УДК 81'42
ББК Ш1

# НАУЧНЫЙ ТЕКСТ: ЛИНГВОКОГНИТИВНЫЙ ПОДХОД

*Т.Н. Хомутова, О.И. Бабина*

# RESEARCH TEXT: LINGUISTIC AND COGNITIVE APPROACH

*T.N. Khomutova, O.I. Babina*

**Рассматриваются вопросы анализа научного текста в русле лингвокогнитивного подхода, проводится корпусное исследование и стратификация лексического уровня научных текстов английского подъязыка «Программирование», результатом чего является частотный словарь и учебный лексический минимум, в состав которых входит общеупотребительная, общенаучная и терминологическая лексика, разрабатываются принципы построения предметной онтологии на базе терминологического минимума, приводится фрагмент онтологии предметной области «Программирование», намечаются возможные пути дальнейшего исследования и использования полученных результатов в прикладных целях.**

***Ключевые слова:*** *научный текст, лингвокогнитивный подход, корпусное исследование, частотность, стратификация лексики, программирование, концепт, предметная онтология.*

**The article investigates the use of a linguistic and cognitive approach to research text analysis. A corpus-based research into lexis of texts in programming is undertaken which results in frequency wordlists and a concise English dictionary for students in programming with general, scientific and terminological layers of lexis defined. Principles of constructing a domain ontology on the basis of the terminological layer are advanced. A fragment of a programming ontology is presented. The perspectives of further research and application of the results obtained are discussed.**

***Keywords:*** *research text, linguistic and cognitive approach, corpus-based research, frequency, lexis stratification, programming, concept, domain ontology.*

С позиций современной лингвистики текст вообще и научный текст в частности рассматривается как средство познания и описания действительности, как инструмент взаимодействия, как знаковый способ материализации знаний и передачи их от человека к человеку, от поколения к поколению, как средство воплощения модели мира, как средство воплощения культуры[1].

В рамках лингвистики текста Л.Г. Бабенко выделяет такие подходы к тексту, как лингвоцентрический, текстоцентрический, антропоцентрический и когнитивный[2].

*Лингвоцентрический* подход изучает элементы текста, а именно функционирование языковых единиц и категорий в условиях текста (Б.А. Ларин, Л.А. Новиков, К.А. Долинин и др.).

*Текстоцентрический* подход рассматривает текст как целостный завершенный объект исследования, структурно-семантическое целое, обладающее собственными текстовыми категориями и свойствами (Г.Н. Золотова, И.Р. Гальперин, Е.А. Гончарова, Е.В. Падучева, В.А. Кухаренко, И.Я. Чернухина и др.).

*Антропоцентрический* подход связан с интерпретацией текста в аспекте его порождения (автор) и восприятия (читатель), а также в аспекте воздействия на читателя. Антропоцентрический подход включает следующие направления изучения текста: психолингвистическое (Л.С. Выготский, Т.М. Дридзе, А.А. Леонтьев, И.А. Зимняя, Н.И. Жинкин, А.Р. Лурия, Л.В. Сахарный, А.М. Шахнарович); прагматическое (А.Н. Баранов); деривационное (Е.С. Кубрякова, Л.Н. Мурзин); коммуникативное (Г.А. Золотова, Н.С. Болотнова) и когнитивное (Е.С. Кубрякова).

**Хомутова Тамара Николаевна,** кандидат филологических наук, доцент, зав. кафедрой лингвистики и межкультурной коммуникации ЮУрГУ. E-mail: tnh@susu.ac.ru

**Бабина Ольга Ивановна,** кандидат филологических наук, доцент кафедры лингвистики и межкультурной коммуникации ЮУрГУ. E-mail: olga_babina@mail.ru

**Tamara N. Khomutova,** Candidate of Philology, Professor, Head of the Department of Linguistics and Cross-Cultural Communication. E-mail: tnh@susu.ac.ru

**Olga I. Babina,** Candidate of Philology, Assistant Professor, Department of Linguistics and Cross-Cultural Communication, SUSU. E-mail: olga_babina@mail.ru

*Когнитивное* направление/подход по мнению Е.С. Кубряковой рассматривает текст как сложный знак, основное средство выражения знаний автора о мире «в связи с процессами говорения и понимания как процессами взаимодействия психических субъектов»[3].

