Harlow $\wedge$ $\wedge$ Russky-Yazyk-Publishers/Moscow))(Город, Москва)(Год, 1989)).*

Пример 3. Пусть *T3* – определение «Евстахиева труба – это канал, ведущий от среднего уха к глотке». *T3* можно поставить в соответствие, в частности, следующую К-цепочку, интерпретируемую как СП текста *T3*:

Определение1(евстахиева-труба, русск-яз, канал2, x1,

$\exists z(\text{чел}) \text{Вестул}(x1, \text{нек среднее-ухо}(\text{Часть}, z),$*

нек глотка)(Часть, z))).*

Пример 4. Пусть *T4* = «Сфигмоманометр – прибор, предназначенный для измерения кровяного давления», тогда *T4* может иметь следующее КП:

(сфигмоманометр $\equiv$ прибор(Назначение, измерение1 **

**(Парам, кровяное-давление)(Субъект, произв чел))).*

Семантическая единица *Назначение* в этом КП обозначает бинарное отношение. Если пара (А, В) принадлежит этому отношению, то А является физическим объектом, а В – формальным семантическим аналогом выражения, описывающего назначение этого физического объекта.

Пример 5. Пусть *T5* – определение «Тромбин – это фермент, который помогает преобразовать фибриноген в фибрин во время коагуляции». Тогда следующая К-цепочка является возможным К-представлением *T5*:

(тромбин $\equiv$ фермент(Назначение, оказание-помощи*(Действие,*

*преобразование1 *(Исх-объект, нек фибриноген)(Результат1,*

нек фибрин)(Процесс, нек коагуляция))).

Примеры, рассматриваемые ниже, покажут выразительные возможности стандартных К-языков в отношении описания семантической структуры дискурсов.

Пример 6. Рассмотрим текст *T6* = «Адениновая основа на одной нити ДНК связана только с тиминовой основой противоположной нити ДНК. Подобным же образом, цитозиновая основа связана только с гуаниновой основой противоположной нити ДНК».

Для построения КП Т6 полезно следующее пояснение: молекула дезоксирибонуклеиновой кислоты (молекула ДНК) содержит тысячи нуклеотидов (комбинаций из трех основных элементов: дезоксирибозы, фосфатов и основы). Существует четыре вида основ: аденин, гуанин, цитозин и тимин. Нуклеотиды ДНК-молекулы образуют цепочку, которая формирует две длинные нити, сплетенные друг с другом. Приняв во внимание это замечание, с первым предложением из Т6 можно связать КП А1 вида

$$\begin{aligned} & \forall x1 \text{ (днк-молекула)} (\text{Связывать}1(\text{произв основа}1* \\ & (\text{Явл, аденин})(\text{Часть, произв нить}1*(\text{Часть, } x1) : y1) : z1, \\ & \text{нек основа}1*(\text{Явл, тимин})(\text{Часть, нек нить}1* \\ & (\text{Часть, } x1)(\text{Противоположн, } y1) : y2) : z2) \wedge \\ & \wedge \neg \exists z3 (\text{основа}1)(\text{Явл}(z3, (\text{аденин} \vee \text{гуанин} \vee \text{цитозин})) \wedge \\ & \wedge \text{Часть}(z3, y2) \wedge \text{Связывать}(z1, z3)) : P1. \end{aligned} \quad (5.4.1)$$

В строке А1 вида (5.4.1) переменные $y1$ и $y2$ используются как метки описаний двух нитей произвольной молекулы ДНК $x1$; переменные $z1, z2, z3$ помечают основы. Переменная $P1$ (имеет сорт «смысл сообщения») используется для обозначения семантического представления первого предложения из Т6. Это позволяет построить компактное СП второго предложения Т6, так как вхождение выражения «подобным же образом» во второе предложение из Т6 означает ссылку на смысл первого предложения. В частности, второе предложение Т6 в контексте первого предложения может иметь К-представление А2 вида

$$\begin{aligned} & (\text{Подобно}(P1, P2) \wedge (P2 \equiv \forall x1 (\text{днк-молекула}) (\text{Связывать} \\ & (\text{произв основа}1*(\text{Явл, цитозин})(\text{Часть, произв нить}1*(\text{Часть, } \\ & x1) : y3) : z4, \text{нек основа}1*(\text{Явл, гуанин})(\text{Часть, нек нить}1* \\ & (\text{Часть, } x1)(\text{Противоположн, } y3) : y4) : z5) \wedge \neg \exists z6 (\text{основа}1) \\ & (\text{Явл}(z6, (\text{аденин} \vee \text{цитозин} \vee \text{тимин})) \wedge \text{Часть}(z6, y4) \wedge \\ & \wedge \text{Связывать}(z4, z6))))). \end{aligned} \quad (5.4.2)$$

Таким образом, с текстом Т6 можно связать К-цепочку А3 вида $(A1 \wedge A2)$, где А1 и А2 – цепочки видов (5.4.1) и (5.4.2) соответственно. Полученную строку можно рассматривать как возможное К-представление для текста Т6.

Цепочка А3 иллюстрирует важную возможность, предоставляемую стандартными К-языками: можно помечать переменными фрагменты цепочек, являющиеся семантическими представлениями сообщений, неопределенных форм глаголов или вопросов. Эта возможность позво-

ляет эффективно описывать структурированные значения дискурсов со ссылками на смысл фрагментов, являющихся сообщениями, целями (советами, пожеланиями) или вопросами. На наличие подобных ссылок в дискурсах часто указывают слова и словосочетания: «эта рекомендация», «например», «то есть», «рассмотренная идея», «другими словами» и ряд других.

Построенное КП $A3$ для $T6$ иллюстрирует еще одну особенность СК-языков: символ « $\equiv$ » соединяет переменную $P2$ и семантическое представление предложения.

Пример 7. Пусть $T7$ = «Термин “цитозин” используется в генетике». Структурированное значение $T7$ может быть представлено в виде К-цепочки

Используется(нек понятие(Название1, «цитозин»), генетика).*

Следовательно, СК-языки позволяют описывать структурированные значения предложений со словами «понятие», «термин» и т.п.

Таким образом, аппарат СК-языков открывает возможности построения новых терминологических языков представления знаний с очень большой выразительной силой и, как следствие, возможности разработки онтологий в произвольных предметных областях, поскольку дает мощные средства построения составных обозначений объектов, множеств, понятий, целей интеллектуальных систем, а также средства отображения ссылок на ранее упомянутые сущности и на смысл предшествующих фраз и более крупных частей связного текста. Перечисленные особенности СК-языков являются основными преимуществами предложенного подхода к формализации предметных областей по сравнению с языком разработки онтологий OWL, используемым в проекте Семантической Паутины.

Поэтому одним из возможных применений аппарата СК-языков является совершенствование средств построения онтологий в проекте Семантической Паутины.

Глава 6

ИССЛЕДОВАНИЕ ВЫРАЗИТЕЛЬНЫХ ВОЗМОЖНОСТЕЙ ОСНОВНЫХ ПОДХОДОВ К ПРЕДСТАВЛЕНИЮ СОДЕРЖАНИЯ ПОСЛАНИЙ КОМПЬЮТЕРНЫХ ИНТЕЛЛЕКТУАЛЬНЫХ АГЕНТОВ

6.1. Использование СК-языков для представления содержания посланий, соответствующих отдельным фразам естественного языка

Проведенный выше анализ позволяет сделать вывод о том, что СК-языки обладают целым рядом свойств, делающих их удобными для представления содержания посланий компьютерных интеллектуальных агентов (КИА). Рассмотрим некоторые из этих свойств.

Свойство 1. С помощью СК-языков можно строить формальные составные обозначения понятий.

Пример 1. Пусть П1 = «тургруппа, состоящая из 12 ученых», тогда возможным К-представлением (КП) выражения П1 является цепочка

тур-группа(Колич-элементов, 12) (Качеств-состав, ученый).*

Пример 2. Если П2 – выражение «керамика, выпущенная в Индии или Шри-Ланке», то возможно первое КП выражения П2

керамика1(Производство, (Индия ∨ Шри-Ланка))*

и второе КП

керамика1(Производство, (нек страна*(Назв, 'Индия') : x1 ∨
∨ нек страна*(Назв, 'Шри-Ланка') : x2)).*

Пример 3. Пусть П3 = «контейнер, содержащий 8 коробок с чайными сервизами из Китая и 4 коробки со столовыми сервизами из Индии или Шри-Ланки». Тогда найдется такой концептуальный базис *B*, что СК-язык *Ls (B)* включает следующее выражение, являющееся возможным семантическим представлением П3:

контейнер1(Содержание1, (нек множество*(Колич-элементов, 8)
(Кач-состав, коробка1*(Содержание2, сервиз1*(Вид, чайный)*

(Страна, нек страна1*(Название, 'Китай') : x1))) : S1 $\wedge$

$\wedge$ нек множество*(Колич-элементов, 4)

(Кач-состав, коробка1*(Содержание2, сервис1*(Вид, столовый)

(Страна, нек страна1*(Название, («Индия» $\vee$

$\vee$ «Шри-Ланка») : x2))) : S2)).

Свойство 2. СК-языки предоставляют широкие возможности построения определений понятий в виде

$$(a) (c \equiv b^*(r_1, d_1) \dots (r_n, d_n))$$

или

$$(b) ((c \equiv b^*(r_1, d_1) \dots (r_n, d_n)) \wedge A),$$

где c – вводимое понятие, b – базовое понятие (считается известным), $n \geq 1$, $r_1, \dots, r_n$ – бинарные реляционные символы; $d_1, \dots, d_n$ – К-цепочки, A – К-цепочка, интерпретируемая как высказывание о свойствах объектов, характеризующихся понятием c .

Пример 1. Пусть П1 – определение «Freight forward (англ. язык) – это груз, оплачиваемый в порту назначения». Тогда семантическое представление определения П1 может являться следующим выражением некоторого СК-языка:

$$(freight-forward \equiv груз1^*(Описание1, < x1, Оплата (x1, нек порт1^* (Пункт-назначения, x1)) >) (Язык, английский)).$$

Пример 2. Пусть П2 = «Малое предприятие – это предприятие с количеством сотрудников, не превышающим 50 человек». Тогда возможным КП определения П2 является следующее выражение:

$$Определение1 (Малое предприятие, x1, (Явл1 (x1, предприятие) \wedge \neg Больше (Колич-элементов (Штам1 (x1)), 50))).$$

Здесь используется другая форма представления определения:

$$Определение1 (c, v, Des (v)),$$

где c – обозначение понятия (поясняемого), v – переменная (обозначает произвольную сущность, характеризующуюся понятием c), $Des(v)$ – К-цепочка (или l -формула), являющаяся СП высказывания о сущности с меткой v .

Свойство 3. С помощью СК-языков можно строить составные обозначения различных сущностей, в том числе обозначения множеств, для этого сначала строится составное обозначение понятия (см. свойство 1), причем применяется правило P[8], а затем добавляется квантор референтности с помощью правила P[1].

Пример 1. К-цепочка

$$коробка1^*(Содержание2, сервис1^*(Вид, чайный)),$$

построенная в результате применения на последнем шаге правила $P[8]$, интерпретируется как составное обозначение понятия «коробка с чайными сервизами». В результате применения правила $P[1]$ можно получить выражения

нек коробка1(Содержание2, сервис1*(Вид, чайный)),* (6.1.1)

все коробки1(Содержание2, сервис1*(Вид, чайный)).* (6.1.2)

Выражение (6.1.1) будем интерпретировать как обозначение коробки, где находится один или несколько чайных сервизов. Выражение (6.1.2) будем рассматривать как обозначение множества, состоящего из всех коробок, содержащих чайные сервизы.

Пример 2. Мы можем следующим образом обозначить конкретную запланированную серию из 5-ти поставок, каждая из которых включает 60 чайных сервизов № 53 и 36 столовых сервизов № 65:

нек множество(Колич-элементов, 5) (Кач-состав, поставка1*
(Состав2, (нек множество(Колич-элементов, 60) (Кач-состав,
сервис1*(Вид, чайный) (Номер, 53)) $\wedge$ нек множество*
(Колич-элементов, 36)(Кач-состав, сервис1(Вид, столовый)(Номер,
65)))))) : S1.*

Свойство 4. СК-языки удобны для построения семантических представлений простых и составных целей.

Пример 1. Пусть $G1$ – цель «Зарезервировать 12 одноместных номеров в трехзвездочных отелях Любляны». Тогда К-представлением этой цели может являться выражение

Резервир1(Объект1, нек множество*(Колич-элементов, 12)
(Кач-состав, номер1*(Вид1, одноместн) (Место1,
произвольн отель*(Вид2, трехзвездочн)
(Локализация, нек город*(Назв, 'Любляна') : $x1$))) : S1) : goal1.*

Пример 2. Пусть $G2$ = «Доставить фирме Smith & Spencer в течение 15–19 октября 2007 г. 5 партий, каждая из которых состоит из 60 чайных сервизов № 53 и 36 столовых сервизов № 65». Тогда цель (или распоряжение) $G2$ может иметь следующее К-представление:

Доставка1(Адресат1, нек фирма1*(Назв, 'Smith & Spencer') : $x1$)
(Время, <15.10.2007, 19.10.2007>) (Объект1, setdescr1),*

где *setdescr1* – построенное выше КП выражения «серия из 5-ти поставок, каждая из которых включает 60 чайных сервизов № 53 и 36 столовых сервизов № 65».

Свойство 5. СК-языки дают возможность представлять коммуникативные акты языка FIPA ACL. Предположим, что компьютерный интеллектуальный агент (КИА) «Клиент-агент» просит КИА «Антологический агент» (электронный поставщик знаний) сообщить, какие бывают виды цитрусовых. Тогда этот коммуникативный акт можно представить в виде следующего выражения некоторого СК-языка:

нек коммуник-акт(Вид, запрос)(Отправитель, Клиент-агент)
 (Получатель, Онтол-агент) (Содержание1, (Вопрос (x1 ≡
 ≡ нек множество*(Кач-состав, произв понятие*(Конкретизация,
 цитрус)))))(Язык-запроса, СК-язык)(Онтология,
 fipa-ontol-service- fruits-ontology)(Метка-ответа, цитрус-запрос)
 (Язык-ответа, СК-язык).*

Свойство 6. СК-языки удобны для построения СП вопросов.

Пример. Пусть В1 = «Сколько стоит двухтонный контейнер?». Тогда возможное К-представление В1:

Вопрос (x1, (x1 ≡ Цена (произв контейнер1(Вес, 2/ тонна))))).*

Свойство 7. СК-языки позволяют использовать ту же форму для построения общих вопросов, т.е. вопросов с ответом ДА/НЕТ.

Пример. Пусть В2 = «Работает ли Сергей Сомов в фирме IBM?»

Тогда К-представление В2 может являться следующим выражением:

Вопрос (x1, (x1 ≡ Ист-знач. ((∃ e1 (ситуация) Явл1 (e1, работа1
 (Агент1, нек человек(Имя, 'Сергей')(Фамилия, 'Сомов') : x2)
 (Место3, нек фирма*(Назв, 'IBM') : x3)) ∧
 ∧ Время (e1, #сейчас#))))).*

6.2. Явное представление причинно-следственных отношений, передаваемых дискурсами

Как уже отмечалось выше, в компьютерной и теоретической лингвистике дискурсом, или связным текстом, называется последовательность взаимосвязанных по смыслу предложений (полных или неполных). Соответствие между группами слов из текста и теми объектами, событиями, процессами, смыслами, которые эти группы слов обозначают, называется референтной структурой текста. СК-языки предоставляют широкие возможности описания смысловой структуры дискурсов, в том числе их референтной структуры.

Пример. Пусть $T1 =$ «Первокурсник Петр Сомов не заметил, что расписание изменилось, поэтому он пропустил первую лекцию по линейной алгебре». В этом тексте проявляется, в частности, следующая особенность дискурсов: личное местоимение «он» используется вместо более длинного сочетания «первокурсник Петр Сомов». Говорят, что у этого последнего выражения и местоимения «он» есть один и тот же референт – некоторый человек, студент вуза. Чтобы явно указать референтную структуру текста, нужно связать метки с сущностями, обозначаемыми некоторыми группами слов из этого текста или неявно упоминаемыми в тексте. Сделаем это таким образом: неявно упоминаемое учебное заведение – метка $x1$; «Первокурсник Петр Сомов», «он» – метка $x2$; «Расписание» – метка $x3$; «первую лекцию по линейной алгебре» – метка $x4$; «не заметил» – метка $e1$ (событие); «изменилось» – метка $e2$; $e3$ – метка ситуации, описываемой первым предложением из $T1$; «пропустил» – метка $e4$ (событие). Будем полагать, что СП текста $T1$ должно включать фрагмент *Причина*($e3, e4$). Пусть

$$Sr1 = ((\text{Ситуация}(e1, \text{—обращение-внимания}*(\text{Агент1, нек чел}*$$

$$*(\text{Имя, 'Петр'})(\text{Фам, 'Сомов'})(\text{Квалиф, студент}*(\text{Курс,1}$$

$$(\text{Уч-заведение, } x1)) : x2)(\text{Время, } t1)(\text{Объект-внимания, } e2)) \wedge$$

$$\wedge \text{Раньше}(t1, \#сейчас\#) \wedge \text{Ситуация}(e2, \text{изменение}*(\text{Предмет,

$$\text{нек расписание : } x3)(\text{Время, } t2)) \wedge \text{Раньше}(t2, t1)) : P1 \wedge$$

$$\wedge \text{Характеризует}(P1, e3)).$$$$

Тогда $Sr1$ – возможное КП первого предложения $P1$ дискурса $T1$. Пусть

$$Sr2 = ((\text{Ситуация}(e4, \text{пропуск1}*(\text{Агент1, } x2)(\text{Объект3, нек лекция}*$$

$$*(\text{Дисциплина, лин-алгебра})(\text{Уч-заведение, } x1) : x4)(\text{Время, } t3)) \wedge$$

$$\wedge \text{Раньше}(t3, \#сейчас\#)).$$

Тогда $Sr2$ – возможное КП второго предложения $P2$ из дискурса $T1$. Пусть $Srd1 = (Sr1 \wedge Sr2 \wedge \text{Причина}(e3, e4))$. Тогда $Srd1$ – возможное

СП дискурса $T1$, являющееся К-представлением текста $T1$.

6.3. Семантические представления дискурсов со ссылками на смысл фраз и более крупных частей текста

Пример. Пусть $T1 =$ «АО “Радуга” подпишет контракт до 15 декабря. Об этом сообщил заместитель директора Игорь Панов». Здесь сочетание «об этом» обозначает ссылку на смысл первого предложения дискурса $T1$. Пусть

$Sr1 = (\text{Ситуация}(e1, \text{подписание1} * (\text{Агент2, нек организация} * (\text{Тип, АО})(\text{Название, «Радуга»}) : x1)(\text{Время, } t1)(\text{Объект3, нек контракт1} : x2)) \wedge \text{Раньше}(t1, 15/\text{декабрь/текущий-год}));$
 $Srd2 = (Sr1 : P1 \wedge \text{Ситуация}(e2, \text{сообщение1} * (\text{Агент1, нек чел} * (\text{Имя, 'Игорь'})(\text{Фамилия, 'Панов'}) : x3)(\text{Время, } t2)(\text{Содержание2, } P1)) \wedge \text{Раньше}(t2, \# \text{сейчас} \#) \wedge \text{Зам.директора}(x3, \text{нек орг} : x4))).$
 Тогда найдется такой к.б. B , что

$B(0, 1, 2, 3, 4, 5, 7, 8) \Rightarrow Srd2 \in Ls(B), Srd2 \& \text{сообщ} \in Ts(B).$

Правило $P[5]$ позволяет приписать переменную v к СП Sr произвольного повествовательного текста и получить формулу вида $Sr : v$, где v – произвольная переменная сорта $P(B)$ -сорта «смысл сообщения». Поэтому выражениям «об этом», «этот метод», «этот вопрос» и т.д. будет соответствовать переменная v сорта $P(B)$ в СП всего дискурса (так же, как и в последнем примере).

6.4. Проблема формирования контрактов и протоколов переговоров в области электронной коммерции

В течение нескольких последних лет в области электронной коммерции возникли два взаимосвязанных научных направления, получивших названия *электронные переговоры* (*e-negotiations*) и *электронное заключение контрактов* (*electronic contracting*). Рождение этих направлений было формально обозначено проведением в начале 2000-х гг. нескольких международных конференций и симпозиумов, в том числе конференции по электронным контрактам и вычислениям, базирующимся на контрактах (Цюрих, Швейцария, 2001); 6-й международной конференции по информационным системам для бизнеса (Колорадо Спрингс, США, 2003); симпозиума по теории и применениям электронных переговоров (Познань, Польша, апрель 2004); 1-го международного симпозиума по электронному заключению контрактов, организованного Международным институтом инженеров по электричеству и электронике (IEEE) в июле 2004 г. в Сан-Диего, Калифорния, США.