Рассмотренные подходы к исследованию текста не исчерпывают всего многообразия точек зрения на этот сложный объект. Необходимо упомянуть синергетический подход к тексту как самоорганизующейся системе (Р.Г. Пиотровский, Г.Г. Москальчук, В.А. Пищальникова), культурологический подход к тексту как единице культуры (М.М. Бахтин, Ю.М. Лотман, Л.Н. Мурзин), гендерный подход (О.Л. Каменская, О.А. Воронина, А.В. Кирилина) и другие, подробное рассмотрение которых выходит за рамки настоящего исследования.

Мы разделяем точку зрения ученых, согласно которой анализ научного текста, как правило, связывается с особенностями научного познания, экстралингвистическими факторами, составляющими процесс познания истины. В связи с этим особое значение для современного изучения научного текста приобретают положения когнитивной лингвистики. Поскольку научный текст рассматривается как средство хранения, передачи и обогащения специальных знаний, это обусловливает не только специфику его содержания и особенности логического структурирования, но и соотнесенность с научной картиной мира, с ментальными структурами хранения знаний, с определенными типами вербализации таких структур лингвистическими средствами[4].

Вместе с тем рассмотрение научного текста как единицы научного дискурса, или единицы научной коммуникации, требует привлечения к его анализу положений коммуникативной лингвистики.

Исходя из этого, в последнее время в лингвистической литературе получила широкое распространение точка зрения, в соответствии с которой решение проблем порождения и понимания письменного научного текста необходимо искать с позиций когнитивно-коммуникативной (или когнитивно-дискурсивной) парадигмы, разрабатываемой школой Е.С. Кубряковой. Когнитивно-дискурсивная парадигма не просто признает двумя главными функциями языка когнитивную и коммуникативную, но и преследует цель «изучать эти функции в постоянном взаимодействии и согласовании друг с другом»[5].

Когнитивная составляющая данной парадигмы позволяет анализировать типы знаний/информации, вербализуемых в научном тексте, и стоящие за ними ментальные единицы и структуры, в то время как дискурсивная составляющая позволяет выявить способы представления информации адресату с учетом прагматической направленности текста, интенций автора и особого контекста коммуникативного акта[6].

Данное исследование выполнено в русле лингвокогнитивного подхода, который мы понимаем как часть когнитивно-дискурсивного подхода, его первый этап, включающий две составляющих. Лингвистическая составляющая предполагает изучение функционирования элементов текста, а именно языковых единиц и категорий, вербализующих в условиях текста ментальные единицы и структуры. Когнитивная составляющая позволяет выявить эти ментальные единицы и структуры и представить их в виде онтологий той или иной предметной области.

Целью настоящей работы является изучение лексического уровня корпуса научных текстов английского подъязыка «Программирование» и разработка принципов построения онтологии предметной области «Программирование» на базе терминологического словаря. Мы понимаем «подъязык» как часть естественного языка, описывающую определенную предметную область, не имеющую лексико-грамматических ограничений (кроме тех, которые заданы тематически однородной областью функционирования языка), и ограничений, накладываемых ситуацией общения, в частности участниками коммуникации, которая может искусственно моделироваться для облегчения процесса общения и обучения иностранному языку для специальных целей[7].

Задачи исследования включают составление частотного словаря английского подъязыка «Программирование» на базе корпуса текстов, определение лексического минимума, его стратификацию, выявление терминологического слоя и разработку принципов построения онтологии предметной области «Программирование».

В работе применяются методы корпусной и математической лингвистики, компонентного, контекстологического, морфолого-семантического, сопоставительного анализа, а также метод экспертных судей.

Как известно, эффективность порождения и понимания научного текста на иностранном языке вторичной языковой личностью в значительной степени зависит от ее коммуникативной компетенции, включающей лингвистическую, социолингвистическую и прагматическую составляющие[8].

В рамках лингвистической компетенции выделяются лексическая, грамматическая, семантическая и фонологическая компетенции.

Под лексической компетенцией, в частности, понимается знание словарного состава языка и способность использования его в речи.

Разработаны специальные шкалы, которые иллюстрируют различные уровни знания словарного состава языка и способностей его использования. Так, уровням самостоятельного и свободного владения языком, то есть уровням B2, C1 и C2, соответствует «хороший» и «обширный» словарный запас, в том числе «по профессиональной/представляющей интерес тематике и на общие

темы»[9]. Вопрос заключается в том, что считать «хорошим» и «обширным» словарным запасом.

Этот вопрос тесно связан с методикой обучения иностранным языкам, а именно с содержанием обучения, с задачей отбора минимального объема лингвистических единиц, обладающих наибольшей употребительностью и информативностью и обеспечивающих наиболее высокий уровень понимания специального текста на иностранном языке при обучении чтению и аудированию и наиболее высокий уровень его порождения при обучении письму и говорению.

Одним из возможных решений этой задачи является создание частотного словаря определенного подъязыка[10]. Частотные словари составляются на основе надежной статистической процедуры и могут служить базой для учебного лексического минимума.

В 1990-е годы под руководством одного из авторов данной работы были созданы 10 частотных словарей различных подъязыков, таких как радиоэлектроника (англ.яз.), электроника (англ.яз), нелинейная оптика (англ.яз), сварка (англ.яз.), металлургия (англ.яз., нем.яз.), энергетика (англ.яз., нем. яз.), строительство (нем.яз), приборостроение (фр.яз.)[11].

Все частотные словари были составлены по единой методике, суть которой заключалась в следующем. Основой частотного словаря являлась тематическая структура специальности, составленная экспертами со специальных кафедр. В соответствии с тематической структурой по рекомендации экспертов осуществлялся подбор текстов из современных английских и американских научных журналов и монографий по специальности. Общий объем текстов по одной специальности составил 50 тыс. словоупотреблений[12]. Тексты для анализа были подобраны по специально разработанной экспертами тематической схеме дозировки текстов в соответствии с их относительной значимостью. Пример тематической структуры и схемы дозировки текстов для составления частотного словаря английского подъязыка электроники приводится в табл. 1.

Дальнейшая работа по созданию частотных словарей предусматривала введение подобранных текстов в ЭВМ и их машинную обработку. Предварительно все тексты были подвергнуты лингвостатистической обработке: были исключены формулы, цифры, имена собственные, сокращения, химические знаки. Весь лексический массив был подвергнут лемматизации, т.е. все слова вводились в исходной форме, кроме слов, образованных суплетивно. Таким образом, за единицу частотного словаря принималась лексема. Как известно, лексема как единица отбора значительно повышает покрываемость текста, так как вбирает все словоформы частотного словаря с тем же корнем[13].

В результате машинной обработки текстов на базе программ для ЭВМ ЕС-1022, разработанных группой «Статистика речи» под руководством Р.Г. Пиотровского, для каждой специальности были получены четыре списка: 1) прямой частотно-алфавитный список лексем до частоты 1; 2) прямой алфавитно-частотный вариант этого списка; 3) обратный частотно-алфавитный список лексем; 4) обратный алфавитно-частотный вариант этого списка.

**Таблица 1**

**Тематическая структура частотного словаря английского подъязыка электроники (1991)[14]**

| № п.п. | Раздел | Кол-во словоупотреблений | Объем, % |
|---|---|---|---|
| 1 | История вычислительной техники | 2 000 | 4 |
| 2 | Вычислительные машины | 25 000 | 50 |
| 2.1 | Классификация ЭВМ | 2 000 | |
| 2.2 | Представление информации | 6 000 | |
| 2.3 | Архитектура ЭВМ | 13 000 | |
| 2.4 | Функции ЭВМ | 4 000 | |
| 3 | Программирование | 20 000 | 40 |
| 3.1 | Технология программирования | 4 000 | |
| 3.2 | Типы программ | 6 000 | |
| 3.3 | Языки программирования | 10 000 | |
| 4 | Применение вычислительной техники | 3 000 | 6 |
| 5 | Общее количество | 50 000 | 100 |

Длина списков для каждой специальности варьировалась в следующих пределах: 2755 лексем (английский подъязык нелинейной оптики), 2871 лексема (английский подъязык радиоэлектроники), 3336 лексем (английский подъязык металлургии), 3500 лексем (английский подъязык сварки), 3825 лексем (английский подъязык электроники), 4003 лексемы (немецкий подъязык металлургии), 4172 лексемы (немецкий подъязык энергетики), 4355 лексем (английский подъязык энергетики), 4568 лексем (немецкий подъязык строительства), 669 лексем (французский подъязык приборостроения).