К центральным задачам, стоящим перед исследователями в этих научных направлениях, относится создание формальных языков для представления содержания коммерческих переговоров, проводимых компьютерными интеллектуальными агентами (КИА), и для построения контрактов, заключаемых КИА в ходе таких переговоров. Эти задачи можно рассматривать как важные частные случаи проблемы раз-

работки формальных языков общего назначения для бизнес-коммуникаций (Kimbrough и Moore, 1997; Moore, 1998; Kimbrough, 2001a, 2001b, 2001c; Hasselbring и Weigand, 2001; Abrahams, Eysers, Kimbrough, 2006).

В работе (Hasselbring и Weigand, 2001) подчеркивается, что если послания в области электронной коммерции должны обрабатываться автоматически, то значения (meanings) посланий должны быть формализованы. Эта идея совпадает с высказанным в статье (Kimbrough и Moore, 1997) мнением о необходимости создания логико-семантических основ конструирования формальных языков для бизнес-коммуникаций.

Кимбру и Мур в указанной выше работе предлагают использовать как можно шире аппарат логики первого порядка для построения выражений таких языков. Однако выразительные возможности класса языков логики первого порядка очень ограничены с точки зрения описания семантической структуры произвольных бизнес-документов.

Анализ показывает, что протоколы коммерческих переговоров и контрактов могут формироваться с помощью выразительных механизмов естественного языка (ЕЯ), используемых для построения произвольных ЕЯ-текстов, относящихся к медицине, технике, юриспруденции и т.д. В частности, тексты из таких документов могут включать: (а) предложения с косвенной речью, (б) неопределенные формы глаголов (или инфинитивы) с зависимыми словами, выражающие цели, предложения («продать 50 ящиков с яблоками»), обещания, обязательства и назначения предметов; (в) конструкции, образованные из инфинитивов с зависимыми словами с помощью логических связок «и», «или», «не» и являющиеся составными обозначениями целей, предложений, обещаний, обязательств и назначений предметов; (г) составные обозначения множеств («партия, состоящая из 50 ящиков с яблоками»); (д) фрагментов, в которых логические связки «и», «или» соединяют не обозначения высказываний, а обозначения предметов; (е) фрагментов, содержащих ссылки на смысл фраз или более крупных фрагментов дискурса («это предложение», «его распоряжение», «это обещание» и т.д.); (ж) обозначения функций, аргументами и/или значениями которых могут быть множества объектов («персонал фирмы А», «поставщики фирмы А», «количество поставщиков фирмы А»); (з) вопросы с ответом «Да» или «Нет»; (и) вопросы с вопросительными словами.

Между тем логика предикатов первого порядка (ЛППП) не позволяет строить формальные аналоги (на семантическом уровне) таких текстов из бизнес-документов, для которых выполняются перечисленные выше свойства (а)–(и).

Сознавая ограниченность выразительных возможностей ЛППП, Кимбру и Мур в публикациях, посвященных разработке формальных языков общего назначения для бизнес-коммуникаций, используют в примерах несколько конструкций, не являющихся разрешенными в ЛППП.

Пример 1. В работе (Kimbrough, 2001b) анализируются возможности формализации смысловой структуры заказов на поставку товаров или продукции. В связи с обсуждением темы формального представления коммерческого контракта эта работа особенно интересна, поскольку такой заказ является мини-контрактом: в заказе присутствует *запрос* на поставку определенных товаров, *обещание* оплатить товары в случае их отправки и *утверждения*, описывающие некоторые факты.

В указанной работе для описания партии, состоящей из двух товаров $g1$ и $g2$, вводится формальная конструкция $[g1, g2]$.

Пример 2. В работе (Kimbrough, 2001a) рассматривается пример построения выражения формального языка бизнес-коммуникаций, являющегося семантическим представлением текста $T1 = \text{«UBS requests that Citibank credit the account of Bruce Weber in the amount of 2500 US dollars at today's rate of exchange»}$ («UBS просит Сити-банк перевести на счет Брюса Вебера сумму 2500 долларов США по текущему обменному курсу»). Тексту $T1$ ставится в соответствие следующее формальное выражение:

$$\begin{aligned} &Request(ubs:1) \wedge Speaker(ubs:1, UBS) \wedge Addressee(ubs:1, CITI) \wedge \\ &\wedge Theme(ubs:1, ubs:2) \wedge Reason(ubs:1, 'HONORARIM200608') \wedge \\ &\wedge Cul(ubs:1, 20060901) \wedge Content(ubs:1, \vdash credit(ubs:2) \wedge \\ &\wedge Agent(ubs:2, CITI) \wedge Benefactive(ubs:2, '12-345-67') \wedge \\ &\wedge Theme(ubs:2, ubs:6) \wedge currency(ubs:6) \wedge unit(ubs:6, 'USD') \wedge \\ &\wedge quantity(ubs:6, 2500) \wedge name('12-345-67', 'Bruce Weber') \wedge \\ &\wedge address('12-345-67', ['WASHINGTON SQUARE', \\ &\quad 'NEW YORK, NY 100001']) \neg). \end{aligned}$$

Рассмотрим основные идеи, лежащие в основе построения этого выражения. Цепочки $ubs:1$, $ubs:2$ являются переменными, обозначающими две ситуации: (а) ситуацию обращения к Сити-банку с распоряжением и (б) ситуацию зачисления на счет Брюса Вебера определенной суммы. Цепочка *Request* интерпретируется как имя одноместного предиката; выражение *Request* ($ubs:1$) связывает с ситуацией тип «Распоряжение». Цепочки *Speaker*, *Addressee*, *Theme*, *Reason*, *Cul*, *Content* интерпретируются как имена бинарных предикатов, соответствующих

тематическим ролям (или концептуальным падежам) *Агент действия, Адресат действия, Тема, Причина, Время, Содержание поручения*.

Для представления пропозиционального содержания текста T1, т.е. содержания поручения, использованы уголовые скобки $\ulcorner$ и $\urcorner$. Этот прием выводит построенное формальное выражение за рамки языка логики предикатов первого порядка, поскольку в ЛППП не допускается использование формулы (т.е. представления высказывания) в качестве одного из аргументов предиката (в данном случае предиката *Content*).

Таким образом, проблема разработки формальных языков, позволяющих отображать содержание протоколов коммерческих переговоров, проводимых КИА, и формировать контракты, заключаемые в ходе таких переговоров, очень сложна. В этой связи представляется разумным использовать для решения этой проблемы наиболее широко применимые теории представления структурированных значений ЕЯ-текстов, предоставляемые математической лингвистикой и математической информатикой.

6.5. Анализ возможностей использования СК-языков для формирования контрактов и протоколов переговоров в области электронной коммерции

СК-языки обладают рядом выразительных возможностей, необходимых для формального представления содержания контрактов. Для иллюстрации важной части таких возможностей рассмотрим сценарий взаимодействия деловых партнеров в ходе обработки страховой компанией поступившего заявления о повреждении автомобиля. Этот сценарий опубликован в работе (Xu, Jeusfeld, 2003), посвященной электронным контрактам. Деловыми партнерами являются страховая компания AGFIL, фирмы с названиями Europ Assist, Lee Consulting Services (сокращенно Lee C.S.), а также сервисные центры и технические эксперты. Фирма Europ Assist предоставляет владельцам страховых полисов возможность круглосуточного (24 часа) экстренного обращения по телефону. Фирма Lee C.S. ежедневно координирует и управляет операциями срочного обслуживания по соглашению с компанией AGFIL.

В целом процесс страхового обслуживания осуществляется следующим образом. Владелец страхового полиса звонит в Europ Assist, чтобы заявить о новом страховом случае. Фирма Europ Assist регистрирует поступившую заявку, предлагает подходящий сервисный центр и направляет извещение компании AGFIL, которая проверяет действи-

тельность полиса и то, покрывает ли полис заявку. После того, как компания AGFIL получает заявку, эта компания направляет детали заявки фирме Lee C.S.

Компания AGFIL посылает владельцу полиса письмо, содержащее полную форму заявления. Фирма Lee C.S. соглашается с затратами на ремонт, если эксперт не требуется в связи с тем, что размер ущерба мал; в противном же случае назначается эксперт. Эксперт осматривает поврежденный автомобиль и договаривается с сервисным центром о затратах на ремонт. После получения от фирмы Lee C.S. договора о ремонте автомобиля сервисный центр начинает ремонт. После завершения ремонта сервисный центр посылает счет фирме Lee C.S., которая сравнивает счет с первоначальной оценкой. Фирма Lee C.S. возвращает все счета компании AGFIL. Эта компания осуществляет выплаты. Если заявка будет признана недействительной, то об этом будут проинформированы все стороны, участвующие в процессе, и процесс будет остановлен.

Данный сценарий позволяет проиллюстрировать ряд свойств СК-языков, делающих их удобным инструментом формального описания контрактов.

Свойство 1. Возможность построения составных обозначений целей.

Пример. Пусть $T1 = \langle \text{«Владелец полиса звонит в фирму Europ Assist, чтобы сообщить о повреждении автомобиля»} \rangle$. Тогда $T1$ может иметь К-представление (КП)

Ситуация ($e1$, *телеф-разговор**(*Агент1*, *нек чел**(*Владеет1*,
нек полис1))(*Объект2*, *нек фирма**(*Назв*, *Europ Assist*)
(*Цель*, *сообщение1**(*Тема1*, *нек повреждение1**(*Объект1*,
нек автомобиль))))).

Свойство 2. Наличие средств компактного представления временных и причинно-следственных отношений между ситуациями.

Свойство 3. Возможность построения компактных семантических образов таких фрагментов предложений, которые получены в результате соединения логическими связками «И», «ИЛИ» обозначений предметов, событий, понятий или целей.

Пример. Пусть $T2 = \langle \text{«После получения накладной по ремонту от фирмы Lee C.S. и заявления от владельца страхового полиса компания AGFIL оплатит сервисному центру стоимость ремонта»} \rangle$. Тогда КП текста $T2$ может являться выражением

(Ситуация (e1, (получение1*(Агент2, нек фирма*(Назв, AGFIL) : x1)
 (Объект1, нек накладная*(Тема, нек ремонт : e2) : x2)
 (Отправитель, нек фирма*(Назв, Lee C.S.) : x3) $\wedge$ получение1*
 *(Агент2, x1)(Объект1, нек заявление1 : x4) (Отправитель,
 нек чел*(Владеет1, нек полис1 : x5) : x6))) $\wedge$ Ситуация (e2, оплата1*
 *(Агент2, x1)(Адресат1, нек сервис-центр : x7)(Сумма,
 Стоимость (e2))) $\wedge$ Раньше (e1, e2)).

Свойство 4. Существование средств формального представления содержания дискурсов со ссылками на смысл фраз и более крупных фрагментов текста.

Пример. Пусть T3 = «Фирма Europ Assist предоставляет телефонное обслуживание владельцу страхового полиса; в частности, указывает сервисный центр для ремонта и извещает компанию AGFIL о заявлении владельца страхового полиса». Тогда T3 может иметь К-представление

(Ситуация (e1, обслуживание1*(Агент2, нек фирма*(Назв,
 Europ Assist) : x1)(Инструмент, нек телефон : x2)(Объект1,
 произвольн чел*(Владеет1, нек полис1 : x3) : x4)) : P1 $\wedge$
 $\wedge$ Конкретизация (P1, (Ситуация (e2, указание1* (Агент2, x1) (Адресат,
 x4))(Объект3, нек сервис-центр*(Назначение1, ремонт) : x5)) $\wedge$
 $\wedge$ Ситуация (e3, извещение1* (Агент2, x1)(Адресат1, нек фирма*
 (Назв, AGFIL) : x6)(Содержание1, нек заявление1
 *(Автор, x4) : x7)))).

Свойство 5. Возможность формального представления содержания контрактных обязательств, зависящих от условий.

Пример. Пусть T4 = «Фирма Lee C.S. назначает эксперта для осмотра автомобиля в течение 41 часа с момента получения заявления о повреждении автомобиля, если стоимость ремонта не превышает 500 USD». Тогда КП текста T4 может являться выражением

Если-то ($\neg$ Больше1 (Стоимость (нек ремонт*(Объект1,
 нек автомобиль : x1) : e1), 500/USD), (Ситуация (e2,
 назначение1* (Агент2, нек фирма*(Назв, Lee C.S.) : x2)(Персона1,
 нек эксперт : x3)(Цель1, нек осмотр*(Объект1, x1) : e3)
 (Момент, t1)) $\wedge$ $\neg$ Больше1 (Разность (t1, t0), 41/ час) $\wedge$

$\wedge \text{Ситуация}(e4, \text{получение1}*(\text{Агент2}, x2)(\text{Объект1}, \text{нек заявление1} * \\ *(\text{Тема}, \text{нек повреждение1}*(\text{Объект1}, x1) : e5))(\text{Время}, t0))))).$

Выходя за рамки обсуждавшегося в данном разделе сценария взаимодействия деловых партнеров в процессе обслуживания заявления о страховом случае, сформулируем два дополнительных свойства СК-языков.

Свойство 6. Наличие средств построения составных обозначений множеств как компонентов семантических представлений ЕЯ-текстов, являющихся протоколами переговоров или контрактами (см. в предыдущем разделе обозначение запланированной серии из 5-ти поставок, каждая из которых включает 60 чайных сервизов № 53 и 36 столовых сервизов № 65).

Свойство 7. Возможность построения объектно-ориентированных СП протоколов переговоров и контрактов, т.е. формальных выражений вида

$$\text{нек информ-объект}*(\text{Вид}, \text{concept})(\text{Содержание1}, \\ \text{content})(r_1, u_1) \dots (r_n, u_n),$$

где *concept* – обозначение понятия «протокол переговоров» или «контракт», *content* – К-представление содержания документа, $r_1, \dots, r_n$ – обозначения внешних характеристик документа (задающих метаданные, например, характеристики *Авторы*, *Дата*, *Язык*), $u_1, \dots, u_n$ – цепочки, интерпретируемые как значения характеристик документа.

Дополнительные полезные свойства СК-языков с точки зрения построения СП контрактов и протоколов переговоров определяются возможностями явного указания тематических ролей (концептуальных падежей) в структуре семантического представления ЕЯ-текста, отображения содержания фраз с прямой и косвенной речью, со словами «понятие», «термин», рассмотрения функций, значениями и/или аргументами которых могут быть множества объектов (Поставщики, Ассортимент, Директор, Персонал и т.д.).

Автором данной работы был проведен сравнительный анализ выразительных возможностей СК-языков и явлений ЕЯ, отражающихся в структуре деловых контрактов и протоколов коммерческих переговоров (Fomichov, 1999a, 2002b). Ряд результатов этого анализа был изложен выше. Проведенный анализ позволяет высказать предположение о том, что выразительных возможностей СК-языков достаточно для построения с их помощью формальных представлений контрактов и протоколов коммерческих переговоров.

С другой стороны, выразительные возможности других известных подходов к формальному представлению содержания ЕЯ-текстов недостаточны для построения СП произвольных контрактов и протоколов переговоров. В частности, это относится к теории представления дискурсов, теории концептуальных графов, эпизодической логике, компьютерной семантике русского языка.

Таким образом, аппарат СК-языков открывает новые возможности построения формальных представлений контрактов и протоколов коммерческих переговоров, осуществляемых компьютерными интеллектуальными агентами.

В то же время из проведенного анализа следует, что СК-языки предоставляют уникальный спектр возможностей для представления результатов семантико-синтаксической обработки лингвистическими процессорами дискурсов, являющихся контрактами или протоколами коммерческих переговоров.

6.6. Краткие сведения о бесконтекстных грамматиках

6.6.1. Определения и примеры

Понятие бесконтекстной (или контекстно-свободной) грамматики, предложенное в конце 1950-х гг., широко применяется в теоретическом программировании для описания синтаксиса алгоритмических языков программирования, а также в лингвистике, теории распознавания образов и теории информационных языков.

Определение. Бесконтекстной (контекстно-свободной) грамматикой называется произвольная упорядоченная четверка Γ вида

$$\Gamma = (N, T, s, P),$$

где N и T – конечные непересекающиеся множества символов, S – произвольный символ из множества N , P – конечное множество выражений вида $x \rightarrow y$, где x принадлежит N , y – произвольная цепочка (возможно, пустая) в алфавите $(N \cup T)$.

Символы из множеств N и T называются нетерминальными символами (или нетерминалами) и терминальными символами (или терминалами), символ S называется начальным символом. Выражения из множества P называются продукциями.

Пример. $N_1 = \{s\}$, $T_1 = \{c, d\}$, P_1 – множество продукций

$$s \rightarrow csd, \tag{1}$$

$$s \rightarrow e. \tag{2}$$

Тогда упорядоченный набор $\Gamma_1 = (N_1, T_1, s, P_1)$ является бесконтекстной грамматикой.

Продукции рассматриваются как подстановки – левая часть может заменяться на правую.

Задача 1. Какое множество цепочек можно вывести из P_1 в Γ_1 , если начинать с символа s ?

Решение (цифры над стрелками обозначают номер применяемой продукции): $s \rightarrow^1 csd \rightarrow^1 ccsdd \rightarrow^1 cccsddd \rightarrow^2 cccddd = c^3d^3$;
 $s \rightarrow^2 e = ee = c^0d^0$.

Следовательно, в Γ_1 с помощью подстановок из системы P_1 можно из символа s вывести любую цепочку, входящую в язык $L_1 = \{c^n d^n \mid n \geq 0\}$. С другой стороны, в грамматике Γ_1 нельзя вывести цепочку, не входящую в L_1 . Следовательно, L_1 – формальный язык, порожденный грамматикой Γ_1 .

Пусть Γ – бесконтекстная грамматика вида (N, T, s, P) . Тогда выводом цепочки y , принадлежащей множеству $(N \cup T)^*$, в грамматике Γ называется последовательность цепочек $x_1, \dots, x_n$ из $(N \cup T)^*$, где $n > 1$, $x_1 = s$ (начальный символ), и для $i = 2, \dots, n$ цепочка x_i получается из цепочки x_{i-1} заменой какого-либо вхождения в x_{i-1} левой части продукции из P на правую часть этой же продукции.

Формальным языком, порождаемым бесконтекстной грамматикой Γ , называется множество всех цепочек в T (в это множество может входить e), выводимых в Γ . Этот язык обозначается через $L(\Gamma)$.

Задача 2. Пусть $L_2 = \{c^k d^{2m} \mid k, m \geq 1\}$. Найти грамматику Γ_2 , порождающую язык L_2 .

Решение. Заметим, что каждая цепочка из L_2 состоит из двух независимых частей. Тогда пусть P_2 – это множество, состоящее из продукций

$$s \rightarrow AB, \tag{1}$$

$$A \rightarrow cA, \tag{2}$$

$$A \rightarrow c, \tag{3}$$

$$B \rightarrow ddB, \tag{4}$$

$$B \rightarrow dd, \tag{5}$$

$N_2 = \{s, A, B\}$, $T_2 = T_1 = \{c, d\}$. Тогда легко видеть, что $\Gamma_2 = (N_2, T_2, s, P_2)$ – бесконтекстная грамматика, порождающая L_2 .

Примеры вывода цепочек в Γ_2 :

если $x_1 = cdd$, то

$$s \rightarrow^1 AB \rightarrow^3 cB \rightarrow^5 cdd;$$

если $x_2 = c^3 d^4$, то

$$s \rightarrow^1 AB \rightarrow^4 AddB \rightarrow^5 Addddd \rightarrow^2 cAd^4 \rightarrow^2 ccAd^4 \rightarrow^3 ccccd^4 = c^3 d^4.$$

Задача 3. Пусть L_3 – множество правильных скобочных структур, образованных символами $(,), [,]$. Например, L_3 включает цепочки $()$, $[]$, $(([])([])[[]])$. Требуется построить грамматику Γ_3 , порождающую язык L_3 .

Решение.

Пусть P_3 – множество продукций

$$s \rightarrow (), \quad (1)$$

$$s \rightarrow [], \quad (2)$$

$$s \rightarrow (s), \quad (3)$$

$$s \rightarrow [s], \quad (4)$$

$$s \rightarrow ss, \quad (5)$$

$$N_3 = \{s\}, T_3 = \{ (,), [,] \}, \Gamma_3 = (N_3, T_3, s, P_3).$$

Тогда легко проверить, что Γ_3 – бесконтекстная грамматика, порождающая язык L_3 .

Задача 4. Условимся считать, что идентификатор – это произвольная цепочка, начинающаяся с буквы a, b, c, d и включающая любое количество символов – букв a, b, c, d или цифр $0, 1, \dots, 9$. Например, цепочки $b, d, b1, d3, c3, bd4, ccbdd$ – идентификаторы.

Построить грамматику Γ_4 , порождающую множество всех идентификаторов.

Решение.

Пусть P_4 – множество продукций

$$\langle \text{идентификатор} \rangle \rightarrow \langle \text{голова} \rangle \langle \text{хвост} \rangle$$

$$\langle \text{голова} \rangle \rightarrow a$$

$$\langle \text{цифра} \rangle \rightarrow 0$$

$$\langle \text{голова} \rangle \rightarrow b$$

$$\langle \text{цифра} \rangle \rightarrow 1$$

$$\langle \text{голова} \rangle \rightarrow c$$

$$\langle \text{цифра} \rangle \rightarrow 2$$

$$\langle \text{голова} \rangle \rightarrow d$$

$$\langle \text{цифра} \rangle \rightarrow 3$$

$$\langle \text{хвост} \rangle \rightarrow \langle \text{символ} \rangle$$

$$\langle \text{цифра} \rangle \rightarrow 4$$

$$\langle \text{хвост} \rangle \rightarrow \langle \text{хвост} \rangle \langle \text{символ} \rangle$$

$$\langle \text{цифра} \rangle \rightarrow 5$$

$$\langle \text{хвост} \rangle \rightarrow e$$

$$\langle \text{цифра} \rangle \rightarrow 6$$

$$\langle \text{символ} \rangle \rightarrow \langle \text{голова} \rangle$$

$$\langle \text{цифра} \rangle \rightarrow 7$$

$$\langle \text{символ} \rangle \rightarrow \langle \text{цифра} \rangle$$

$$\langle \text{цифра} \rangle \rightarrow 8$$

$$\langle \text{цифра} \rangle \rightarrow 9,$$

$$N_4 = \{ \langle \text{идентификатор} \rangle, \langle \text{голова} \rangle, \langle \text{хвост} \rangle, \langle \text{символ} \rangle, \langle \text{цифра} \rangle \},$$

$$T_4 = \{ a, b, c, d, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 \}.$$

Тогда упорядоченный набор

$$\Gamma_4 = (N_4, T_4, \langle \text{идентификатор} \rangle, P_4)$$

является грамматикой, порождающей множество всех идентификаторов.

6.6.2. Бесконтекстные грамматики в форме Бэкуса-Наура

Бесконтекстные грамматики в форме Бэкуса-Наура используются в теоретическом программировании для описания синтаксиса алгоритмических языков программирования.

Для преобразования бесконтекстной грамматики в форму Бэкуса-Наура необходимо (Johnsonbaugh, 2001):

(1) символ $\rightarrow$ заменить на $::=$,

(2) подмножество продукций вида

$$x \rightarrow y_1$$

$$x \rightarrow y_2$$

$$x \rightarrow y_3$$

заменить на выражение вида $x ::= y_1 | y_2 | y_3$, и т.д.,

(3) цепочку ϵ не указывать в продукциях, вместо нее ничего не пишется.

Учитывая все эти правила, можно грамматику Γ_4 заменить на грамматику Γ_5 в форме Бэкуса-Наура.

Тогда продукции будут выглядеть следующим образом:

$$\langle \text{идентификатор} \rangle ::= \langle \text{голова} \rangle \langle \text{хвост} \rangle$$

$$\langle \text{голова} \rangle ::= a | b | c | d$$

$$\langle \text{хвост} \rangle ::= | \langle \text{символ} \rangle | \langle \text{хвост} \rangle \langle \text{символ} \rangle$$

$$\langle \text{символ} \rangle ::= \langle \text{голова} \rangle | \langle \text{цифра} \rangle$$

$$\langle \text{цифра} \rangle ::= 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9$$

В 1960–1980-е гг. бесконтекстные грамматики широко использовались в форме Бэкуса-Наура для описания синтаксиса алгоритмических языков программирования, например, синтаксиса языков АЛГОЛ-60, Фортран, Кобол, Паскаль.

В 1990-е гг. бесконтекстные грамматики нашли новые важные применения: для описания синтаксиса языка Java, в проекте Семантической Всемирной Паутины (Semantic Web) для описания синтаксиса семантически-структурированных информационных языков, в частности, синтаксиса языка описания метаданных об информационных ресурсах RDF = Resource Description Framework (RDF, 1999; RDF SSL, 2000) и для описания синтаксиса стандарта языка общения компьютерных ин-

теллектуальных агентов (КИА) и стандарта языка представления содержания посланий КИА, разработанных в рамках международного Фонда интеллектуальных физических агентов (FIPA, 1998a, 1998b; FIPA SL CL, 2002).

6.7. Синтаксис языка представления содержания посланий компьютерных интеллектуальных агентов FIPA SL Content Language

Рассмотрим основную часть бесконтекстной грамматики в модифицированной форме Бэкуса-Наура, задающей синтаксическую структуру выражений языка представления содержания посланий КИА FIPA SL Content Language (FIPA SL CL, 2002), разработанного в рамках международного Фонда интеллектуальных физических агентов (FIPA). В продукциях этой грамматики используется знак '=' вместо знака '::=' (читаемого «определяется как»).

Каждое выражение вида D^+ в правой части продукции, где D – нетерминал, означает возможность использования любой конечной последовательности вхождений нетерминала D в этой части продукции. Каждое выражение вида Z^* в правой части продукции, где Z – нетерминал или терминал, означает возможность использования любой конечной или пустой последовательности вхождений символа Z в этой части продукции. Терминальные символы (или терминалы) в правых частях продукций представляются цепочками, обрамленными кавычками слева и справа; например, левая и правая полукруглые скобки являются терминалами, поэтому в правых частях продукций грамматики используются цепочки "(" и ")". В конце каждой продукции ставится точка.

Отличие рассматриваемой основной части бесконтекстной грамматики от полной грамматики из стандарта (FIPA SL CL, 2002) заключается, во-первых, в переводе на русский язык обозначений нетерминалов; во-вторых, в пропуске продукций, позволяющих вывести из нетерминалов *Строка*, *Целое*, *ИдентификаторПеременной* и некоторых других цепочки в терминальном алфавите; в-третьих, в конце каждой продукции ставится не точка, а точка с запятой, что более привычно.

Фрагмент грамматики языка FIPA SL Content Language

Содержание "(" ОбозначениеСодержания+ ")";

ОбозначениеСодержания = ИдентифицирующееВыражение
| ОбозначениеДействия
| Высказывание;

Высказывание = Правильно-постр-формула;

Правильно-постр-формула =
Атомарная-формула

| "(" УнарныйЛогическийОператор Правильно-постр-формула ")"
| "(" БинарныйЛогическийОператор Правильно-постр-формула
Правильно-постр-формула ")"
| "(" Квантор Переменная Правильно-постр-формула ")"
| "(" МодальныйОператор Агент Правильно-постр-формула ")"
| "(" ОператорДействия ОбозначениеДействия ")"
| "(" ОператорДействия ОбозначениеДействия Правильно-постр-
формула " ");

УнарныйЛогическийОператор = "not";

БинарныйЛогическийОператор = "and" P "or" P "implies" P "equiv";

Атомарная-формула = Пропозициональный-символ

| "(" Бинарный-Оператор-Над-Термами ТермИлиИдентифицир-вы-
ражение ТермИлиИдентифицир-выражение ")"
| "(" ПредикатныйСимвол ТермИлиИдентифицир-выражение+ ")"
| "истина" | "ложь" ;
Бинарный-Оператор-Над-Термами = "=" | "результат" ;
Квантор = "для-любого" | "существует" ;

МодальныйОператор = "B" | "U" | "PG" | "I" ;

ОператорДействия = "возможно" | "выполнено";

ТермИлиИдентифицир-выражение = Терм | Идентифицир-выраже-
ние;

Терм = Переменная P ФункциональныйТерм | ОбозначениеДействия
| Константа | Последовательность | Множество;

Идентифицир-выражение = "(" РеференциальныйОператор
ТермИлиИдентифицир-выражение Правильно-постр-формула " ");

РеференциальныйОператор = "iota" | "any" | "all";

ФункциональныйТерм = "(" ФункциональныйСимвол
ТермИлиИдентифицир-выражение* ")"

Р "(" ФункциональныйСимвол Параметр* ")";

Константа = ЧисловаяКонстанта | Строка | ДатаВремя ;

ЧисловаяКонстанта = Целое | ВещественноеЧислоСПлавающейЗа-
пятой;

Переменная = ИдентификаторПеременной;

ОбозначениеДействия = "(" «действие» Агент ТермИлиИдентифи-
цир-выражение ")"

| "(" "Р" ОбозначениеДействия ОбозначениеДействия ")"

| "(" ";" ОбозначениеДействия ОбозначениеДействия ")";

Пропозициональный-символ = Строка;

ПредикатныйСимвол = Строка;

ФункциональныйСимвол = Строка;

Агент = ТермИлиИдентифицир-выражение;

Последовательность = "(" "последовательность" ТермИлиИдентифи-
цир-выражение* ")";

Множество = "(" "множество" ТермИлиИдентифицир-выражение* ")";

Параметр = НазваниеПараметра ЗначениеПараметра;

ЗначениеПараметра = ТермИлиИдентифицир-выражение.

Рассмотрим интерпретацию модальных операторов *B (Belief)*,
U (Uncertainty), *I (Intention)*, *PG (Persistent Goal)*.

Выражение вида (*B Agent Expression*) означает, что интеллектуаль-
ная система *Agent* верит в то, что выражение *Expression* истинно.

Смысл выражения вида (*U Agent Expression*) заключается в том,
что интеллектуальная система *Agent* не уверена в том, что выражение
Expression истинно. Это значит, что *Agent* не верит в истинность выра-
жения *Expression* и в то же время не верит в ложность выражения
Expression, но при этом больше верит в то, что выражение *Expression*
истинно, чем в то, что выражение *Expression* ложно.

Выражение вида (*I Agent Expression*) означает, что интеллектуаль-
ная система *Agent* намерена предпринять какие-то действия для того,
чтобы выражение *Expression* стало истинным.

Смысл выражения вида (*PG Agent Expression*) заключается в том,
что одна из постоянных целей интеллектуальной системы *Agent* зак-

ключается в том, чтобы сделать истинным выражение *Expression*, но не обязательно стремится сделать это в ближайшее время.

Выражение вида

$action-1; action-2; \dots; action-N$

обозначает последовательность, состоящую из N действий, а выражение вида

$action-1 \mid action-2 \mid \dots \mid action-N$

обозначает какое-то действие, являющееся результатом выбора из действий $action-1, action-2, \dots, action-N$.

6.8. Моделирование выразительных механизмов языка представления содержания посланий FIPA SL Content Language средствами СК-языков

Рассмотрим те свойства СК-языков, которые обеспечивают возможность моделирования выразительных механизмов языка FIPA SL Content Language.

Свойство 1. Возможность моделирования выразительных механизмов языка логики предикатов первого порядка.

Эта возможность обеспечивается правилами построения выражений СК-языков $P[2]$ (построение термов с помощью имен функций и обозначений их аргументов), $P[4]$ (построение атомарных формул с помощью имен отношений и обозначений их атрибутов), $P[6]$ (присоединение к формуле логической связки НЕ), $P[7]$ (построение составных формул из исходных формул с помощью логических связок И, ИЛИ), $P[9]$ (присоединение к формуле логических кванторов существования и всеобщности с указанием области действия квантора).

Возможность построения формул вида (*эквивал* $A \ B$), где A, B – правильно построенные формулы, обеспечивается правилом $P[3]$, позволяющим строить выражения СК-языков вида ($Y \equiv Z$), где Y и Z – формулы, являющиеся соответственно аналогами формул A и B и интерпретируемые как семантические представления высказываний.

Для формул со связкой «импликация» («*implies*») вида (*импликация* $A \ B$) можно строить их аналоги – выражения СК-языков вида *Импликация* (C, D), где *Импликация* – реляционный символ из первичного информационного универсума с типом $\{(P, P)\}$, где P – выделенный сорт «смысл сообщения», C и D – выражения некоторого СК-языка, являющиеся соответственно аналогами формул A и B .

Свойство 2. Возможность моделирования структуры правильно построенных атомарных формул вида ($= \text{Term1 Term2}$).

Эта возможность обеспечивается правилом $P[3]$, позволяющим строить выражения СК-языков вида $(y = z)$, где y и z – формулы, являющиеся соответственно аналогами термов *Терм1* и *Терм2* и интерпретируемые как семантические описания сущностей из рассматриваемых предметных областей, обозначениями которых являются термы *Терм1* и *Терм2* (см. примеры ниже).

Свойство 3. Возможность моделирования структуры правильно построенных формул с йота-оператором (такие формулы используются для описания объектов или множеств объектов, обладающих определенными свойствами).

Пример. Рассмотрим выражение языка FIPA SL Content Language $(iota ? x (France President ? x))$, интерпретируемое как семантическое представление выражения «президент Франции». Тогда можно построить эквивалентное К-представление этого выражения

$нек\ чел*(Президент, нек\ страна1*(Назв, 'Франция')) : x1$.

Для построения этого КП нужно выполнить следующие шаги: с помощью правил $P[0]$, $P[1]$, $P[2]$, $P[3]$ построить цепочку

$(Назв(нек\ страна1) \equiv 'Франция')$;

по правилу $P[2]$ построить цепочку

$Президент(нек\ страна1*(Назв, 'Франция'))$;

с помощью правил $P[1]$, $P[3]$ построить цепочку

$(нек\ чел \equiv Президент(нек\ страна1*(Назв, 'Франция')))$;

с помощью правил $P[9]$, $P[1]$ построить цепочку

$нек\ чел*(Президент, нек\ страна1*(Назв, 'Франция'))$;

применяя правила $P[0]$ и $P[5]$, получить цепочку

$нек\ чел*(Президент, нек\ страна1*(Назв, 'Франция')) : x1$.

Свойство 4. Возможность моделирования структуры правильно построенных формул с кванторами «произвольный» (какой-нибудь, ану) и «все» (all).

Эта возможность обеспечивается правилами построения выражений СК-языков $P[8]$ (построение составных обозначений понятий) и $P[1]$ (присоединение интенциональных кванторов к простым или составным обозначениям понятий).

Пример. Выражениям «какой-нибудь трехзвездочный отель в Вене» и «все трехзвездочные отели Вены» можно поставить в соответствие К-представления

$произвольн\ отель*(Колич-звезд, 3)(Место, нек\ город*(Назв, 'Вена'))$,

$все\ отель*(Колич-звезд, 3)(Место, нек\ город*(Назв, 'Вена'))$.

Свойство 5. Возможность моделирования структуры правильно построенных формул с модальными операторами.

Рассмотренная выше основная часть грамматики, задающей синтаксис языка представления содержания посланий КИА FIPA SL Content Language, использует обозначения модальных операторов B (*Belief*), U (*Uncertainty*), I (*Intention*), PG (*Persistent goal*); интерпретация этих операторов была рассмотрена выше.

Теория СК-языков позволяет ввести в первичный информационный универсум элементы B , U , I , PG , имеющие один и тот же тип $\{(интс, P)\}$, где *интс* – сорт «интеллектуальная система», P – выделенный сорт «смысл сообщения», и рассматривать эти обозначения модальных операторов как бинарные реляционные символы.

Свойство 6. Возможность описания последовательности действий.

Пример. Пусть $T1$ = «Игорь Новиков поручил Андрею Титову подготовить проект договора с фирмой «Рассвет» и направить его факсом генеральному директору этой фирмы до 14:00 ближайшего четверга». Тогда $T1$ может соответствовать следующее К-представление:

(Ситуация($e1$, поручение1*(Время, нек момент*(Раньше, сейчас) : $t1$)
(Агент1, нек чел*(Имя, Игорь)(Фам, Новиков) : $x1$) (Адресат1,
нек чел*(Имя, Андрей)(Фам, Титов) : $x2$)(Цель1, реализация1*
(События, ($e2$ and $e3$)))) $\wedge$ Ситуация($e2$, подготовка1(Время, $t2$)
(Агент1 $x2$)(Продукт1, нек проект*(Целевой-объект1, нек договор1*
(Партнер1, нек фирма(Назв, 'Рассвет') : $x3$) : $x4$))) $\wedge$
 $\wedge$ Ситуация($e3$, пересылка1*(Время, нек момент*
*(Позже, $t2$)(Раньше, Четверг($t1$, 14:00)) : $t3$) (Агент1, $x2$)
(Адресат1, нек чел : $x5$)) $\wedge$ ($x5 \equiv$ Ген-директор($x3$))).

С формальной точки зрения, признаками описания последовательности действий являются совпадение агентов действий и явное указание соотношения между моментами выполнения этих действий.

Свойство 7. Возможность описания альтернативных действий.

Эта возможность обеспечивается правилом построения выражений СК-языков $P[7]$, позволяющим соединять логической связкой ИЛИ обозначения действий.

Свойство 8. Возможность построения аналогов термов языка FIPA SL Content Language, обозначающих результат выполнения действия.

К-представления вида

нек объект*(Описание, Результат (agent, action, v)) : v ,

где *Описание* – имя бинарного отношения, *Результат* – имя трехместного отношения, *agent* – обозначение интеллектуальной системы или природной силы, выполнившей действие, *action* – обозначение действия (см. Свойство 6), *v* – переменная, являющаяся меткой результата действия, могут использоваться для обозначения результатов выполнения действий.

Свойство 9. Возможность построения составных обозначений множеств.

Пример. Пусть E1 – выражение «Три контейнера с велосипедами». Тогда выражению E1 может соответствовать K-представление

нек множ(Колич-элемент, 3)(Качеств-состав, контейнер1 *
(Семантич-характ-содержания-емкости, велосипед)).

Ряд других примеров построения составных обозначений множеств был рассмотрен выше.

6.9. Преимущества СК-языков по сравнению с языком FIPA SL Content Language

Совокупность примеров построения K-представлений ЕЯ-текстов, рассмотренных в главах 3–6, показывает, что аппарат СК-языков обладает следующими основными преимуществами по сравнению с с языком представления содержания посланий компьютерных интеллектуальных агентов FIPA SL Content Language:

(а) позволяет строить объектно-ориентированные концептуальные представления ЕЯ-текстов (т.е. формальные выражения, отражающие не только содержание текста, но и его метаданные – сведения об авторе, цели формирования текста, языке и т.д.), что необходимо для формального представления коммуникативных актов и коммерческих контрактов;

(б) ориентирован на построение семантических представлений (СП) не только отдельных предложений, но и связных текстов (в частности, позволяет отображать ссылки на смысл фраз и более крупных фрагментов дискурса);

(в) позволяет строить формальные аналоги сложных составных обозначений понятий и множеств, а также сложных составных обозначений объектов и ситуаций;

(г) предоставляет возможность отображения смысловой структуры предложений со словом «понятие», что необходимо, в частности, для

формального представления информации энциклопедического характера;

(д) позволяет строить формальные представления составных целей;

(е) конструктивно отражает существование нескольких дополнительных способов использования логических связей в русском, английском и многих других языках по сравнению с языком логики предикатов;

(ж) как следствие перечисленных и некоторых других свойств, удобен для построения семантических представлений коммерческих контрактов и протоколов переговоров.

6.10. Сравнение выразительных возможностей СК-языков с возможностями основных известных подходов к формальному представлению содержания ЕЯ-текстов

6.10.1. Сравнение с основными подходами, разработанными в нашей стране

Основными средствами формального представления содержания ЕЯ-текстов, разработанными в нашей стране и использовавшимися в 1990–2000-е гг. для проектирования лингвистических процессоров, являются, помимо предложенных автором данной монографии стандартных К-языков (СК-языков), расширенные семантические сети (Кузнецов, 1976–1989; Кузнецов и др., 2000; Кузнецов, Мацкевич, 2001, 2003; Кузнецов, Шарнин, 2003; Kuznetsov, Matskevich, 2002; Соловьева, Со-мин, 1993), формальные выражения, предоставляемые компьютерной семантикой русского языка, и неоднородные семантические сети (Осипов, 1990, 1997).

Расширенную семантическую сеть (РСС) можно представить как конечное множество выражений вида $R(c_1, c_2, \dots, c_n, d)$, где $n \geq 1$, R – имя n -арного отношения, $c_1, c_2, \dots, c_n$ – атрибуты отношения R , d – метка, являющаяся уникальным именем (в рамках рассматриваемой базы данных) выражения $R(c_1, c_2, \dots, c_n)$. Выражения вида $R(c_1, c_2, \dots, c_n, d)$ называются элементарными фрагментами (ЭФ).

Каждый ЭФ вида $R(c_1, c_2, \dots, c_n, d)$ можно аппроксимировать выражением некоторого СК-языка $R(c_1, c_2, \dots, c_n) : d$, где при построении формулы $R(c_1, c_2, \dots, c_n)$ на последнем шаге применялось правило $P[4]$, а формула $R(c_1, c_2, \dots, c_n) : d$ построена в результате применения правила $P[5]$ к операндам $R(c_1, c_2, \dots, c_n)$ и d , причем d является переменной.

Использование элементарных фрагментов обеспечивает большую однородность представления информации в виде РСС, что создает

предпосылки для унификации процедур обработки знаний, представленных с помощью РСС. Однако принципиальным недостатком использования РСС для отображения содержания текстов является огромный разрыв между структурой ЕЯ-текстов и структурой их семантических представлений. В связи с этим, располагая только аппаратом РСС, разработчик семантико-синтаксического анализатора ЕЯ-текстов остается один на один с многочисленными проблемами алгоритмизации перехода от ЕЯ-текста к его семантическому представлению.