Разницу в объеме можно объяснить различиями в предметных онтологиях исследованных областей знания, а также морфологическими особенностями языков.

При составлении учебных лексических минимумов была принята точка зрения, согласно которой общее понимание текста достигается, если его покрываемость учебным словарем составляет более 85 %[15]. В данном случае мы имеем в виду понимание текста на уровне лексических единиц.

Естественно, что понимание представляет собой «сложный многоэтапный процесс, включающий перцептивно-когнитивно-аффективную переработку воспринимаемого активным и пристрастным субъектом соответствующей деятельности и требующий взаимодействия разных видов знаний: языковых и энциклопедических, явно данных в тексте и выводных, осознаваемых и учитываемых без их вывода на «табло сознания»[16]. Вместе с тем нельзя не признать, что слово играет особую роль в понимании текста. А.А. Залевская образно сравнивает слово с лазерным лучом при считывании голограммы: оно делает доступным для человека определенный условно-дискретный фрагмент континуальной и многомерной индивидуальной картины мира во всем богатстве связей и отношений, полнота которых обеспечивается в разной мере осознаваемой опорой на выводные знания и переживания разных видов, одновременно слово рассматривается как «средство доступа к единому информационному тезаурусу человека»[17]. Поэтому исследование лексического уровня, специфики его единиц, особенностей их организации и т.д. является весьма важным в изучении научного текста и проблем его порождения и понимания.

Объем учебного лексического минимума устанавливался экспериментально методом проверки покрываемости текстов.

Анализ полученных результатов показал, что оптимальным частотным списком, обеспечивающим покрываемость текстов выше 85 %, для большинства исследованных специальностей является частотный список на 800 лексем. Для специальности ЭВМ длина такого списка составила 842 единицы. Этот список и был принят в качестве учебного лексического минимума соответствующего подъязыка.

Главной целью лексического минимума, как указывалось выше, является минимизация и оптимизация учебного материала, что проявляется в отборе и стратификации основного лексического ядра, необходимого для практического овладения языком научных текстов по электронике.

Минимизация алфавитно-частотного словаря-минимума по электронике объемом 842 л.е. проводилась методом выделения лексических гнезд и составления учебного словаря-минимума для чтения текстов по электронике. В результате отбора и стратификации лексики учебный лексический минимум имел следующую структуру: словарь-минимум, список терминов, список строевой лексики, список интернациональных слов, списки синонимов, антонимов, омонимов, список звукографических интерферем. В качестве приложения были составлены таблицы суффиксальных и префиксальных словообразовательных моделей существительных, прилагательных, глаголов и наречий.

Как показал опыт практической работы, учебные лексические минимумы, составленные по рассмотренной методике, стали обязательным компонентом учебных методических комплексов по специальности, с их помощью у студентов и аспирантов формируется необходимая лексическая компетенция, при этом освоение лексики научных текстов по специальности проходит эффективно и в оптимальном объеме.

Следует отметить, что наука не стоит на месте. Это в равной мере касается как лингвистики, так и других наук. Меняется состав предметных областей, а вместе с ним меняется и лексический состав соответствующего подъязыка. Частотные словари и учебные лексические минимумы требуют обновления. Достижения вычислительной и корпусной лингвистики, новые информационные технологии позволяют добиться в этом оптимальных результатов.

В эпоху информационного взрыва одной из самых активно развивающихся областей научного знания является «Программирование». Программирование представляет собой «процесс и искусство создания компьютерных программ и/или программного обеспечения с помощью языков программирования»[18]. Специальность программиста является в настоящее время одной из самых востребованных на рынке труда. Более того, каждый пользователь ЭВМ должен быть в определенной степени знаком с терминами подъязыка программирования для работы с компьютером. О глобализации английского подъязыка программирования свидетельствует ассимиляция английской терминологии в среде профессионалов разных стран. Все это делает актуальным исследование научных текстов английского подъязыка программирования, в том числе его лексического уровня.

Для исследования лексического уровня научных текстов английского подъязыка «Программирование» был собран корпус текстов, включающий 47 статей из электронных версий четырнадцати современных научных журналов по программированию на английском языке таких, как *Artificial Intelligence; Computer Languages, Systems and Structures; Computer Speech and Language; Data & Knowledge Engineering; Parallel Computing; Science of Computer Programming; Theoretical Computer Science; The Computer Journal of Parallel and Distributed Programming* и других за 2007–2008 гг., общим объемом около 500 тыс. словоупотреблений.

Анализ лексического уровня текстов по программированию проводился с использованием следующего программного обеспечения:

- приложение *SMAT* для построения частотных списков n-грамм для корпуса текстов[19];
- программный комплекс *Ling Assistant* для построения спектрального распределения частот, предварительного разбиения по частям речи и сортировки списков лексических единиц[20];
- приложение *Statistics* – одно из приложений комплекса *Ling Assistant* для сортировки частотных списков и составления спектрального распределения частот в частотных словарях;

• словарные базы электронных словарей *Multitran, MacMillan* и *Lingvo 10.0* с приложениями *LingvoComputer* и *Computers* (для терминов), LingvoScience (для общенаучной лексики) и *LingvoUniversal* (для общеупотребительной лексики).

На основе частотного анализа корпуса текстов по программированию с помощью программы *SMAT* были получены 4 словаря: 1) прямой частотно-алфавитный; 2) обратный частотно-алфавитный; 3) прямой алфавитно-частотный; 4) обратный алфавитно-частотный. Длина словаря составила 17940 словоформ. Лемматизация словарного состава до ввода корпуса текстов в ЭВМ не проводилась.

Чтобы сделать словарь пригодным для анализа, мы пришли к выводу о необходимости его минимизации. Оптимальным вариантом минимизации, как указывалось выше, является получение оптимального списка словоформ, покрывающих 85 % словоупотреблений всех текстов, то есть списка, пригодного для использования в прикладных целях.

В результате обработки указанного корпуса текстов программой *Statistics* были отобраны наиболее частотные однословные словоформы, покрывающие 85 % словоупотреблений текстов, что составило 1761 словоформу с частотой от 33958 (определенный артикль *the*) до 33 (54 слова, среди которых такие, как *indicate, reset, subgraph, packet, geometrical* и т.д.) (табл. 2). Такой объем мы сочли достаточным для анализа, поскольку, как мы уже упоминали, именно покрываемость текста не ниже 85 % словоупотреблений свидетельствует о том, что понимание текста на лексическом уровне достигнуто.

**Таблица 2**

**Спектральный анализ покрываемости корпуса текстов частотным словарем английского подъязыка «Программирование» (2009)**

| Частота | Кол-во слов | Процент от общего кол-ва слов | Суммарный процент |
|---|---|---|---|
| 33958 | 1 | 7 | 7,011 |
| … | … | … | … |
| 34 | 25 | 0,18 | 84,87 |
| 33 | 54 | 0,37 | 85,24 |
| 32 | 29 | 0,26 | 85,5 |
| … | … | … | … |
| 1 | 7566 | 1,6 | 100 |

Значительное превышение длины словаря-минимума по сравнению с нашими предыдущими исследованиями объясняется, в первую очередь, тем, что за прошедшие со времени создания словаря по электронике 15–20 лет предметная область «Программирование» значительно расширила свои рамки, что, безусловно, не могло не привести к расширению словарного состава. Во-вторых, объем выборки в десять раз превысил объемы предыдущих наших выборок, что также способствовало увеличению длины словаря. Кроме того, при создании нового словаря не проводилось предварительной лемматизации, а также не исключались формулы, цифры, имена собственные и сокращения.

Поэтому следующим шагом при составлении лексического минимума по программированию стала минимизация словарного состава, а именно исключение из словаря цифр и формул. Имена собственные из словаря не исключались, поскольку в большинстве случаев относились к понятиям данной предметной области, например, *Boolean, Java, Petri,* и использовались в терминологических словосочетаниях *Boolean algebra/ expression/ logic, Java class/ compiler/ language, Petri net* и других. Сокращения присоединялись к полным словам, а в случае отсутствия последних в словаре сохранялись в качестве словарных единиц, например, *fig → figure, fig; ref → reference, ref*, но *etc – et cetera; sm – shared memory; uml – unified modeling language* и т.д.

Далее была проведена лемматизация словарного состава, то есть слова приводились к исходной форме, например, *takes, taking, taken → take; expressions → expression; easymap's → easymap* и т.п.

После этих процедур длина словаря-минимума составила 1178 лексем, что вполне сопоставимо с длиной в 842 единицы словаря-минимума по электронике (1991).