Исходный запас идей для развития компьютерной семантики русского языка (КСРЯ) был изложен в 5-й главе монографии (Тузов, 1984). В последующие годы эти идеи получили развитие, в частности, в публикациях (Тузов, 2001; Каневский, Тузов, 2002; Лезин, Тузов, 2003).

Центральная идея КСРЯ заключается в следующем. Каждое слово русского языка (РЯ) интерпретируется как название (имя) функции, связанной с этим словом и называемой его семантикой. Семантическое представление предложения является суперпозицией функций, причем в качестве аргументов исходных функций берутся обозначения понятий (называемые лексемами).

Анализ публикаций по КСРЯ позволяет заметить, что более естественно было бы говорить в таких ситуациях о суперпозиции функций и отношений. Например, выражения $Caus(x, y)$ (x является причиной события y), $Loc(x, y)$ (x расположен в y) наиболее естественно рассматривать как атомарные формулы, в которых элементы $Caus$, Loc интерпретируются либо как имена бинарных отношений, либо как имена бинарных предикатов.

Выражения семантического языка КСРЯ можно аппроксимировать выражениями СК-языков, полученными применением правил $P[2]$ и $P[4]$, предназначенных для использования имен функций и имен n -арных отношений ($n \geq 1$), к исходным цепочкам вида $ref\ concept$, где ref – квантор референтности (информационная единица, соответствующая слову «некоторый»), $concept$ – обозначение понятия из первичного информационного универсума.

Например, в работе (Тузов, 2001) семантическое представление предложения «Собака охраняет кофейную плантацию» является выражением

$Oper09_a1(СОБАКА\$14224112\sim! \%1, ОХРАНА\$182036(Rel_o1$
 $(ПЛАНТАЦИЯ\$12411\sim! \%1, КОФЕ\$14/112)))$.

К-представлением этого предложения может быть, например, выражение

$\exists e1(\text{ситуация}) \text{ Является}(e1, \text{охрана1}*(\text{Агент1, нек собака})(\text{Объект1, нек плантация}*(\text{Растения, нек множ}*(\text{Кач-состав, дерево}*(\text{Вид, кофейн}))))))$.

СК-языки удобны и для построения шаблона семантической модели предложения (см. Лезин, Тузов, 2003). Например, шаблон семантической модели предложения «Вручая книгу, старик окинул мальчика быстрым оценивающим взглядом» является последовательностью формальных выражений, включающей выражения

КНИГА \$14110 : X1, СТАРИК \$12411 : X2, ВРУЧЕНИЕ \$15210 : X3
(Oper, СУБЪЕКТ.X2, ОБЪЕКТ.X1, АДРЕСАТ.Z3).

Эти выражения можно аппроксимировать К-цепочками

*Явл(X1, книга), Явл(X2, старик), Явл(e1, вручение1**
**(Агент, X2)(Объект1, X1)(Адресат, Z3)).*

Аппарат СК-языков, предложенный в данной книге, обладает следующими основными преимуществами по сравнению с КСРЯ:

(а) ориентирован на построение СП не только отдельных предложений, но и связных текстов (в частности, позволяет отображать ссылки на смысл фраз и более крупных фрагментов дискурса);

(б) позволяет строить формальные аналоги сложных составных обозначений понятий и множеств;

(в) предоставляет возможность отображения смысловой структуры предложений со словом «понятие», что необходимо для формального представления информации энциклопедического характера;

(г) позволяет строить формальные представления составных целей;

(д) конструктивно отражает существование нескольких дополнительных способов использования логических связей в русском, английском и многих других языках по сравнению с языком логики предикатов.

Этими же основными преимуществами аппарат СК-языков обладает и по сравнению с аппаратом неоднородных семантических сетей (Осипов, 1990, 1997).

Таким образом, аппарат СК-языков значительно расширяет возможности формального отображения содержания ЕЯ-текстов по сравнению с КСРЯ и по сравнению с аппаратом неоднородных семантических сетей.

В последнее десятилетие у зарубежных исследователей, работающих в области компьютерной лингвистики, значительно усилился интерес к методам формального исследования семантики ЕЯ-тестов. В этот период наибольшей популярностью пользовались три подхода: теория представления дискурсов (ТПД), возникшая в начале 1980-х гг. (Kamp, 1981), теория концептуальных графов (ТКГ), своим появлением обязанная работам (Sowa, 1984, 1991), и эпизодическая логика (ЭЛ), предложенная Л. Шубертом и Ч.Х. Хуан (Schubert, 1999, 2000; Shubert, Hwang, 1989, 2000; Hwang, 1992; Hwang, Schubert, 1993a–1993b).

Анализ показывает, что структура любых формальных выражений, использовавшихся в теории концептуальных графов или эпизодической логике для отображения содержания ЕЯ-текстов и представления знаний о мире, может быть аппроксимирована выражениями стандартных К-языков (СК-языков).

Например, выражение $[\text{книга} : \{ * \} @ 50]$, являющееся в ТКГ СП словосочетания «50 книг», может быть аппроксимировано К-цепочкой $\text{нек} \text{множ}^* (\text{Колич-элемент}, 50) (\text{Качеств-состав}, \text{книга})$, где *нек* – информационная единица (квантор референтности), соответствующая словам с лексемой «некоторый», *Колич-элемент* и *Качеств-состав* – бинарные реляционные символы, обозначающие отношения «Количество элементов множества» и «Качественный состав множества».

В то же время теория К-представлений обладает несколькими общими глобальными преимуществами по сравнению с перечисленными выше, а также другими известными подходами к формализации содержания ЕЯ-текстов.

Во-первых, модель, построенная в данной книге, представляет в математической форме *гипотезу* об общих внутренних (или ментальных) механизмах формирования сложных структур концептуального уровня (или семантических структур) из первичных единиц концептуального уровня. Ни ЭЛ, ни ТКГ не предпринимают попытки подобного рода. Во-вторых, построенная выше модель формулирует *гипотезу о полной системе квазилингвистических ментальных операций*, т.е. внутренних операций, позволяющих строить концептуальные структуры, выражающие смысл произвольных реальных предложений и дискурсов на ЕЯ, относящихся к любым областям деятельности человека. Другие же известные подходы к формальному описанию содержания ЕЯ-текстов лишь отмечают расширение выразительных возможностей (как правило, языка логики предикатов первого порядка), не выдвигая

гипотезы о построении модели полной системы квазилингвистических ментальных операций и не обсуждая эту проблему.

В-третьих, форма описания как языка концептуальных графов в ТКГ, так и логических форм в ЭЛ не является строго математической. Стиль описания концептуальных графов в работе (Sowa, 2000) напоминает стиль описания языка программирования, например, языка Паскаль. Это потенциально затрудняет разработку на основе этой теории алгоритмов обработки знаний, совместимых с процедурами вычислительной логики. Набор форм Бэкуса-Наура, используемый в диссертации Ч.Х. Хуан (Hwang, 1992) для описания базового логического синтаксиса, включает выражение *<1-местная-предикатная-константа>:= счастливый | человек | определенный | вероятный | ...*, а также несколько других выражений сходной структуры. Единственный способ избежать употребления многоточий в продукциях заключается в рассмотрении некоторого аналога понятия концептуального базиса, введенного выше.

Дополнительными общими преимуществами аппарата СК-языков по сравнению с ТКГ и ЭЛ являются:

- (1) более четкое структурирование предметных областей на основе определения множества типов,
- (2) рассмотрение отношения совместимости на множестве сортов предметной области, позволяющее связывать со многими сущностями не одну, а несколько «координат» по разным «семантическим осям»,
- (3) наличие средств формального описания смысловой структуры таких дискурсов, в которых есть ссылки на смысл предыдущих фраз или более крупных частей текста,
- (4) возможность моделирования смысловой структуры фраз с прямой и косвенной речью,
- (5) возможность рассмотрения информационной единицы, соответствующей слову «понятие», что расширяет арсенал средств формального представления энциклопедической информации,
- (6) возможности использования логических связок «и», «или» для соединения обозначений объектов или понятий или целей,
- (7) возможность использования в формулах имен функций, аргументами и/или значениями которых могут быть множества объектов или понятий.

Наряду с общими преимуществами по сравнению с ТКГ и ЭЛ, теория СК-языков обладает рядом индивидуальных преимуществ по сравнению с каждым из этих подходов. Основными дополнительными преимуществами по сравнению с ТКГ являются:

(1) возможности построения составных обозначений целей, команд,

(2) значительно большие возможности построения составных обозначений множеств объектов и множеств понятий. К дополнительным преимуществам по сравнению с ЭЛ, в частности, относятся возможности построения составных обозначений понятий и множеств объектов.

Примеры текстов, анализируемых в публикациях по теории представления дискурсов (ТПД), по своей сложности близки к сложности предложения «Если у человека есть автомобиль, то он является владельцем кредитной карты». Поэтому, по всей видимости, причины широкой популярности ТПД являются не столько научными, сколько психологическими: простота восприятия изображений, состоящих из блоков с простыми формулами, импонирует широкому кругу лингвистов, интересующихся семантикой ЕЯ. Что же касается ценности ТПД для приложений, то можно согласиться с мнением профессора Л. Аренберга (Ahrenberg, 1992, с. 7) о том, что «вопреки своему названию, ТПД, в сущности, может быть охарактеризована как формальная семантика для коротких последовательностей предложений, но не как теория дискурсов». В отличие от ТПД, предложенная в данной работе модель применима к построению СП широкого многообразия дискурсов на ЕЯ, целей и действий систем, построению определений понятий. Каждое из перечисленных выше преимуществ построенной модели по сравнению с ТКГ или ЭЛ является преимуществом модели по сравнению с ТПД.

Важным аспектом расширения в ЭЛ выразительных возможностей логики предикатов является введение специального оператора *Ka* для образования *видов действий*. Например, выражение «красить стену» в работах (Hwang, Schubert, 1993a; Schubert, Hwang, 2000a) рассматривается как обозначение вида действия. Поэтому с помощью оператора *Ka* строится выражение *Ka (красить стена)*, интерпретируемое как терм языка логики предикатов. Затем образуется выражение

(*настоящ-время [Джон любит (Ka (красить стена))]*).

Это выражение, с одной стороны, интерпретируется как СП предложения «Джон любит красить стену». С другой стороны, в указанных работах это выражение рассматривается как аналог некоторой формулы языка логики предикатов первого порядка.

Попытаемся ответить на вопрос, насколько математически корректным является выбранный способ расширения выразительных возможностей языка логики предикатов первого порядка с помощью оператора *Ka*.

С одной стороны, неопределенные формы глаголов с зависимыми словами позволяют выражать цели интеллектуальных систем, назначения вещей (например, «подъемные краны нужны для того, чтобы поднимать, перемещать и опускать грузы»), советы, желания, умения и т.д. Кроме того, к этому виду выражений легко привести и императивные предложения (в частности, команды).

С математической точки зрения, оператор Ка в указанных работах рассматривается как функция, одним аргументом которой является предикат «Красить», а другим – терм (информационная единица, соответствующая слову «стена»). Но одно из принципиальных ограничений логики предикатов первого порядка заключается в том, что не рассматриваются функции, аргументом которых могли бы быть предикаты.

Кроме того, выражения, образованные неопределенными формами глаголов с зависимыми словами, могут быть сколь угодно сложными и длинными. Чтобы учесть все возможные случаи, нужно рассматривать счетное множество функций, аргументами которых могут быть как предикаты, так и термы.

Представляется, что такой подход полностью противоречил бы как нашей языковой интуиции, так и принципам математической логики. Именно поэтому в работах по ЭЛ рассматриваются только очень простые выражения подобного вида. Как правило, это неопределенная форма глагола с зависимым существительным.

Между тем в теории СК-языков предлагается такое оригинальное решение этой проблемы, которое полностью соответствует нашей языковой интуиции и не является насилием над принципами какого-либо математического подхода. Это решение заключается в возможности построения сколь угодно сложных выражений, интерпретируемых как семантические представления выражений, образованных инфинитивами вместе с зависимыми словами.

Например, цель «Поступить в Государственный университет – Высшую школу экономики, окончить его, подготовить и защитить кандидатскую диссертацию по электронному бизнесу» можно представить в виде выражения некоторого СК-языка

(поступление1(Учеб-заведение, нек университет**

**(Название, 'ГУ-ВШЭ') : x1) ∧ окончание1*(Учеб-заведение, x1) ∧
∧ подготовка1*(Продукт-деятельности, нек диссертация**

**(Уч-степень, кандидат-наук)(Область1, электронный бизнес) : x2) ∧
∧ защита2*(Инф-объект, x2)).*

Литература

1. Арлазаров В.Л., Журавлев Ю.М., Ларичев О.И., Лохин В.М., Макаров И.М., Рахманкулов В.З., Финн В.К. Теория и методы создания интеллектуальных компьютерных систем // Информационные технологии и вычислительные системы. 1998. № 1. С. 3–13.
2. Вагин В.Н. Дедукция и обобщение в системах принятия решений. М.: Наука, Главная редакция физико-матем. литературы, 1988. 383 с.
3. Валгина Н.С., Розенталь Д.Э., Фомина М.И. Современный русский язык / Под ред. Н.С. Валгиной. Изд. 6-е, перераб. и доп. М.: Лотос, 2003. 528 с.
4. Воробьева Г.Ф., Панюшева М.С., Толстой И.В. Современный русский язык. Синтаксис. М.: Русский язык, 1975. 192 с.
5. Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. СПб.: ПИТЕР, 2000. 382 с.
6. Гаврилова Т.А. Использование онтологий в системах управления знаниями // Труды Международного конгресса «Искусственный интеллект в XXI веке». Россия, Дивногорское, 2001. С. 21–32.
7. Городецкий В.И. Многоагентные системы : основные свойства и модели координации поведения // Информационные технологии и вычислительные системы. 1998. № 1. С. 22–34.
8. Добров Б.В., Лукашевич Н.В., Невзорова О.А., Федунов Б.Е. Методы и средства автоматизированного проектирования прикладной онтологии // Известия АН. Теория и системы управления. 2004. № 2. С. 58–68.
9. Жолковский А.К., Мельчук И.А. К построению действующей модели языка «Смысл – Текст» // Машинный перевод и прикладная лингвистика. 1969. Вып. 11. С. 5–35.
10. Ершов А.П. Машинный фонд русского языка: внешняя постановка // Машинный фонд русского языка: идеи и суждения / Под ред. Ю.Н. Караулова. М: Наука, 1986. С. 7–12.

11. Ершов Ю.Л., Палютин Е.А. Математическая логика. М.: Наука, 1979. 320 с.

12. Каневский Е.А., Тузов В.А. Некоторые вопросы пополнения семантического словаря терминами предметной области // Труды Междунар. семинара Диалог'2002 по компьютерной лингвистике и ее приложениям. Т. 2. Прикладные проблемы.

http://www.dialog-21.ru/archive.asp?y=2002&vol=6078 & parent_menu_id=711.

13. Клыков Ю.И., Горьков Л.Н. Банки данных для принятия решений. М.: Советское радио, 1980. 208 с.

14. Клышинский Э.С. Одна модель построения агента // ИУ'99. 1999. С. 126.

15. Кузнецов И.П. Кибернетические диалоговые системы. М.: Наука, 1976. 299 с.

16. Кузнецов И.П. Механизмы обработки семантической информации. М.: Наука, 1978.

17. Кузнецов И.П. Семантические представления. М.: Наука, 1986. 290 с.

18. Кузнецов И.П. (ред). Система обработки декларативных структур знаний ДЕКЛАР-2. М.: Ин-т проблем информатики АН СССР. Препринт. 1989. 106 с.

19. (Кузнецов и др., 2000). Кузнецов И.П., Кузнецов В.П., Мацкевич А.Г. Система выявления из документов значимой информации на основе лингвистических знаний в форме семантических сетей // Труды Междунар. семинара Диалог'2000 по компьютерной лингвистике и ее приложениям. Т. 2. Прикладные проблемы.

http://www.dialog-21.ru/archive.asp?y=20010&vol=6078 & parent_menu_id=711.

20. Кузнецов И.П., Мацкевич А.Г. Лингвистический процессор для автоматического выявления из текстов значимой информации с ее компоновкой в рамках указанных шаблонов // Труды Междунар. семинара Диалог'2001 по компьютерной лингвистике и ее приложениям. Т. 2. Прикладные проблемы.

http://www.dialog-21.ru/archive.asp?y=20010&vol=6078 & parent_menu_id=711.

21. Кузнецов И.П., Мацкевич А.Г. Особенности организации базы предметных и лингвистических знаний в системе АНАЛИТИК // Компьютерная лингвистика и интеллектуальные технологии. Труды Междунар. конференции Диалог'2003 (Протвино, 11–16 июня 2003 г.). С. 373–378.

22. Кузнецов И.П., Шарнин М.М. Интеллектуальный редактор знаний на основе расширенных семантических сетей // Системы и средства информатики / Отв. ред. И.А. Мизин. Вып. 5. М.: Наука, 1993. С. 14–21.
23. Лезин Г.В., Тузов В.А. О представлении результатов семантико-синтаксического анализа текста концептуальными моделями данных // Компьютерная лингвистика и интеллектуальные технологии. Труды Междун. конференции Диалог'2003 (Протвино, 11–16 июня 2003 г.). С. 383–393.
24. Леонтьева Н.Н. Семантика связного текста и единицы информационного анализа // НТИ. Сер. 2. 1981. № 1. С. 21–29.
25. Лукашевич Н.В. АвиаОнтология: анализ современного состояния ресурса // Компьютерная лингвистика и интеллектуальные технологии. Труды междунар. конференции Диалог'2004 (Верхневолжский, 2–7 июня 2004 г.). С. 424–430.
26. Лукашевич Н.В., Добров Б.В. Тезаурус русского языка для автоматической обработки больших текстовых коллекций // Компьютерная лингвистика и интеллектуальные технологии. Труды Междунар. семинара Диалог'2002 Т. 2. Прикладные проблемы / Под ред. А.С. Нариньяни. М.: Наука. 2002. С. 338–346.
http://www.dialog-21.ru/archive.asp?y=2002&vol=6078&parent_menu_id=711.
27. Люстиг И.В., Фомичев В.А. Принципы формального отображения семантики лексических единиц, предложений и дискурсов в интеллектуальной поисковой системе Medsearch // Компьютерная лингвистика и интеллектуальные технологии. Труды Междунар. конф. Диалог'2004 (Верхневолжский, 2–7 июня 2004 г.). С. 431–435.
28. Мальцев А.И. Алгебраические системы. М.: Наука, 1970. С. 392.
29. Мельчук И.А. Опыт теории лингвистических моделей «Смысл – Текст». М.: Наука, 1974. 314 с.
30. Нариньяни А.С. Кентавр по имени ТЕОН: Тезаурус + Онтология // Труды Междунар. семинара Диалог'2001 по компьютерной лингвистике и ее приложениям. Т. 1. Теоретические проблемы.
http://www.dialog-21.ru/archive.asp?y=2001&vol=6077&parent_menu_id=711.
31. Нариньяни А.С. ТЕОН-2: От тезауруса к онтологии и обратно // Труды Междунар. семинара Диалог'2002 по компьютерной лингвистике и ее приложениям. Т. 1. Теоретические проблемы.
http://www.dialog-21.ru/archive.asp?y=2002&vol=6077&parent_menu_id=711.

32. Невзорова О.А., Федунев Б.Е. Система анализа технических текстов «ЛоТА»: основные концепции и проектные решения // Изв. АН. Теория и системы управления. 2001. № 3.

33. Осипов Г.С. Построение модели предметных областей. Неоднородные семантические сети // Изв. АН СССР. Техн. Кибернетика. 1990. № 5. С. 32–45.

34. Осипов Г.С. Приобретение знаний интеллектуальными системами. М.: Наука, Физматлит, 1997.

35. Плесневич Г.С. Логика моделей «классы – бинарные отношения». I // Изв. РАН. Теория и системы управления. 1997. № 5. С. 17–26.

36. Плесневич Г.С. Логика моделей «классы – бинарные отношения». II // Изв. РАН. Теория и системы управления. 1998. № 5. С. 69–80.

37. Плесневич Г.С. Метод аналитических таблиц для логики событий // Интеллектуальное управление: новые интеллектуальные технологии в задачах управления (ICIT'99) – Труды Междун. конференции. Переславль-Залесский, 6–9 декабря 1999. М.: Наука, Физматлит, 1999. С. 72–76.

38. Плесневич Г.С. Понятийно-ориентированные языки в инженерии знаний // Новости искусственного интеллекта. 2003. № 6. С. 3–9.

39. Попов Э.В. Общение с ЭВМ на естественном языке. М.: Наука, Главная редакция физико-математической литературы, 1982. 360 с.

40. Попов Э.В. Общение с базами данных на ограниченном естественном языке: прошлое, настоящее, будущее // Новости искусственного интеллекта. 2002. № 1. С. 21–26.

41. Поспелов Д.А. Большие системы. Ситуационное управление. М.: Знание, 1975. 62 с.

42. Поспелов Д.А. Логико-лингвистические модели в системах управления. М.: Энергоиздат, 1981. 231 с.

43. Поспелов Д.А. Ситуационное управление: теория и практика. М.: Наука, 1986. 288 с.

44. Поспелов Д.А. Многоагентные системы – настоящие и будущее // Информационные технологии и вычислительные системы. 1998. № 1. С. 14–21.

45. Рыбина Г.В., Берзин В.Ю. Модель диалога интеллектуальных агентов // Труды Междунар. семинара Диалог'2002 по компьютерной лингвистике и ее приложениям. Т. 1. Теоретические проблемы. http://www.dialog-21.ru/archive.asp?y=2002&vol=6077 & parent_menu_id=711.

46. Смирнов А.В., Пашкин М.П., Шилов Н.Г., Леватова Т.В. Онтологии в системах искусственного интеллекта: способы построения и организации (Ч. 1) // Новости искусственного интеллекта. 2002. № 1(49). С. 3–13.

47. Смирнов А.В., Пашкин М.П., Шилов Н.Г., Леватова Т.В. Онтологии в системах искусственного интеллекта: способы построения и организации (Ч. 2) // Новости искусственного интеллекта. 2002. № 2(50). С. 3–9.

48. Соловьева Н.С., Сомин Н.В. О семантическом представлении связанных текстов // Системы и средства информатики / Отв. ред. И.А. Мизин. Вып. 5. М.: Наука, 1993. С. 22–29.

49. Тарасов В.Б. Агенты, многоагентные системы, виртуальные сообщества: стратегическое направление в информатике и искусственном интеллекте // Новости искусственного интеллекта. 1998. № 2. С. 5–64.

50. Тузов В.А. Математическая модель языка. Л.: Изд-во ЛГУ, 1984. 176 с.

51. Тузов В.А. Компьютерная семантика русского языка // Труды Междунар. семинара Диалог'2001 по компьютерной лингвистике и ее приложениям. Т. 2. Прикладные проблемы. РосНИИ Искусственного Интеллекта, 2001. С. 356–363.

http://www.dialog-21.ru/archive.asp?y=2001&vol=6078&parent_menu_id=711).

52. Фомичев В.А.(1981a) К теории логико-алгебраического моделирования механизмов речеобразования смыслового уровня. I. Постановка задачи и идея подхода к ее решению. М., 1981. 85 с. Деп. в ВИНТИ ГКНТ и АН СССР 27.10.81, № 4939 – 81 Деп. – РЖ Математика. 1982. № 2, реферат 2В 1386 ДЕП.

53. Фомичев В.А. (1981б) О разработке и применении теории логико-алгебраического моделирования ряда естественно-языковых механизмов текстообразования смыслового уровня // IX Всесоюзный симпозиум по кибернетике (Тезисы симпозиума. Сухуми. 10–15 ноября 1981 г.). Т. 1. Представление знаний. М.: Науч. Совет по комплексной проблеме «Кибернетика» при Президиуме АН СССР, 1981. С. 186–188.

54. Фомичев В.А. Формальные системы для моделирования человеко-машинного взаимодействия на естественном языке // Международный симпозиум по искусственному интеллекту ИФАК. Т. I. Л.: АН СССР, 1983. С. 223–243. Одновременно этот доклад опубликован на английском языке: Fomitchov V.A. Formal systems for natural language

man-machine interaction modelling // International Symposium on Artificial Intelligence IFAC (Intern. Federation of Automatic Control). V. 1. L.: Ac. Sc. USSR, 1983. P. 223–243.

55. Фомичев В.А. Аппарат К-исчислений и К-языков как инструмент исследования ряда ключевых вопросов теории вычислительных систем 5-го поколения // ИНФОРМАТИКА-87. 2-я Всесоюзная конф. по актуальным проблемам информатики и выч. техники. Тезисы докладов. Ереван, 20–22 окт. 1987. Ереван: Изд-во АН Армянской ССР, 1987. С. 182–183.

56. Фомичев В.А. Представление информации средствами К-исчислений: Учеб. пособие. М.: Московский институт электронного машиностроения (МИЭМ), 1988.

57. Фомичев В.А. О средствах построения математической теории естественно-языкового общения // Матем. обеспечение вычислительных, информационных и управляющих систем / Под ред. В.Н. Маркова. М.: МИЭМ, 1988. С. 21–25.

58. Фомичев В.А. К-исчисления, К-языки и проблематика создания фонда естественно-языковых процессоров // Вторая Всесоюз. конф. по созданию Машинного Фонда русского языка (Материалы конф.). М.: Институт русского языка АН СССР, 1988. С. 99–106.

59. Фомичев В.А. Перспективы проектирования экспертных систем на основе аппарата К-исчислений и К-языков // Вопросы применения экспертных систем / Под ред. В.В. Соломатина и Е.В. Марковой. Минск: НПО «Центрсистем», 1988. С. 65–72.

60. Фомичев В.А. О теоретической базе создания фонда естественно-языковых процессоров. М., 1988. 13 с. Статья депонирована в ВИНТИ ГКНТ и АН СССР 28.12.1988, № 9102-B88. Библиогр. указатель ВИНТИ «Депонированные рукописи». 1989. № 4. б/о 170. Техническая кибернетика. 1989. № 5. 5.81.452 ДЕП.

61. Фомичев В.А. Принципы построения модели «Текст – Знания – Смысл» для реализации новой информационной технологии в нескольких предметных областях. М., 1988. II с. Деп. в ВИНТИ АН СССР 28.12.1988, № 9103-B88.

62. Фомичев В.А. Математический синтаксис концептуального уровня как инструментальное средство реализации новой информационной технологии. М., 1988. 13 с. Деп. в ВИНТИ ГКНТ и АН СССР 28.12.1988, № 9104-B88. Библиогр. указатель ВИНТИ «Депонированные рукописи». 1989. № 4. б/о 162.

63. Фомичев В.А. О возможностях структурного документирования алгоритмов средствами стандартных К-языков. М., 1990. 24 с. Деп. в ВИНТИ АН СССР 19.01.1990, № 416-B90.

64. Фомичев В.А. Перспективы применения аппарата К-исчислений и К-языков к разработке лексикона программирования // Математич. и программное обеспечение вычислит., информ. и управляющих систем / Под ред. В.Н. Афанасьева. М.: Московский институт электронного машиностроения, 1990. С. 109–113.

65. Фомичев В.А. К-языки и разработка новых информационных технологий // Новые информационные технологии в системотехнике / Под ред. Л.С. Болотовой. М.: Радио и Связь, 1990. С. 53–62.

66. Фомичев В.А. К-языки и проектирование переносимых подсистем общения экспертных систем // Экспертные системы на персональных компьютерах: Материалы семинара / Общество «Знание» РСФСР. Московский дом научно-технической пропаганды. М., 1990. С. 33–37.

67. Фомичев В.А. Об актуальности и структуре кибернетических моделей нового вида – математических моделей систем обработки естественного языка и их подсистем. М.: МИЭМ, 1991. 32 с. Деп. в ВИНТИ АН СССР и ГКНТ 18.07.1991, № 3066-B91 (реферат 12 Г117 в РЖ «Математика», 1991, № 12).

68. Фомичев В.А. О математических моделях для проектирования анализаторов дискурсов // Научно-технические средства информатизации, автоматизации и интеллектуализации в народном хозяйстве. Материалы семинара / Под ред. А.Е. Петрова. М.: Центральный Российский Дом Знаний, 1991. С. 62–70.

69. Фомичев В.А. (2001). Теория К-исчислений как универсальная формальная метаграмматика для описания содержания посланий компьютерных интеллектуальных агентов // Материалы VII Международного семинара «Дискретная математика и ее приложения» (29 января – 2 февраля 2001 г.), Ч. II. М.: МГУ им. М.В. Ломоносова, Механико-математический факультет, 2001. С. 203–206.

70. Фомичев В.А. (2002а). Алгоритм преобразования матричного семантико-синтаксического представления текста в К-формулу // Проблемы теоретической кибернетики. Тезисы докладов XIII Международной конференции (Казань, 27–31 мая 2002 г.). Ч. II. М.: МГУ им. М.В. Ломоносова, РАН, ИПМ им. М.В. Келдыша, МО и ПО РФ, Казанский государственный университет, 2002. С. 181.

71. Фомичев В.А. (2002б). Математические основы представления смысла текстов для разработки лингвистических информационных технологий. Ч. I. Модель системы первичных единиц концептуального уровня // Информационные технологии. 2002. № 10 (октябрь). С. 16–25.
72. Фомичев В.А. (2002в). Математические основы представления смысла текстов для разработки лингвистических информационных технологий. Ч. II. Система правил для построения семантических представлений фраз и сложных связных текстов // Информационные технологии. 2002. № 11 (ноябрь). С. 34–45.
73. Фомичев В.А. (2005а). Формализация проектирования лингвистических процессоров. М.: МАКС Пресс, 2005. 368 с.
74. Фомичев В.А. (2005б). Семантико-синтаксические анализаторы вопросо-ответных Интернет-систем нового поколения как инструмент повышения качества высшего образования // Качество. Инновации. Образование. 2005. № 1. С. 67–72.
75. Фомичев В.А. (2005в). Стандартные К-языки как универсальный и гибкий инструмент формирования контрактов и протоколов переговоров в области электронной коммерции // Информационные технологии. 2005. № 3. С. 26–29.
76. Фомичев В.А. (2005г). Понятие текстообразующей системы как компонент нового формального аппарата для проектирования лингвистических процессоров // Информационные технологии. 2005. № 8. С. 22–27.
77. Фомичев В.А. (2005д). Класс формальных языков и алгоритм для построения семантических аннотаций Веб-документов // Вестник МГТУ им. Н.Э. Баумана. Сер. «Приборостроение». 2005. № 3 (60). С. 73–86.
78. Фомичев В.А. (2005е). Новый метод преобразования естественно-языковых текстов в семантические представления // Информационные технологии. 2005. № 10. С. 25–35.
79. Фомичев В.А. (2005ж). Новый класс языков для представления содержания посланий компьютерных интеллектуальных агентов // Качество и ИПИ (CALS)-технологии. 2005. № 1. С. 34–39.
80. Фомичев В.А. (2005з). Формализация структуры основных словарей лингвистической базы данных // Качество и ИПИ (CALS)-технологии. 2005. № 3. С. 30–38.
81. Фомичев В.А. (2007). Новый теоретический подход к разработке лингвистических процессоров рекомендательных систем // Новые

информационные технологии в автоматизированных системах. Материалы десятого научно-практического семинара. М.: МИЭМ, МГТУ им. Н.Э. Баумана, Институт прикладной математики им. М.В. Келдыша РАН, 2007. С. 143–145.

82. Abrahams A.S., Evers D.M., Kimbrough S.O. On defining and capturing events in business communication systems // IDT brownbag, 4/22/06. Part 1 in PDF; available online at <http://opim-sky.wharton.upenn.edu/~sok/sokpapers/2007/flbc-implementation-foils-20060422.pdf>.

83. Aczel P., Israel D., Katagiri Y., Peters S. Situation Theory and Its Applications // CSLI Lecture Notes. Stanford: CSLI Publications. 1993. Vol. 3. No. 37.

84. Ahrenberg L. On the integration and scope of segment-based models of discourse // In papers from the Third Nordic Conf. on Text Comprehension in Man and Machine, Linköping Univ. 1992. P. 1–16.

85. Allen J.F., Schubert L. K., Ferguson G.M., Heeman P.A., Hwang C.H., Kato T., Light M., Martin N.G., Miller B.W., Poesio M., Traum D.R. The TRAINS project: A case study in building a conversational planning agent // J. Experimental and Theoretical Artificial Intelligence. 1995. V. 7. P. 7–48.

86. Alshawi H. & van Eijck J. Logical Forms in the Core Language Engine // In Proc. 27th Ann. Meeting of the ACL, Vancouver, Canada. 1989. P. 25–32.

87. Alshawi H. Resolving quasi logical forms // Computational Linguistics. 1990. V. 16. No. 3. P. 133–144.

88. Alshawi H. The Core Language Engine. MIT Press, Cambridge, MA, 1992.

89. Barwise J., Cooper R. Generalized Quantifiers and Natural Language // Linguistics & Philosophy. 1981. Vol. 4. No. 2. P. 159–219.

90. Barwise J., Cooper R. Extended Kamp notation: a graphical notation for Situation Theory // In Aczel, Israel, Katagiri, & Peters (1993). 1993. P. 29–53.

91. Barwise J., Perry J. Situations and Attitudes // Cambridge, MA: The MIT Press, 1983.

92. Brachman R.J., Schmolze J.G. An overview of the KL-ONE knowledge representation system // Cognitive Science. 1985. Vol. 9. No. 2. P. 171–216.

93. Carpenter B. The Logic of Typed Feature Structures. Cambridge Tracts in Theoretical Computer Science, Cambridge University Press, 1992. V. 32.

94. Carpenter B., Penn G. Attribute Logic Engine (ALE), Version 3.2. <http://www.cs.toronto.edu/~gpenn/ale.html>.

95. Cresswell M.J. Structured Meanings: the Semantics of Prepositional Attitudes. Cambridge. MA: MIT Press, 1985.

96. Chierchia G. Structured Meanings, Thematic Roles and Control // Chierchia G., Partee B.H., Turner R. (Eds.). Properties, types and meaning, 2. Semantic issues. Dordrecht etc.: Kluwer. 1989. P. 131–166.

97. Clifford J. QE-III: a Formal Approach to Natural Language Querying // Proceedings of the National Conference on Artificial Intelligence (AAAI-83). Los Altos, California: AAAI; William Kaufman, Inc. 1983. P. 79–83.

98. Clifford J. Natural Language Querying of Historical Data Bases // Computational Linguistics. 1988. V. 14. No. 4. P. 10–34.

99. Cooper R. Three Lectures on Situation Theoretic Grammar // Filgueiras M. et al. (Eds), Natural Language Processing. Berlin etc. Springer-Verlag. 1991. P. 102–140.

100. (CYC 2001) The Upper Cyc Ontology, 2001; <http://www.cyc.com/cyc-2-1/toc.html>.

101. DAML Ontology Library. The DARPA's Technology Integration Center (TIC), Arlington, VA, 2001 (<http://www.daml.org/ontologies>).

102. DAML plus OIL 2001: Reference description of the DAML + OIL (March 2001) ontology markup language (<http://www.daml.org/2001/03/daml+oil>).

103. Eijck D.J.N. van, Kamp H. Representing Discourse in Context. Amsterdam, The University of Amsterdam, 1996.

104. Eklund P.W., Ellis G., Mann G. (Eds.) (1996). Conceptual structures; Knowledge Representation as Interlingua. Proceedings of the 4th Intern. Conf. on Conceptual Structures, ICCS'96, Sydney, Australia, August 1996. Berlin, Heidelberg, New York: Springer-Verlag, 1996. 321 p. (Lecture Notes in Artificial Intelligence, 1115).

105. Emelyanov V.V. Organization of the agents interaction in multi-agents of production coordination system // Binder Z. (Ed.). Management and Control of Production and Logistics 2000 (MCPL 2000). A Proceedings volume from the 2nd IFAC/IFIP/IEEE Conference, Grenoble, France, 5–8 July 2000. Vol. 2. Pergamon, An Imprint of Elsevier Science, Oxford, UK., New York, Tokyo. P. 485–489.

106. Fensel D. et al. OIL in a Nutshell // Knowledge Acquisition, Modeling, and Management, Proceedings of the European Knowledge

Acquisition Conference (EKAW-2000) / Eds. R.Dieng et al. Lecture Notes in Artificial Intelligence (LNAI), Springer-Verlag, 2000.

107. Fenstad J.E., Halvorsen P.K., Langholm T., Van Benthem J. Situations, Language and Logic. Dordrecht: D. Reidel, 1987.

108. Fenstad J.E., Langholm T., Vestre E. Representations and interpretations. In Rosner & Johnson, 1992. P. 31–95.

109. Finin T., Weber J., Wiederhold G., Genesereth M., Fritzson R., McKay D., Mc Guire J., Pelavin R., Shapiro S., Beck C. Draft. Specification of the KQML. Agent-Communication Language. The DARPA Knowledge Sharing Initiative, External Interfaces Working Group, 1993.

110. Finin T., Labrou Y., Mayfield J. (1997). KQML as an agent communication language. In J. Bradshaw (Ed.), Software Agents, MIT Press, Cambridge, 1997.

111. FIPA (1998a). FIPA'98 Specification. Part 2 – Agent Communication Language. Geneva, The Foundation for Intelligent Physical Agents, 1998. <http://www.fipa.org/spec/FIPA98.html>.

112. FIPA (1998b). FIPA'98 Specification. Part 12 – Ontology Service. Geneva, The Foundation for Intelligent Physical Agents, 1998. <http://www.fipa.org/spec/FIPA98.html>.

113. FIPA SL CL (2002). FIPA SL Content Language Specification; <http://www.fipa.org/specs/fipa00008/SC000081.htm>.

114. Fomitchov V. (1984): Formal systems for natural language man-machine interaction modelling // Artificial Intelligence. Proc. of the IFAC Symposium, Leningrad, USSR, 4–6 October 1983 / Edited by V.M. Ponomaryov (IFAC Proc. Series. 1984. No. 9). Oxford, UK; New York, etc.: Pergamon Press, 1984. P. 203–209.

115. Fomitchov V. (1992): Mathematical models of natural-language-processing systems as cybernetic models of a new kind. Cybernetica (Belgium), XXXV (1), 63–91.

116. Fomitchov V.A. (1993a): Towards a mathematical theory of natural-language communication // Informatica. An Intern. J. of Computing and Informatics (Slovenia), 17(1), 21–34.

117. Fomitchov V.A. (1993b): K-calculuses and K-languages as powerful formal means to design intelligent systems processing medical texts // Cybernetica (Belgium), XXXVI (2), 161–182.

118. Fomitchov V.A. (1994): Integral Formal Semantics and the design of legal full-text databases // Cybernetica (Belgium), XXXVII (2), 145–177.

119. Fomitchov V.A. (1995): A variant of a Universal Metagrammar of Conceptual Structures. Algebraic systems of conceptual syntax (Invited talk). In A. Nijholt, G. Scollo, R. Steetskamp (eds.), *Algebraic Methods in Language Processing. Proceedings of the Tenth Twente Workshop on Language Technology joint with First AMAST Workshop on Language Processing*, University of Twente, Enschede, The Netherlands, December 6–8, 1995, 195–210.

120. Fomitchov V.A. (1996a): A mathematical model for describing structured items of conceptual level // *Informatica (Slovenia)*. Vol. 20. No. 1. 5–32.

121. Fomitchov V.A. (1996b): An outline of a formal metagrammar for describing structured meanings of complicated discourses. II Intern. Conf. on Mathematical Linguistics (ICML'96), Abstracts, Tarragona, 2–4 de maig de 1996. Grup de Recerca en Linguística Matemàtica i Enginyeria del Llenguatge (GRLMC), Report 7/96, Universitat Rovira i Virgili, 1996, 31–32.

122. Fomitchov V.A. (1997a): K-calculuses and the problem of conceptual information retrieval in textual data bases. Knowledge Transfer (Volume II). Edited by A.Behrooz (Proc. of the International Conference «Knowledge Transfer – 1997 (KT97)», Symposium «Information Technology», University of London, 14–16 July 1997). London, University of London, 52–58.

123. Fomitchov V.A. (1997b). Formal studies of anticipative systems as a basis of effective humanitarian and computer applications // *Advances in Modelling of Anticipative Systems*, Vol. II / Edited by G.E. Lasker, D. Dubois, B. Teiling. The International Institute for Advanced Studies in Systems Research and Cybernetics. University of Windsor, Canada, 1997. P. 73–77.

124. Fomitchov V.A. (1998a). K-calculuses and matrix grammars as powerful tools for designing natural language processing systems // Abstracts of Papers, Interdisciplinary Symposium «Artificial Intelligence, Cognitive Science, and Philosophy for Social Progress» (President – Chairperson: Mr. Vladimir Fomichov, Russia), 15th International Congress on Cybernetics, Namur, Belgium, August 24–28, 1998. Namur, International Association for Cybernetics, 1998. P. 1.

125. Fomitchov V.A. (1998b). A comprehensive mathematical framework for designing agent communication languages. Proceedings of the International Conference «Information Society (IS'98)», Ljubljana, Slovenia, 6–7 October 1998, 81–84.

126. Fomitchov V.A. (1998c). Theory of Restricted K-calculuses as a Comprehensive Framework for Constructing Agent Communication Languages; Special Issue on NLP and Multi-Agent Systems, edited by V.A. Fomichov and A.P. Zeleznikar, *Informatica // An International Journal of Computing and Informatics* (Slovenia). Vol. 22. No. 4. P. 451–463.