Следующим этапом работы стала стратификация словаря-минимума. Как известно, лексика языка науки неоднородна и состоит из трех слоев: общеупотребительного, общенаучного и терминологического[21].

Общеупотребительная лексика английского языка составляет нейтральную ткань повествования. Она представлена служебными словами (артиклями, предлогами, союзами), словами общелитературного языка, для которых типично употребление в различных функциональных стилях, и, как правило, не является предметом специального рассмотрения в языке науки.

Общенаучная лексика отражает научные понятия, соотносящиеся с объектами, процессами и явлениями в различных областях научного знания, способствует логическому и последовательному изложению материала и составляет основу научного стиля изложения.

Терминологическая лексика функционирует в научных текстах определенной научной области и отражает специфику последней. Под специальным термином понимается слово или словосочетание специального языка, создаваемое для точного выражения специальных понятий и обозначения специальных предметов.

В словарном составе нашего алфавитно-частотного словаря-минимума по программированию мы выделили три указанных слоя: общеупотребительную лексику, общенаучную лексику и терминологическую, или специальную лексику. Отнесение словарных единиц к тому или иному слою проводилось с помощью приложения *SMAT* методом сопоставления полученного нами словаря

на 1178 л.е. с соответствующими авторитетными электронными словарями. Для этой процедуры использовались словарные базы электронных словарей *Multitran, MacMillan* и *Lingvo 10.0* с приложениями *LingvoComputer* и *Computers* (для терминов), *LingvoScience* (для общенаучной лексики) и *LingvoUniversal* (для общеупотребительной лексики).

Для описания стратификационной структуры лексики английского подъязыка «Программирование» мы ввели следующие лингвистические переменные: T (термины), S (общенаучная лексика) и G (общеупотребительная лексика). Если слово встречалось в словаре терминов, оно получало помету T, в словаре общенаучной лексики – помету S, в словаре общеупотребительной лексики – помету G. Если слово встречалось в двух словарях, то оно получало соответствующие пометы GS, TG, TS, встречаемость слова сразу в трех словарях помечалась TGS. Если слово встречалось только в одном словаре, оно получало помету этого словаря с индексом 1: T1, S1, G1. Фрагмент росписи алфавитно-частотного словаря приводится в табл. 3.

Результаты стратификационного анализа приводятся в табл. 4.

Таким образом, в результате стратификационного анализа были получены следующие данные: общеупотребительный слой составил 1127 лексических единиц, общенаучный – 973 единицы, терминологический – 773 единицы.

Сопоставление словаря-минимума (2009) с рассмотренным выше словарем-минимумом для специальности ЭВМ (1991)[22] показало их совпадение на 35,75 %, а совпадение терминов на 30,42 %, при этом процентное содержание терминологической лексики оказалось практически одинаковым (табл. 5).

Достаточно низкий процент совпадений связан, в первую очередь, с различиями в тематической структуре словарей. Как указывалось выше, тексты по программированию составляли лишь 40 % в корпусе текстов, использованных для составления словаря 1991 года (см. табл. 1), в то время как для составления словаря 2009 года использовались 100 % текстов по программирова-

**Таблица 3**

**Фрагмент росписи алфавитно-частотного словаря-минимума английского подъязыка «Программирование» (2009)**

| Словарь/ л.е.{частота} | Lingvo | | | Multitran | | | MacMillan | | |
|---|---|---|---|---|---|---|---|---|---|
| | Term | General | Science | Term | General | Science | Term | General | Science |
| absolute {47} | | G | | | G | | | G | |
| abstract {119} | T | G | | T | G | S | | G | |
| abstraction {89} | T | G | S | T | G | S | | G | |
| access{229} | T | G | S | T | G | S | T | G | |
| accessor{65} | T | | | T | G | S | T | G | |

**Таблица 4**

**Стратификационный анализ лексики словаря-минимума английского подъязыка «Программирование» (2009)**

| Слой | G1 | S1 | T1 | «Плавающая лексика» | | | | G | S | T |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | GS | TG | TS | TGS | | | |
| Кол-во л.е. | 125 | 7 | 31 | 273 | 49 | 13 | 680 | 1127 | 973 | 773 |
| % от 1178 | 10,6 | 0,6 | 2,6 | 86,2 (1015 л.е.) | | | | 95,7 | 82,6 | 65,6 |