127. Fomitchov V.A. (1999a). Theory of Restricted K-calculuses as a Universal Informational Framework for Electronic Commerce // Database, Web and Cooperative Systems. Vol. 1. The Proceedings of 1999 International Symposium on Database, Web and Cooperative Systems, August 3–4, 1999 in Baden-Baden, Germany – DWACOS'99 / Edited by George E. Lasker, University of Windsor and Yanchun Zhang, University of Southern Queensland. The International Institute for Advanced Studies in Systems Research and Cybernetics, University of Windsor, Windsor, Ontario, Canada. P. 41–46.

128. Fomitchov V.A. (1999b). A Universal Resources and Agents Framework for Electronic Commerce and Other Applications of Multi-Agent Systems; 7th International Workshop on Computer Aided Systems Theory and Technology 1999 – EUROCAST'99. September 29th – October 2nd, 1999, Vienna, Austria, Vienna University of Technology. Preprints (ed. P. Kopacek). P. 99–102.

129. Fomitchov V.A. (2000). An ontological mathematical framework for electronic commerce and semantically-structured Web // Zhang Y., Fomichov V.A., Zeleznikar A.P. (Eds.). Special Issue on Database, Web, and Cooperative Systems. *Informatica. An International Journal of Computing and Informatics* (Slovenia, Europe), 2000. Vol. 24. No. 1. P. 39–49.

130. Fomitchov V.A. (2001). New content languages for electronic commerce and digital libraries // Binder Z. (Ed.). Management and Control of Production and Logistics 2000 (MCPL 2000). A Proceedings volume from the 2nd IFAC/IFIP/IEEE Conference, Grenoble, France, 5–8 July 2000. Vol. 2. Pergamon, An Imprint of Elsevier Science, Oxford, UK., New York, Tokyo. P. 503–508.

131. Fomitchov V.A. (2002a). The Method of Constructing the Linguistic Processor of the Animation System AVIAROBOT // Preconference Proceedings «Collaborative Decision-Support Systems» (Focus Symposium in conjunction with the 14th International Conference on Systems Research, Informatics and Cybernetics – InterSymp-2002, July 29 – August 3, 2002, Germany). Focus Symposia Chair: Jens Pohl – CAD Research Center, Cal Poly, San Luis Obispo, CA, USA, 2002. P. 91–102.

132. Fomitchov V.A. (2002b). Theory of K-calculuses as a Powerful and Flexible Mathematical Framework for Building Ontologies and Designing Natural Language Processing Systems // Troels Andreasen, Amihai Motro, Henning Christiansen, Henrik Legind Larsen (Eds.), Flexible Query Answering Systems. 5th International Conference, FQAS 2002, Copenhagen, Denmark, October 27–29, 2002. Proceedings; LNAI 2522 (Lecture Notes in Artificial Intelligence, Vol. 2522), Springer: Berlin, Heidelberg, New York, Barcelona, Hong Kong, London, Milan, Paris, Tokyo. P. 183–196.

133. Fomitchov V.A. (2004). Theory of Standard K-languages as a Model of a Universal Semantic Networking Language // Preconference Proceedings «Intelligent Software Systems for the New Infostructure» (Focus Symposium in-conjunction with the 16th International Conference on Systems Research, Informatics and Cybernetics – InterSymp-2004, July 29 – August 5, 2004, Germany). Focus Symposia Chair: Jens Pohl – CAD Research Center, Cal Poly, San Luis Obispo, CA, USA, 2004. P. 51–61.

134. Fomitchov V.A. (2005). Standard K-Languages as a Powerful and Flexible Tool for Building Contracts and Representing Contents of Arbitrary E-Negotiations // Bauknecht K., Proell B., Werthner H. (Eds.). The 6th Intern. Conf. on Electronic Commerce and Web Technologies «EC-Web 2005», Copenhagen, Denmark, Aug. 23–26, 2005, Proceedings. Lecture Notes in Computer Science. Vol. 3590. Springer Verlag. 2005. P. 138–147.

135. Fomitchov V.A. (2006). The Theory of SK-Languages as a Powerful Framework for Constructing Formal Languages for Electronic Business Communication // Preconference Proceedings «Advances in Intelligent Software Systems» (Focus Symposium in conjunction with the 18th International Conference on Systems Research, Informatics and Cybernetics – InterSymp-2006, August 7–12, 2006, Germany). Focus Symposia Chair: Jens Pohl – CAD Research Center, Cal Poly, San Luis Obispo, CA, USA, 2006. P. 123–130.

136. Fomitchov V.A. (2007). Theory of K-representations as a Framework for Constructing a Semantic Networking Language of a New Generation // Preconference Proceedings «Representation of Context in Software: Data $\Rightarrow$ Information $\Rightarrow$ Knowledge» (Focus Symposium in conjunction with the 19th International Conference on Systems Research, Informatics and Cybernetics – InterSymp-2007, July 30 – August 4, 2007, Germany). Focus Symposia Chair: Jens Pohl – CAD Research Center, Cal Poly, San Luis Obispo, CA, USA, 2007. P. 61–70.

137. Fomitchov V.A., Akhromov Ya.V. (2003). Standard K-languages as a Powerful and Flexible Tool for Constructing Ontological Intelligent

Agents // Preconference Proceedings «Collaborative Decision-Support Systems» (Focus Symposium in conjunction with the 15th International Conference on Systems Research, Informatics and Cybernetics – InterSymp-2003, July 28 – August 01, 2003, Germany). Focus Symposia Chair: Jens Pohl – CAD Research Center, Cal Poly, San Luis Obispo, CA, USA, 2003. P. 167–176.

138. Fomitchov V.A., Chuykov A.V. (2000). An intelligent agent with linguistic skills for the Web-search of production manufacturers // InterSymp-2000, 12th International Conference on Systems research, Informatics and Cybernetics (July 31 – August 4, 2000, Germany). Preconference Proceedings «Advances in Computer-Based and Web-Based Collaborative Systems (Focus Symposia, August 1 and 2, 2000)», Focus Symposia Chairs Jens Pohl and Thomas Fowler, IV. Collaborative Agent Design (CAD) Research Center, Cal Poly, San Luis Obispo, CA, USA, 2000. P. 137–147.

139. Fomitchov V.A., Kochanov A.A. (2001). Principles of Semantic Search for Information on the Web by the Intelligent Agent LingSearch-1 // Preconference Proceedings «Advances in Computer-Based and Web-Based Collaborative Systems» (Focus Symposia in conjunction with the 13th International Conference on Systems Research, Informatics and Cybernetics – InterSymp-2001, July 31 – August 1, 2001, Germany). Focus Symposia Chairs: Jens Pohl and Thomas Fowler, IV. Collaborative Agent Design (CAD) Research Center, Cal Poly, San Luis Obispo, CA, USA, 2001. P. 121–131.

140. Fomitchov V.A., Lustig I.V. (2001). A Computer Intelligent Agent for Semantic Classification of E-mail Messages // Preconference Proceedings «Advances in Computer-Based and Web-Based Collaborative Systems» (Focus Symposia in conjunction with the 13th International Conference on Systems Research, Informatics and Cybernetics – InterSymp-2001, July 31 – August 1, 2001, Germany). Focus Symposia Chairs: Jens Pohl and Thomas Fowler, IV. Collaborative Agent Design (CAD) Research Center, Cal Poly, San Luis Obispo, CA, USA, 2001. P. 29–37.

141. Fomitchova O.S., Fomitchov V.A. (2004). A New Approach to Designing Children-Oriented Web-sites of Art Museums // In Cybernetics and Systems 2004, Volume II. Proceedings of the Seventeenth European Meeting on Cybernetics and Systems Research (University of Vienna, Austria, 13–16 April 2004) / Edited by Robert Trappl. University of Vienna and Austrian Society for Cybernetic Studies, 2004. P. 757–762.

142. Gaerdenfors P. (Ed.) Generalised Quantifiers // Linguistic and Logical Approaches. Dordrecht: D. Reidel, 1987.

143. Gazdar G., Klein E., Pullum G., Sag I. Generalized Phrase Structure Grammar. Oxford: Blackwell, 1985.
144. Gazdar G., Mellish C.S. Natural Language Processing in PROLOG: an Introduction to Computational Linguistics. Addison-Wesley, 1989.
145. Genesereth M.R., Fikes R.E. et al. (1992). Knowledge Interchange Format Version 3 Reference Manual. Technical Report Logic-92-1, Computer Science Department, Stanford University.
146. Genesereth M.R. (1999). Knowledge Interchange Format. Geneva, FIPA, 1999; <http://www.fipa.org>.
147. Grishman R. Computational Linguistics: an Introduction. Cambridge: Cambridge University Press, 1986.
148. Groenendijk J., Stokhof M. Two Theories of Dynamic Semantics // J. van Eijck (Ed.), Logics in Artificial Intelligence. Berlin etc.: Springer-Verlag. 1991. P. 55–64.
149. Groenendijk J., Stokhof M. Dynamic Montague Grammar // Kalman L., Polos L. (Eds.). Papers from the Second Symposium on Logic and Language. Budapest: Akademiai Kiado, 1990. P. 3–48.
150. Groenendijk J., Stokhof M. Dynamic Predicate Logic // Linguistics and Philosophy. 1991. Vol. 14. No. 1. P. 39–101.
151. Gruber T.R. A translation approach to portable ontology specifications // Knowledge Acquisition Journal. 1993. V. 5. P. 199–220.
152. Guarino N. Formal ontology and information systems // Proceedings of FOIS'98. Trento, Italy. Amsterdam: IOS Press, 1998. P. 3–15.
153. Guilfoyle C., Jeffcoate J., Stark H. (1997). Agents on the Web: Catalyst for E-Commerce. London: Ovum Ltd., April 1997.
154. Hasselbring W., Weigand H. (2001) Languages for electronic business communication: state of the art // Industrial Management and Data Systems. Vol. 1001. Issue 5. P. 217–226.
155. Heim I. E-Type Pronouns and Donkey Anaphora // Linguistics & Philosophy. 1990. Vol. 13. No. 2. P. 137–177.
156. Herzog O., Rollinger C.-R. (Eds.) (1991): Text Understanding in LILOG. Integrating Computational Linguistics and Artificial Intelligence. Final Report on the IBM Germany LILOG-Project. Berlin etc.: Springer-Verlag.
157. Hirst G. Semantic Interpretation and Ambiguity // Artificial Intelligence. 1988. Vol. 34. P. 131–177.

158. Horrocks I. A denotational semantics for Standard OIL and Instance OIL. Department of Computer Science, University of Manchester, UK, 2000; <http://www.ontoknowledge.org/oil/downl/semantics.pdf>.

159. Horrocks I., van Harmelen F., Patel-Schneider P. DAML + OIL Release (March 2001). <http://www.daml.org/2001/03/daml+oil+index>.

160. Hwang C.H. A Logical Approach to Narrative Understanding. Ph.D. Dissertation, U. of Alberta, Edmonton, Canada, 1992.

161. Hwang C.H., Schubert L.K. Meeting the interlocking needs of LF-computation, deindexing, and inference: An organic approach to general NLU // Proc. 13th Int. Joint Conf. on AI (IJCAI'93). Chambéry, France, 1993. 28 Aug. – 3 Sept.

162. (Hwang, Schubert 1993a) Hwang C.H., Schubert L.K. Episodic Logic: a situational logic for natural language processing // In Aczel, Israel, Katagiri & Peters. 1993. P. 303–338.

163. (Hwang, Schubert 1993b) Hwang C.H., Schubert L.K. Episodic Logic: a comprehensive, natural representation for language understanding // Minds and Machines. 1993b. Vol. 3. P. 381–419.

164. Java A., Finin T., Nirenburg S. (2006). Text Understanding Agents and the Semantic Web // Proceedings of the 39th Hawaii International Conference on System Sciences. http://ebiquity.umbc.edu/file_directory/papers/205.pdf.

165. Johnson M. Attribute-Value Logic and the Theory of Grammar. Stanford, CSLI, 1988.

166. Johnsonbaugh R. Discrete Mathematics. Fifth edition. Upper Saddle River, New Jersey: Prentice Hall, 2001. 621 p.

167. Jowsey E. Montague Grammar and First Order Logic // Edinburgh Working Papers in Cognitive Science. 1987. Vol. 1. Univ. of Edinburgh.

168. Kamp H. A theory of truth and semantic representation // Groenendijk J., Janssen T., Stokhof M. (Eds.). Formal Methods in the Study of Natural Language. Part 1. Amsterdam: Mathematical Centre. 1981. P. 227–322.

169. Kamp H., Reyle U. (1993): Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory. Dordrecht: Kluwer Academic Publishers.

170. Kamp H., Reyle U. (1996). A Calculus for First Order Discourse Representation Structures // Journal for Logic, Language and Information (JOLLI), 1996.

171. Kimbrough S.O. (2001a). A Note on Getting Started with FLBC: Towards a User Guide for FLBC. August 25, 2001; <http://grace.wharton.upenn.edu/~sok/sokpapers/2001-2/lap/lap-acm-questions-flbc-20010825.pdf>.
172. Kimbrough S.O. (2001b). A Note on the FLBC Treatment of an Example X12 Purchase Order. 2001; <http://opim-sun.wharton.upenn.edu/~sok/flbcinfo/span-fullEDI-sok-sam.pdf>.
173. Kimbrough S.O. (2001c). Reasoning about the Objects of Attitudes and Operators: Towards a Disquotation Theory for Representation of Propositional Content // Proceedings of the Tenth International Conference on Artificial Intelligence and Law – ICAIL'2001; <http://grace.wharton.upenn.edu/~sok/sokpapers/2000-1/icail/sok-icail01.pdf>.
174. Kimbrough S.O., Moore S.A. (1997). On automated message processing in electronic commerce and work support systems: speech act theory and expressive felicity // ACM Transactions on Information Systems. Vol. 15. No. 4. P. 321–367.
175. Kuznetsov I.P., Matskevich A.G. System for extracting semantic information from natural language text // Труды Междунар. семинара Диалог'2002 по компьютерной лингвистике и ее приложениям. Т. 2. Прикладные проблемы. http://www.dialog-21.ru/archive.asp?y=2002&vol=6078&parent_menu_id=711.
176. Labrou Y. (1996). Semantics for an Agent Communication Language. Ph.D. thesis, University of Maryland, Baltimore County, August 1996.
177. Labrou Y., Finin T. (1997). Semantics and conversation for an agent communication language. Proc. of the 15th International Joint Conference on Artificial Intelligence (IJCAI-97), Nagoya, Japan, August 1997.
178. Labrou Y., Finin T. (1998). Semantics for an Agent Communication Language. In Agent Theories, Architectures, and Languages. Ed. by M.P. Singh, A.S. Rao, M.J. Wooldridge. Lect. Notes in AI. Vol. 1365. Springer-Verlag. 209–214.
179. Lenat D.B. (1995). CYC: A large-scale investment in knowledge infrastructure // Communications of the ACM. 1995. Vol. 38. No. 11 (<http://www.cyc.com>).
180. (Loom 2001). Loom Project Home Page, 2001. <http://www.isi.edu/isd/LOOM/LOOM-HOME.html>.
181. McGuinness D.L., van Harmelen F. (Eds.) (2004). OWL Web Ontology Language Overview. W3C Recommendation 10 February 2004; <http://www.w3.org/TR/owl-features/>.

182. Meyer R. (1994). Probleme von Zwei-Ebenen-Semantiken // Kognitionswissenschaft. 1994. Vol. 4. No. 1. P. 32–46.
183. Montague R. Universal Grammar. Theoria. 1970. Vol. 36. P. 373–398.
184. Montague R. English as a Formal Language // Thomason R.H. (Ed.). Formal Philosophy. Selected papers of Richard Montague. New Haven and London: Yale University Press, 1974. P. 188–221.
185. Montague R. The Proper Treatment of Quantification in Ordinary English. Ibid. P. 247–270.
186. Moore S.A. (1998). Position Statement // Proceedings of the International Workshop on Component-Based Electronic Commerce, The Haas School of Business, July 25, 1998; <http://www-personal.umich.edu/~samooore/research/flbc/statement-for-workshop.html>.
187. Narin'yan A.S. Towards an Integral Model of Language Competence // Computational Models of Natural Language Processing. Amsterdam: North-Holland, 1984. P. 275–295.
188. Nebel B., Peltason C. Terminological reasoning and information management // In Information Systems and Artif. Intel.: Integration Aspects. First Workshop. Ulm, FRG, March 1990, Proc., ed. D. Karagiannis. Berlin etc.: Springer-Verlag, 1991. P. 181–212.
189. Partee B.H., ter Meulen A., Wall R. Mathematical Methods in Linguistics // Dordrecht etc.: Kluwer, 1990.
190. Partee B.H. (Ed.). Montague Grammar. New York etc.: Academia Press, 1976.
191. Peregrin J. On a Logical Formalization of Natural Language // Kybemetika. 1990. Vol. 26. No. 4. P. 327–341.
192. Peres J. Basic Aspects of the Theory of Generalized Quantifiers // Filgueiras M. et al. (Eds.). Natural Language Processing. Berlin etc. Springer-Verlag, 1991. P. 141–157.
193. Plesniewicz G.S., Tarassov V.B. (2001). Networked enterprises as multi-agent systems: concept-oriented logic approach // Binder Z. (Ed.), Management and Control of Production and Logistics 2000 (MCPL 2000). A Proceedings volume from the 2nd IFAC/IFIP/IEEE Conference, Grenoble, France, 5–8 July 2000. Vol. 2. Pergamon, An Imprint of Elsevier Science, Oxford, UK., New York, Tokyo. P. 497–502.
194. Pletat U., Von Luck K. (1990) Knowledge representation in LILOG // Blasius K.H., Hedtstuck U., C.-R. Rollinger (Eds.). Sorts and Types in Artificial Intelligence. Berlin etc.: Springer-Verlag, 1990. P. 140–164.

195. Pletat U. (1991). The knowledge representation language L_{LILOG} . In Herzog and Rollinger (1991). P. 357–379.

196. Pohl J. (2003). The Emerging Knowledge Management Paradigm: Some Organizational and Technical Issues // Preconference Proceedings «Collaborative Decision-Support Systems» (Focus Symposium in conjunction with the 15th International Conference on Systems Research, Informatics and Cybernetics – InterSymp-2003, July 28 – August 01, 2003, Baden-Baden, Germany). Focus Symposia Chair: Jens Pohl – CAD Research Center, Cal Poly, San Luis Obispo, CA, USA, 2003. P. 11–26.

197. Quillian M.R. Semantic memory // Semantic Information Processing, M. Minsky (ed.). Cambridge, Massachusetts, 1968. P. 27–70.

198. (RDF 1999). Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation. January 1999, <http://www.w3.org/TR/WD-rdf-syntax>.

199. (RDF SSL 2000). Resource Description Framework (RDF) Schema Specification 1.0. W3C Candidate Recommendation 27 March 2000; <http://www.w3.org/TR/rdf-schema>.

200. Schank R.C. Conceptual Dependency: a theory of natural language understanding // Cognitive Psychology. 1972. V. 3. N. 4. P. 552–631.

201. Schank R.C., Goldman N.M., Rieger C.J., Riesbeck C.K. Conceptual Information Processing // Amsterdam, Oxford: North-Holland Publ. Company; New York: American Elsevier Publ. Comp., Inc., 1975. Книга переведена на русский язык: Шенк Р., Голдман Н.М., Ригер Ч. Дж., Ризбек К.К. Обработка концептуальной информации / Пер. с англ. под ред. В.М. Брябрина. М.: Энергия, 1980. 361 с.

202. Shapiro S.C. Formalizing English // International Journal of Expert Systems. 1996. V. 9. N. 1. P. 151–171.

203. Schubert L. K., Hwang C. H. An episodic knowledge representation for narrative texts // Proceedings of the First Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'89), Toronto, Canada, 1989. P. 444–458.

204. Schubert L.K., Hwang C.H. Episodic Logic meets Little Red Riding Hood: A comprehensive, natural representation for language understanding // Iwanska L., Shapiro S.C. (eds.). Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language, MIT/AAAI Press, Menlo Park, CA, and Cambridge, MA, 2000. P. 111–174.

205. Schubert L.K. Dynamic Skolemization // Bunt H., Muskens R. (eds.). *Computing Meaning*. Vol. 1. *Studies in Linguistics & Philosophy Series*. Kluwer Academic Press, Dordrecht (also Boston, London), 1999. P. 219–253.
206. Schubert L.K. The situations we talk about // Minker J. (ed.). *Logic-Based Artificial Intelligence*. Kluwer, Dordrecht, 2000. P. 407–439.
207. Semantic Web (2001). Semantic Web Activity Statement. W3C, <http://www.w3.org/2001/sw/activity>.
208. Sembock T.M.T., Van Rijsbergen C.J. SILOL: A Simple Logical-linguistic Document Retrieval System // *Information Processing & Management*. 1990. Vol. 26. No. 1. P. 111–134.
209. Simmons R.F. Semantic networks: their computation and use for understanding English sentences // Schank R., Colby K. (eds.). *Computer Models of Thought and Language*. San Francisco: Freeman, 1973. P. 63–113.
210. Sowa J.F. *Conceptual Structures: Information Processing in Mind and Machine* // Addison-Wesley Publ. Comp.: Reading, MA, 1984.
211. Sowa J.F. Toward the expressive power of natural language // Sowa J.F. (Ed.), *Principles of Semantic Networks. Explorations in the Representation of Knowledge*. Morgan Kaufman Publ., Inc. 1991. P. 157–189.
212. Sowa J.F. (2000). *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Pacific Grove, CA: Brooks/Cole. 594 p.
213. Thayse A., Gribomont P., Louis G., Snyers D., Wodon P., Gochet P., Gregoire E., Sanchiz E., Delsarte P. *Approche logique de l'Intelligence Artificielle. 1. De la logique classique a la programmation logique*. Dunot informatique. Bordas, Paris, 1988.
214. Thomason R.H. A Model Theory for Prepositional Attitudes. *Linguistics & Philosophy*. 1980. Vol. 4. No. 1. P. 47–70.
215. Thome R., Schinzer H. (1998). Market survey of electronic commerce. *Informatica (Slovenia)*. Vol. 22. No. 1. P. 11–19.
216. Uchida H., Zhu M., Della Senta T. *The UNL, A Gift for a Millennium*. A book published by The United Nations University /Institute of Advanced Studies (UNU/IAS), 1999.
217. Uchida H., Zhu M. (2001). *The Universal Networking Language beyond Machine Translation*. UNDL Foundation.
218. UNL (2005). *Universal Networking Language (UNL) Specifications, Version 2005*. www.undl.org/unlsys/unl/unl2005/.

219. UNL (2006). Universal Networking Language (UNL) Specifications, Version 2005, Edition 2006, 30 August 2006, UNL Center of UNDL Foundation. <http://www.unl.org/unlsys/unl/unl2005-e2006/>.

220. Voice (2001). «Voice Browser» Activity – Voice enabling the Web! W3C paper, 2001. <http://www.w3.org/Voice/>.

221. Wilks Y. An artificial intelligence approach to machine translation // Schank R., Colby K. (Eds.). Computer Models of Thought and Language. San Francisco: Freeman, 1973. P. 114–151.

222. Winograd T. Understanding Natural Language. New York: Academic Press; Edinburgh: Edinburgh Univ. Press, 1973.

223. Woods W., Kaplan R. The lunar sciences natural language information system // BBN Report. No. 2265. Bolt, Beranek and Newman, Cambridge, Mass, 1971.

224. (Wooldridge et al. 1995). Wooldridge M., Jennings N. (1995). Intelligent Agents: Theory and Practice // Knowledge Engineering Review. Vol. 2. No. 10.

225. Wooldridge M. (1998). Verifiable semantics for agent communication languages. Proc. of the International Conference on Multi-Agent Systems (ICMAS-98), Paris, France, July 2–7, 1998, IEEE Press.

226. Xu L., Jeusfeld M.A. A concept for monitoring of electronic contracts. Tilburg University, The Netherlands, 2003. (<http://infolab.uvt.nl/pub/itrs010.pdf>). 19 p.

227. Zhu M., Uchida H. Universal Word and UNL Knowledge Base // Proceedings of the International Conference on Universal Knowledge Language (ICUKL-2002), 25–29 November 2002, Goa of India; <http://www.unl.ias.edu/publications/UW%20and%20UNLKB.htm>.

Доказательства Лемм 1, 2 и Утверждения 3.5 из главы 3

1. Доказательство Леммы 1

По соображениям удобства ниже повторно приводятся базовое определение, используемое в Лемме 1, и формулировка леммы.

Определение. Пусть B – произвольный концептуальный базис (к.б.), $n \geq 1$, для $i = 1, \dots, n$ $c_i \in D(B)$, $s = c_1 \dots c_n$, $1 \leq k \leq n$. Тогда через $lt_1(s, k)$ и $lt_2(s, k)$ обозначим количество вхождений символа '(' и символа '<' соответственно в подцепочку $c_1 \dots c_k$ цепочки $s = c_1 \dots c_n$. Через $rt_1(s, k)$ и $rt_2(s, k)$ обозначим количество вхождений символа ')' и символа '>' в подцепочку $c_1 \dots c_k$ цепочки s . Если в подцепочку $c_1 \dots c_k$ не входит символ '(' или символ '<', то соответственно $lt_1(s, k) = 0$, $lt_2(s, k) = 0$, $rt_1(s, k) = 0$, $rt_2(s, k) = 0$.

Лемма 1. Пусть B – произвольный к.б., $y \in Ls(B)$, $n \geq 1$, для $i = 1, \dots, n$ $c_i \in D(B)$, $y = c_1 \dots c_n$. Тогда:

(а) при $n > 1$ для каждого $k = 1, \dots, n-1$ и каждого $m = 1, 2$ $lt_m(y, k) \geq rt_m(y, k)$;

(б) $lt_m(y, n) = rt_m(y, n)$.

Доказательство.

Условимся использовать в доказательстве следующие соглашения об обозначениях. Во-первых, будем считать равнозначными обозначения $lt_1(y, k)$ и $lt1(y, k)$, $lt_2(y, k)$ и $lt2(y, k)$, $rt_1(s, k)$ и $rt1(y, k)$, $rt_2(y, k)$ и $rt2(y, k)$, где y – цепочка, k – номер позиции символа из $D(B)$ в цепочке y . Во-вторых, для $m = 0, 1, 2, \dots, 10$ правило $P[m]$ может обозначаться также через P_m . В-третьих, длину (количество элементов из $Ds(B)$) произвольной цепочки $s \in Ds^+(B)$ будем обозначать через $l(s)$ или *Длина*(s).

Пусть B – произвольный к.б., $y \in Ls(B)$, $n \geq 1$, для $i = 1, \dots, n$ $c_i \in D(B)$, $y = c_1 \dots c_n$. Если $n = 1$, то из правил $P_0, P_1, \dots, P_{10}$ непосредственно вытекает, что $y \in X(B) \cup V(B)$.

По определению концептуального базиса, символы, входящие в $X(B)$ или $V(B)$, отличны от символов '(', ')', '<', '>'. Поэтому $lt1(y, n) = rt1(y, n) = 0$, $lt2(y, n) = rt2(y, n) = 0$.

Если $n > 1$, то y строится с помощью каких-то правил из списка $P_1, \dots, P_{10}$ (помимо правила P_0). Докажем лемму индукцией по количеству q применений правил $P_1, \dots, P_{10}$ для построения y .

Часть 1: Рассмотрим случай $q = 1$.

1.1. Пусть цепочка y получена применением один раз правила P_2 или P_4 . Тогда $y = f(u_1, \dots, u_m)$ или $y = r(u_1, \dots, u_m)$, где $m \geq 1, f \in F(B), r \in R(B), u_1, \dots, u_m \in X(B) \cup V(B)$.

Тогда $lt1(y, 1) = rt1(y, 1) = 0, lt1(y, n) = rt1(y, n) = 1$, для $1 < k < n$ $lt2(y, k) = 1, rt2(y, k) = 0$, для $k = 1, \dots, n$ $lt2(y, k) = rt2(y, k) = 0$.

1.2. Если цепочка y получена применением P_3 , то найдутся такие $u_1, u_2 \in X(B) \cup V(B)$, что $y = (u_1 \equiv u_2)$. Тогда для $k = 1, \dots, n-1$ $lt1(y, k) = 1, rt1(y, k) = 0, lt1(y, n) = rt1(y, n) = 1$, для $k = 1, \dots, n$ $lt2(y, k) = rt2(y, k) = 0$.

1.3. Предположим, что цепочка y получена применением ровно один раз правила P_7 из связки $bin \in \{\wedge, \vee\}$ и элементов $d_1, \dots, d_m \in X(B) \cup V(B)$, где $m > 1$.

Тогда $y = (d_1 \text{ bin } d_2 \text{ bin } \dots \text{ bin } d_m)$. Следовательно, для $k = 1, \dots, n-1$ $lt1(y, k) = 1, rt1(y, k) = 0, lt1(y, n) = 1, rt1(y, n) = 1$, для $k = 1, \dots, n$ $lt2(y, k) = rt2(y, k) = 0$.

1.4. Если цепочка y построена применением один раз правила P_1 или P_6 , то y не включает символов '(', ')', '<', '>'. Тогда $lt1(y, k) = rt1(y, k) = lt2(y, k) = rt2(y, k) = 0$ для $k = 1, \dots, n$.

1.5. Анализ правила P_9 показывает, что цепочка y не может быть построена из элементов $X(B)$ и $V(B)$ применением ровно один раз правила P_9 (и, возможно, P_0). Действительно, если это не так, то $y = Qv(\text{concept})des$, где $Q \in \{\forall, \exists\}, v \in V(B), \text{concept} \in X(B), des \& P \in Ts(B)$, где P -сорт «смысл сообщения». Но цепочка des должна включать переменную v , где $tp(v) = [\text{сущн}]$. Поэтому $des = v$. Последнее соотношение противоречит условию $des \& P \in Ts(B)$.

1.6. Если цепочка y получена применением один раз правила P_{10} , то y не включает символов '(', ')', и первым символом y является '<'. Тогда для $k = 1, \dots, n$ $lt1(y, k) = 1, rt1(y, k) = 0$, для $k = 1, \dots, n-1$ $lt2(y, k) = 1, rt2(y, k) = 0, lt2(y, n) = rt2(y, n) = 1$.

Часть 2: Пусть существует такое $q \geq 1$, что утверждения (а), (б) Леммы 1 справедливы для каждой цепочки y , построенной из элементов $X(B)$ и $V(B)$ применением не более q раз правил $P_1, \dots, P_{10}$ (и, возможно, применением произвольного количества раз правила P_0). Докажем, что утверждения (а), (б) выполняются и для всякой цепочки y , полученной применением $q + 1$ раз правил из $P_1, \dots, P_{10}$. Рассмотрим возможные случаи.

2.1. $y = f(a_1, \dots, a_m)$ или $y = r(a_1, \dots, a_m)$, где $m > 1$. Для $a_1, \dots, a_m$ применим условия (а), (б). Поэтому, очевидно, $lt1(y, 1) = rt1(y, 1) = 0, lt1(y, 2) = 1, rt1(y, 2) = 0$; для $i = 3, \dots, n-1$ $lt1(y, i) > rt1(y, i), lt1(y, n) = rt1(y, n)$.

2.2. Если $y = (a_1 \equiv a_2)$, то из индуктивного предположения следует, что $ltl(y, 1) = 1$, $rtl(y, 1) = 0$, для $i = 2, \dots, n-1$ $ltl(y, n) = rtl(y, n)$.

2.3. Пусть существует такая связка $b \in \{\wedge, \vee\}$, такое $m > 1$ и такие цепочки $a_1, \dots, a_m$, что цепочка y получена из $b, a_1, \dots, a_m$ применением ровно один раз P_7 . Тогда $y = (a_1 b a_2 b \dots b a_m)$.

Пусть $Длина(a_1) = n_1, \dots, Длина(a_m) = n_m$. Очевидно, для $a_1, \dots, a_m$ справедливы утверждения (а), (б) Леммы 1. Заметим, что связка b занимает в y позиции $n_1 + 2, n_1 + n_2 + 3, n_1 + \dots + n_m + (m + 1)$. Тогда $ltl(y, 1) = 1$, $rtl(y, 1) = 0$; для $p = n_1 + 2, n_1 + n_2 + 3, n_1 + n_2 + \dots + n_m + m + 1$ $rtl(y, p) = ltl(y, p) - 1$; для $i = 1, \dots, m - 1$ и для каждого такого p , что $n_1 + \dots + n_i + i + 1 < p < n_1 + \dots + n_i + n_{i+1} + i + 2$ $rtl(y, p) \leq ltl(y, p) - 1$; $ltl(y, n) = rtl(y, n)$.

2. Доказательство Леммы 2

Для удобства анализа излагаемого ниже доказательства приведем повторно формулировку Леммы 2 из раздела 3.8.

Лемма 2. Пусть B – произвольный к.б., $y \in Ls(B)$, $n > 1$, $y = c_1 \dots c_n$, где для $i = 1, \dots, n$ $c_i \in D(B)$, цепочка y включает запятую или какой-либо из символов $\equiv, \wedge, \vee$, и k – такое произвольное натуральное число, что $1 < k < n$. Тогда:

(а) если c_k – один из символов $\equiv, \wedge, \vee$, то $ltl(y, k) > rtl(y, k) \geq 0$;

(б) если c_k – запятая, то выполняется по крайней мере одно из соотношений

$$\begin{aligned} ltl(y, k) > rtl(y, k) &\geq 0, \\ ltl_2(y, k) > rtl_2(y, k) &\geq 0. \end{aligned}$$

Доказательство.

Пусть для B и y выполняются предположения Леммы 2, и $Symb = \{\equiv, \wedge, \vee\}$. Поскольку y включает хотя бы один из элементов множества $Symb$, постольку цепочка y построена из элементов множества $X(B) \cup V(B)$ и нескольких вспомогательных символов применением r раз, где $r \geq 1$, каких-то правил из списка $P_1, \dots, P_{10}$ (и, возможно, применением какого-то количества раз правила P_0).

Докажем лемму индукцией по r .

Случай 1. Пусть $r = 1$. Тогда справедливость доказываемого утверждения для y , непосредственно вытекает из рассмотрения всех ситуаций, анализировавшихся в Части 1 доказательства Леммы 1.

Случай 2. (индуктивный переход). Пусть существует такое $r \geq 1$, что утверждение Леммы справедливо для произвольного к.б. B и для каждого такого $y \in L_s(B)$, что при построении y использовалось m пра-

вил из списка $P_1, \dots, P_{10}$, где $1 \leq m \leq r$. Докажем, что тогда утверждение леммы справедливо и для каждого такого $z \in Ls(B)$, что при построении u использовалось $r + 1$ правило из списка $P_1, \dots, P_{10}$.

Случай 2.1а. Пусть на последнем шаге построения цепочки z использовалось правило P_7 . Тогда цепочка z представима в виде $z = (y_1 q y_2 q \dots q y_m)$, где $m > 1$, $q \in \{\wedge, \vee\}$, для $i = 1, \dots, m$ $y_i \in Ls(B)$. По предположению индукции, при построении каждой из цепочек $y_1, \dots, y_m$ использовались не более r раз правила $P_1, \dots, P_{10}$.

Пусть $2 < k < l(y_1) + 1$. Тогда, очевидно, символ c_k является внутренней частью цепочки y_1 . По индуктивному предположению, если $lt1(y_1, k - 1) > 0$, то $lt1(y_1, k - 1) > rt1(y_1, k - 1)$; если $lt2(y_1, k - 1) > 0$, то $lt2(y_1, k - 1) > rt2(y_1, k - 1)$.

Но $lt1(z, k) = lt1(y_1, k - 1) + 1$, $rt1(z, k) = rt1(y_1, k - 1)$, поэтому $lt1(z, k) > rt1(z, k)$.

Кроме того, $lt2(z, k) = lt2(y_1, k - 1)$, $rt2(z, k) = rt2(y_1, k - 1)$.

Поэтому, в силу индуктивного предположения, если $lt2(z, k) > 0$, то $lt2(z, k) > rt2(z, k)$.

Случай 2.1б. Предположим, как и ранее, что на последнем шаге построения цепочки z применялось правило P_7 , и цепочка z представима в виде $(y_1 q y_2 q \dots q y_m)$ для некоторых $m > 1$, $q \in \{\wedge, \vee\}$, $y_1, \dots, y_m \in Ls(B)$. Кроме того, пусть найдется такое i , $1 \leq i \leq m$, что

$$1 + l(y_1) + 1 + \dots + l(y_i) + 2 < k < 1 + l(y_1) + 1 + \dots + l(y_i) + 1 + l(y_{i+1}),$$

где $l(y_h) = \text{Длина}(y_h)$, $h = 1, \dots, i + 1$. Это означает, что символ c_k является внутренним элементом некоторой подцепочки $y_{i+1} \in Ls(B)$.

Пусть k -я позиция рассматриваемой цепочки z является p -й позицией цепочки y_{i+1} . Тогда

$$lt1(z, k) = 1 + lt1(y_1, l(y_1)) + \dots + lt1(y_i, l(y_i)) + lt1(y_{i+1}, p),$$

$$rt1(z, k) = rt1(y_1, l(y_1)) + \dots + rt1(y_i, l(y_i)) + rt1(y_{i+1}, p).$$

В силу Леммы 1, для $j = 1, \dots, i$ $lt1(y_j, l(y_j)) = rt1(y_j, l(y_j))$, $lt2(y_j, l(y_j)) = rt2(y_j, l(y_j))$, $lt1(y_{i+1}, p) \geq rt1(y_{i+1}, p)$. Поэтому $lt1(z, k) > rt1(z, k)$.

Случай 2.1в. Пусть существуют такие $m > 1$, $y_1, \dots, y_m \in Ls(B)$, $q \in \{\wedge, \vee\}$, что $z = (y_1 q y_2 q \dots q y_m)$, и найдется такое i , $1 \leq i < m$, что

$$k = 1 + l(y_1) + 1 + \dots + l(y_i) + 1 = l(y_1) + \dots + l(y_i) + i + 1.$$

Это означает, что c_k является вхождением логической связки q , разделяющей y_i и y_{i+1} . Из Леммы 1 следует, что для $s = 1, \dots, i$ $lt1(y_s, l(y_s)) = rt1(y_s, l(y_s))$. Поэтому

$$lt1(z, k) = 1 + lt1(y_1, l(y_1)) + \dots + lt1(y_i, l(y_i)),$$

$$rt1(z, k) = rt1(y_1, l(y_1)) + \dots + rt1(y_i, l(y_i)).$$

Тогда $lt1(z, k) = rt1(z, k) + 1$.

Случай 2.2. На последнем шаге построения цепочки z использовалось правило P_3 . Тогда найдутся такие $y_0, y_2 \in Ls(B)$, что z представима в виде $(y_0 \equiv y_2)$. Очевидно, этот случай рассматривается аналогично случаю 2.1 при $m = 2$.

Случай 2.3. На последнем шаге построения цепочки z использовалось правило P_2 или P_4 . Тогда найдутся такие $y_1, \dots, y_n \in Ls(B)$, что z представима в виде $f(y_1, \dots, y_n)$ или в виде $r(y_1, \dots, y_n)$ соответственно, где (f – n -арный функциональный символ, r – n -арный реляционный символ). Очевидно, этот случай рассматривается аналогично случаю 2.1.

Случай 2.4. На последнем шаге построения цепочки z использовалось правило P_8 или P_9 или P_{10} . Эти случаи снова рассматриваются аналогично случаю 2.1.

3. Доказательство Утверждения 3.5

Пусть B – произвольный к. б., $z \in Ls(B) \setminus (X(B) \cup V(B))$. Тогда из структуры правил $P[0], P[1], \dots, P[10]$ и определения множества $Ls(B)$ непосредственно вытекает существование таких k , где $1 \leq k \leq 10, n \geq 1$, и $y_0, y_1, \dots, y_n \in Ls(B)$, что

$$y_0 \& y_1 \& \dots \& y_n \& z \in Ynr_{10}^k(B).$$

Поэтому основное внимание нужно уделить доказательству единственности такого набора $(k, n, y_0, \dots, y_n)$. Напомним, что, в соответствии с определением из раздела 2.8, для произвольного к.б. B $D(B) = X(B) \cup V(B) \cup \{', ', '(', ')', ':', '*', '<', '>'\}$, $Ds(B) = D(B) \cup \{', '\}$, $D^+(B)$ и $Ds^+(B)$ – множества всех непустых конечных последовательностей элементов из $D(B)$ и $Ds(B)$ соответственно.

Предположим, что B – произвольный к. б., z – произвольная формула из $Ls(B) \setminus (X(B) \cup V(B))$, $1 \leq k \leq 10, n \geq 1$, и $y_0, y_1, \dots, y_n$ – такие формулы из $Ls(B)$, что

$$y_0 \& y_1 \& \dots \& y_n \& z \in Ynr_{10}^k(B). \quad (П1)$$

Докажем, что в таком случае набор $(k, n, y_0, \dots, y_n)$ является единственным набором, для которого выполняется условие (П1). С этой целью потребуются рассмотреть довольно много различных возможных случаев. Длину (количество элементов из $Ds(B)$) произвольной цепочки $s \in Ds^+(B)$ будем обозначать через $l(s)$. Так как элементы первичного информационного универсума $X(B)$ и переменные из $V(B)$ считаются символами, то длина каждой цепочки из $X(B)$ и $V(B)$ равна 1. Число 0 будем считать длиной пустой цепочки.

Случай 1. Первый символ цепочки z , обозначаемый через $z[1]$, является левой скобкой '('. Тогда, очевидно, $k = 3$ либо $k = 7$. Если $k = 3$,

$n = 2$, и z – цепочка вида $(y_0 \equiv y_2)$, причем y_1 – это символ $\equiv$. Если же $k = 7$, то $n \geq 2$, y_0 – логическая связка $\vee$ или $\wedge$, и z – цепочка вида $(y_1 y_0 y_2 y_0 \dots y_0 y_n)$, т.е. цепочка вида $(y_1 \wedge y_2 \wedge \dots \wedge y_n)$ при $y_0 = \wedge$, и цепочка вида $(y_1 \vee y_2 \vee \dots \vee y_n)$ при $y_0 = \vee$.

Допустим, что z – цепочка вида $(y_0 \equiv y_2)$, где $y_0, y_2 \in D^+(B)$, и докажем, что z нельзя представить другим способом. Доказательство будем осуществлять методом от противного.

Случай 1а. Предположим, что найдутся такие $w_1, w_2 \in Ls(B)$, что $l(y_1) > l(w_1)$, и z можно представить в виде $(w_1 \equiv w_2)$. Тогда цепочка $w_1 \equiv$ является началом цепочки y_0 . Пусть найдутся такие k, m , что $1 \leq k < m$, $y_1 = c_1 \dots c_m$, $w_1 = c_1 \dots c_k$, где $c_1 \dots c_m \in D(B)$. Тогда $c_{k+1} = \equiv$ и, поскольку никакая цепочка из $Ls(B)$ не оканчивается символом $\equiv$, постольку $k + 1 < m$.

Так как y_0 включает символ $\equiv$, то в y_1 в некоторой позиции $j < k + 1$ встречается левая скобка '('. Поэтому $lt_1(y_1, k + 1) > 0$. Но из соотношения $c_{k+1} = \equiv$ следует (в силу Леммы 2), что $lt_1(y_0, k + 1) > rt_1(y_0, k + 1)$. Очевидно, $lt_1(y_0, k + 1) = lt_1(w_1, k)$, $rt_1(y_0, k + 1) = rt_1(w_1, k)$. Поэтому, $lt_1(w_1, k) > rt_1(w_1, k)$. Однако, в силу Леммы 1, из $w_1 \in Ls(B)$ и $k = l(w_1)$ вытекает, что $lt_1(w_1, k) = rt_1(w_1, k)$. Таким образом, получено противоречие.

Аналогичные рассуждения можно провести и в случае предположения о существовании таких $w_1, w_2 \in Ls(B)$, что $l(w_1) > l(y_0)$, $l(w_2) < l(y_2)$, и $z = (w_1 \equiv w_2)$. Таким образом, таких w_1 и w_2 не существует.

Случай 1б. Как и ранее, будем предполагать, что цепочка z представлена в виде $(y_0 \equiv y_2)$, где $y_0, y_2 \in D^+(B)$. Допустим также, что найдутся такие $q \in \{\wedge, \vee\}$, $m > 1$, и $w_1, \dots, w_m \in Ls(B)$, что $z = (w_1 q w_2 q \dots q w_m)$.

Очевидно, что $l(w_1) \neq l(y_2)$. Пусть $l(y_0) < l(w_1)$. Тогда цепочка $y_0 \equiv$ является началом цепочки $w_1 \in Ls(B)$. Отсюда легко получаем противоречие, повторяя рассуждения Случая 1а.

Если $l(w_1) < l(y_0)$, то цепочка $w_1 q$ является началом цепочки $y_0 \in Ls(B)$. В этом случае снова применима Лемма 2 и рассуждения, использовавшиеся при рассмотрении Случая 1а. Таким образом, если $z \in Ls(B)$ и для некоторых $y_0, y_2 \in Ls(B)$ z – цепочка вида $(y_0 \equiv y_2)$, где $y_0, y_2 \in D^+(B)$, то z не могла быть получена с помощью какого-то другого правила (на последнем шаге ввода), кроме правила $P[3]$, и только из «блоков» y_0, y_2 .

Случай 1в. Пусть существуют такие $q \in \{\wedge, \vee\}$, $m > 1$, и $y_1, \dots, y_m \in Ls(B)$, что $z = (y_1 q y_2 q \dots q y_m)$. Тогда необходимо рассмотреть две ситуации:

(1) найдутся такие $u_1, u_2 \in Ls(B)$, что z – цепочка вида $(u_1 \equiv u_2)$;

(2) существуют такие $p > 1$, $w_1, \dots, w_p \in Ls(B)$, $d \in \{\wedge, \vee\}$ такие, что $z = (w_1 d w_2 d \dots d w_p)$, причем набор $(q, y_1, \dots, y_m)$ отличен от набора $(d, w_1, \dots, w_p)$.

Рассмотрим ситуацию (1). Пусть $l(u_1) < l(y_1)$, тогда $u_1, y_1 \in Ls(B)$, и цепочка $u_1 \equiv$ является началом цепочки y_1 . Эта ситуация была рассмотрена в Случае 1а, где мы получили противоречие.

Если $l(y_1) < l(u_1)$, то $u_1, y_1 \in Ls(B)$, и цепочка $y_1 q$ является началом цепочки u_1 ; поэтому, очевидно, $l(u_1 q) < l(u_1)$.

Воспользуемся Леммой 2. Пусть $y_1 q = c_1 \dots c_r c_{r+1}$, где для $i = 1, \dots, r+1$ $c_i \in D(B)$, $c_{r+1} = q$. Так как c_{r+1} – символ ' $\equiv$ ', то, по Лемме 2, $lt_1(u_1, r+1) > rt_1(u_1, r+1)$. Но $lt_1(u_1, r+1) = lt_1(u_1, r) = lt_1(y_1, r)$, $rt_1(u_1, r+1) = rt_1(u_1, r) = rt_1(y_1, r)$. Однако, поскольку $y_1 \in Ls(B)$, то, в силу Леммы 1, $lt_1(y_1, r) = rt_1(y_1, r)$. Получили противоречие.

Проанализируем ситуацию (2). Предположим, что $l(w_1) < l(y_1)$. Тогда $w_1, y_1 \in Ls(B)$, и цепочка $w_1 d$ является началом y_1 , где d – один из символов $\wedge, \vee$. Но отсюда можно получить противоречие точно так же, как и при разборе ситуации (1). Очевидно, что случай $l(y_1) < l(w_1)$ симметричен только что рассмотренному случаю.

Предположим, что $q = d$, и найдется такое j , $1 \leq j < m$, что $y_1 = w_1, \dots, y_j = w_j$, но $w_{j+1} \neq y_{j+1}$. Тогда либо цепочка $w_{j+1} d$ является началом цепочки y_{j+1} , либо $y_{j+1} d$ является началом w_{j+1} . Поэтому мы оказываемся в ситуациях, рассмотренных непосредственно выше.

Таким образом, рассмотрение Случая 1 закончено. Мы доказали следующее промежуточное утверждение:

Пусть B – произвольный к.б., $z \in Ls(B) \setminus (X(B) \cup V(B))$, и первый символ цепочки z является левой скобкой ' $($ '. Тогда найдется единственный такой набор $(k, n, y_0, \dots, y_n)$, где $1 \leq k \leq 10$, $n \geq 1$, $y_0, y_1, \dots, y_n \in Ls(B)$, что $y_0 \& y_1 \& \dots \& y_n \& z \in Lnr_{10}^k(B)$; при этом $k = 3$ или $k = 7$.

Случай 2. Пусть B – произвольный к.б., z – произвольная формула из $Ls(B) \setminus (X(B) \cup V(B))$, $1 \leq k \leq 10$, $n \geq 1$, и $y_0, y_1, \dots, y_n$ – такие формулы из $Ls(B)$, что выполнено соотношение (П1). Предположим, что z начинается с функционального символа, т.е. $z[1] \in F(B)$, и предпоследний символ z отличен от ':'. Тогда из структуры правил $P[0], P[1], \dots, P[10]$ непосредственно вытекает, что цепочка z построена в результате применения правила $P[2]$ на последнем шаге вывода. Поэтому $z = f(y_1, \dots, y_n)$, где $n \geq 1$, $y_1, \dots, y_n \in Ls(B)$.

Если $f \in F_1(B)$, т.е. f – одноместный функциональный символ, то $n = 1$ и y_1 определяется однозначно по z . Предположим, что $n > 1$, $f \in F_n(B)$, и найдется такая последовательность $u_1, \dots, u_n$, отличная от последовательности $y_1, \dots, y_n$, что $u_1, \dots, u_n \in Ls(B)$, и цепочка $(y_1, \dots, y_n)$

совпадает с цепочкой $(u_1, \dots, u_n)$. Применяя Лемму 2 аналогично тому, как это было сделано при рассмотрении Случая 1в, можно легко показать, что два разных представления $(y_1, \dots, y_n)$ и $(u_1, \dots, u_n)$ цепочки $z \in Ls(B)$ невозможны точно так же, как и два разных представления $(y_1 d y_2 d \dots d y_n)$ и $(u_1 d u_2 d \dots d u_n)$, где $d \in \{\wedge, \vee\}$.

Случай 3. Для B и z выполнены предположения формулировки Утверждения 2.5, и z начинается с реляционного символа, причем z не является функциональным символом (т.е. $z[1] \in R(B) \setminus F(B)$), и предпоследний символ z отличен от двоеточия. Этот случай аналогичен Случаю 2, только z строится в результате применения правила P_4 на последнем шаге. Поэтому для z найдется единственный набор $(k, n, y_0, \dots, y_n)$, такой, что $1 \leq k \leq 10$, $n \geq 1$, $y_0, y_1, \dots, y_n \in Ls(B)$, и $y_0 \& y_1 \& \dots \& y_n \& z \in Ynr_1^k(B)$, причем $k = 4$.

Случай 4. Пусть B – произвольный к. б., $z \in Ls(B) \setminus (X(B) \cup V(B))$, для $i = 1, \dots, l(z)$ $z[i]$ – i -й символ цепочки, $z[1] \in X(B)$, $z[2] = '*'$, предпоследний символ z отличен от двоеточия. Тогда из правил $P_0, P_1, \dots, P_{10}$ непосредственно следует, что при построении цепочки z на последнем шаге использовалось правило P_8 . Поэтому найдутся такие $p \geq 1$ и набор $(r_1, b_1, \dots, r_p, b_p)$, где для $i = 1, \dots, p$ $r_i \in R_2(B)$, $b_i \in Ls(B)$, что, если $a = z[1]$, то

$$z = a^*(r_1, b_1) \dots (r_p, b_p). \quad (П2)$$

Остается доказать, что для всякого набора $(k, n, y_0, \dots, y_n)$, такого, что $1 \leq k \leq 10$, $n > 1$, для $i = 0, \dots, n$ $y_i \in Ls(B)$, из соотношения $y_0 \& y_1 \& \dots \& y_n \& z \in Ynr_1^k(B)$ следует, что набор $(k, n, y_0, \dots, y_n)$ совпадает с набором $(8, 2p + 1, a, r_1, b_1, \dots, r_p, b_p)$.

Из соотношения (П2) вытекает, что необходимо рассмотреть только следующую ситуацию: найдутся такие $m \geq 1$, $h_1, \dots, h_m \in R_2(B)$, $d_1, \dots, d_m \in Ls(B)$, что

$$Z = a^*(h_1, d_1) \dots (h_m, d_m), \quad (П3)$$

причем представление (П3) отличается от представления (П2).

Элементы $a, r_1, \dots, r_p, h_1, \dots, h_m$ интерпретируются как символы. Поэтому, очевидно, $r_1 = h_1$. Пусть $d_1 \neq b_1$, и $l(b_1) < l(d_1)$. Тогда цепочка b_1 является началом цепочки d_1 . Используя Лемму 2 и Лемму 1 аналогично тому, как это делалось при разборе Случаев 1а и 1в, получим противоречие. Следовательно, $b_1 = d_1$.

Пусть для некоторого i , такого, что $1 \leq i b < \min\{p, m\}$, для $j = 1, \dots, i$ $r_j = h_j$, $b_j = d_j$, и $r_{j+1} = h_{j+1}$, $b_{j+1} \neq d_{j+1}$. Тогда, если $l(b_{j+1}) < l(d_{j+1})$, то цепочка b_{j+1} является началом цепочки d_{j+1} , а подобную ситуацию мы только что рассмотрели. Если же $l(b_{j+1}) > l(d_{j+1})$, то цепочка d_{j+1} является началом цепочки b_{j+1} , и мы, по существу, имеем дело с той же

ситуацией. Таким образом, представления (П2) и (П3) цепочки z совпадают.

Случай 5. Для B и z выполняются предположения формулировки Утверждения 2.5, $z[1]$ – квантор $\forall$ или $\exists$, предпоследний символ $z \neq ':$. Тогда, очевидно, на последнем шаге вывода z применялось правило P_9 . Поэтому z – цепочка вида

$$Q \vee (conc) A, \quad (\text{П4})$$

где $Q = z[1]$, $v \in V(B)$; $conc, A \in Ls(B)$; $conc \in X(B)$ или $conc$ – цепочка, при построении которой на последнем шаге применялось правило P_8 ; $A \& P \in Ts(B)$, где P – сорт «смысл сообщения» базиса B .

Пусть $l(z) = m$ (т.е. m – длина z). Если $conc \in X(B)$, то однозначность представления z очевидна: $Q = z[1]$, $v = z[2]$, $'(= z[3]$, $conc = z[4]$, $')$ $= z[5]$, $A = z[6] \dots z[m]$. Если же $conc \in Ls(B) \setminus (X(B) \cup V(B))$, то воспользуемся Леммой 1. Пусть $l(conc) = p$, где $p > 1$. Тогда для $i = 1, \dots, p$ $lt_1(conc, i) \geq rt_1(conc, i)$ и $lt_1(conc, p) = rt_1(conc, p)$. Поэтому для $q = 1, \dots, p + 3$ $lt_1(z, q) > rt_1(z, q)$ и $lt_1(z, p + 4) = rt_1(z, p + 4)$. Следовательно, позиция правой скобки $')$ сразу после $conc$ определяется как такое наименьшее число $s \geq 1$, что $lt_1(z, s) > rt_1(z, s)$ и $lt_1(z, s) = rt_1(z, s)$.

Таким образом, если $z \in Ls(B)$ и $z[1]$ – квантор Q , где $Q = \forall$ или $Q = \exists$, то существует единственный такой набор $(k, n, y_0, \dots, y_n)$ где $1 \leq k \leq 10, n \geq 1, y_0, y_1, \dots, y_n \in Ls(B)$, что

$$y_0 \& y_1 \& \dots \& y_n \& z \in Ynr_{10}^k(B);$$

при этом $k = 9, n = 4, y_0 = Q, y_1 = v, y_3 = conc, y_4 = A$ (см. представление (П4)).

Случай 6. Пусть для B и z выполнены предположения формулировки Утверждения 3.5, $z[1] = '<'$, и предпоследний символ z отличен от двоеточия. Тогда очевидно, что на последнем шаге построения z использовалось правило P_{10} . Предположим, что существует такая последовательность $m, u_1, \dots, u_n$ и отличная от нее последовательность $n, y_1, \dots, y_n$, где $m, n > 1, u_1, \dots, u_m, y_1, \dots, y_n \in Ls(B)$, что z может быть представлена в виде $<u_1, \dots, u_m>$ и $<y_1, \dots, y_n>$.

Применяя Лемму 2 подобно тому, как это было сделано при рассмотрении Случая 2, получаем, что два различных представления z невозможны точно так же, как невозможны два различных представления $f(u_1, \dots, u_m)$ и $f(y_1, \dots, y_n)$ одной и той же цепочки из $Ls(B)$.

Случай 7. Пусть для B и z справедливы предположения формулировки Утверждения 3.5, $z[1]$ – интенциональный квантор из $Int(B)$, z не имеет окончания : v , где $v \in V(B)$, $m = l(z)$. Тогда очевидно, что z построена с помощью правила с номером $k = 1$ из «блоков» $y_1 = z[1]$ и $y_2 = z[2] \dots z[m]$.

Случай 8. Пусть B – произвольный к. б., $z \in Ls(B) \setminus (X(B) \cup V(B))$, $z[1] = \neg$. Тогда найдется такая цепочка $y \in Ls(B)$, что $z = \neg y$. Анализ структуры правил $P_0, P_1, \dots, P_{10}$ показывает, что возможны только две ситуации: (1) y не имеет окончания : v , где $v \in V(B)$, и тогда z построена из набора $(\neg, y)$ в результате применения правила P_6 ; (2) $y = w : v$, где $v \in V(B)$, и тогда z построена из набора $(\neg w, v)$ в результате применения один раз правила P_5 .

Ситуация же, когда z построена из набора $(\neg, w : v)$ в результате применения правила P_6 , невозможна, потому что правило P_6 позволяет строить выражения $\neg a$ только в том случае, когда либо $a \in X(B)$, либо на последнем шаге построения a не использовалось правило с номером 2, 5 или 10.

Случай 9. Пусть B – произвольный к.б., $z \in Ls(B) \setminus (X(B) \cup V(B))$, окончанием z является подцепочка : v , где $v \in V(B)$, и z не начинается с символа ' $\neg$ '. Покажем, что в этом случае на последнем шаге построения z применялось правило P_5 .

Предположим, что это не так, и на последнем шаге построения z применялось правило P_m , где $1 \leq m \leq 10$, $m \neq 5$, $m \neq 6$. Если $m = 1$, то $z = q \text{ des}$, где q – интенсиональный квантор из $Int(B)$, $\text{des} \in X(B)$ или

$$\text{Des} = a^*(r_1, b_1) \dots (r_p, b_p),$$

где $p \geq 1$, $a \in X(B)$, $r_1, \dots, r_n \in R(B)$, $b_1, \dots, b_p \in Ls(B)$.

Так как окончанием z является подцепочка : v , где $v \in V(B)$, то ситуация $z = q \text{ des}$, где $q \in Int(B)$, $\text{des} \in X(B)$, исключается. Но $v \neq ')$ ', следовательно, получаем противоречие, и ситуация $m = 1$ невозможна.

Если $m = 2, 3, 4, 7, 8$, то последним символом z является скобка ')', поэтому каждая из этих ситуаций невозможна. Если $m = 10$, то последним символом z является угловая скобка '>', что тоже невозможно.

Допустим, что последним применялось правило $P[9]$. Тогда найдутся такой квантор $qex \in \{\exists, \forall\}$, $v \in V(B)$, $\text{des} \in Ls(B)$, $A \in Ls(B)$, что $z = qex \ v \ (\text{des}) \ A$.

Согласно определению правила P_9 , цепочка A не имеет окончания вида : z , где z – произвольная переменная из $V(B)$. Следовательно, в Случае 9 на последнем шаге построения z применялось правило $P[5]$. Таким образом, рассмотрены все возможные случаи, и Утверждение 3.5 доказано.

УДК 004.457

ББК 32.973.26-018.2

Ф20



Издание осуществлено в рамках
Инновационной образовательной программы ГУ ВШЭ
«Формирование системы аналитических компетенций
для инноваций в бизнесе и государственном управлении»

Фомичёв В. А. Математические основы представления содержания посланий компьютерных интеллектуальных агентов [Текст] / В. А. Фомичёв ; Гос. ун-т — Высшая школа экономики. — М. : ТЕИС, 2007. — 174 [2] с. — 1000 экз. — ISBN 978-5-7598-0516- (в обл.).

В монографии предложен оригинальный, широко применимый (по гипотезе автора — универсальный) математический подход к представлению содержания посланий компьютерных интеллектуальных агентов (КИА) в многоагентных системах, в частности, в системах электронной коммерции. Выразительные возможности определяемого в книге класса СК-языков (стандартных концептуальных языков) превышают возможности основных известных подходов к этой проблеме, в том числе возможности языка представления содержания посланий КИА, разработанного в рамках международного Фонда интеллектуальных физических агентов (FIPA). Обоснованы возможности использования аппарата СК-языков в проектировании лингвистических процессоров (компьютерных систем, осуществляющих смысловую обработку письменных текстов или устной речи на естественном языке) в произвольных предметных областях в качестве инструмента математического описания смысловой структуры не только предложений, но и сложных связных текстов (или дискурсов), относящихся к деловой прозе: текстов по медицине, технике, экономике, юриспруденции и т.д.

Значительная часть материалов монографии была опубликована в научных журналах «Информационные технологии», «Качество и ИПИ (CALS)-технологии», «Качество. Инновации. Образование», «Informatica» (Словения), «Cybernetica» (Бельгия) и трудах международных научных конференций и симпозиумов, проходивших в России, Австрии, Великобритании, Германии, Дании, Нидерландах, Словении, Франции.

Книга будет полезна разработчикам логико-информационных основ многоагентных систем и электронной коммерции, специалистам в области компьютерной лингвистики, а также студентам и начинающим ученым.

OCR by Palek

ISBN 978-5-7598-0516-8 © Фомичёв В.А., 2007

© Оформление. Издательство «ТЕИС», 2007