**Таблица 5**

**Сопоставительный анализ лексики частотных словарей по ЭВМ (1991) и программированию (2009)**

| Лексика | Кол-во л.е. | | | Совпадение | | Несовпадение | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Есть только в 1991 г. | | Есть только в 2009 г. | | Всего отличающихся единиц | |
| | 1991 | 2009 | Всего разных единиц | Абс. | % | Абс. | % | Абс. | % | Абс. | % |
| Вся лексика | 842 | 1178 | 1488 | 532 | 35,75 | 310 | 20,83 | 646 | 43,41 | 956 | 64,25 |
| Термины | 556[23] /66,0 % | 773 /65,6 % | 1019 /68,5 % | 310 | 30,42 | 246 | 24,14 | 463 | 45,44 | 709 | 69,58 |

нию. Низкий процент совпадений можно также объяснить и изменениями в самой предметной области «Программирование». Так, при составлении словаря 1991 года тематическая структура области «Программирование» по мнению экспертов включала три раздела (см. табл. 1). В 2007 году по данным Ассоциации вычислительной техники классификация предметной области «Программирование» включала 5 разделов и более 300 подразделов[24]. Это еще раз подчеркивает необходимость проведения работы по созданию нового словаря английского подъязыка программирования как для научных, так и прикладных целей.

Как следует из табл. 5, список терминов лексического минимума по электронике (1991), который насчитывал 496 л.е. (58,9 % словаря), после проверки с помощью современных лексикографических ресурсов увеличился на 60 л.е. и составил 556 л.е., или 66,0 % словаря. Подавляющее большинство терминов (404 л.е., или 72,7 %) составляли имена существительные. Этот показатель совпадает с результатами исследований, согласно которым терминироваться могут все части речи, однако, основная доля терминов приходится на существительные[25].

*Список терминов* нового лексического минимума по программированию (2009) насчитывает 773 лексических единицы, что составляет 65,6 % словаря. Например, *bit, byte, cache, compiler, cascade, descriptor, location* и т.д. Имена существительные составляют в этом списке 69,5 % (537 л.е. с учетом конверсионных пар, например, *copy, code, frame, index* и др.).

*Общеупотребительная* лексика составляет 95,7 % словаря (1127 л.е.) и довольно равномерно распределена в частотных списках, что свидетельствует о значении общеупотребительной лексики, в состав которой входит строевая лексика, как для порождения, так и понимания иноязычного текста. Например, *a(an), the, do, have, can, may, make, any, all, describe, different, end* и т.д.

*Общенаучная* лексика представлена 973 словами, что составляет 82,6 % лексического минимума. К общенаучной лексике относятся слова, обозначающие стадии научного познания, методы и приемы, инструментарий, математические термины и т.д. Например, *axiom, derivation, evolution, lemma, modification, principle* и т.д.

При стратификации лексического минимума выяснилось, что отнесение слова к какому-либо одному из слоев лексики возможно в относительно небольшом количестве случаев (163 л.е., или 13,8 %), в то время как подавляющее большинство слов встречались сразу в нескольких словарях, то есть относились разными авторами к общеупотребительной, общенаучной и терминологической лексике, или к каким-либо двум из перечисленных слоев (см. табл. 4). В нашем материале такой «плавающей» лексики встретилось 1015 л.е., или 86,2 %, из них 680 л.е., или 57,7 % встретились во всех трех словарях. Например, лексема *heuristics* представлена в словаре терминологической и общенаучной лексики, лексема *compiler* представлена в словаре терминологической и общеупотребительной лексики, лексема *reflective* встречается в словаре общенаучной и общеупотребительной лексики, лексема *restriction* представлена во всех трех словарях. Этот факт можно объяснить, с одной стороны, ролью субъективного фактора, то есть фактора автора словаря, его фоновых знаний, учета мнения экспертов и т.д. С другой стороны, в корпусе текстов слова, составляющие лексический минимум, объективно обладают многозначностью и в зависимости от контекста могут относиться то к одному, то к другому слою.

О явлении так называемой «двойной детерминации» или «консубстанциональности» терминов и слов общего языка писали известные лингвисты Р.А. Будагов, О.С. Ахманова, А.И. Комарова[26] и др. По мнению указанных лингвистов, различие в объеме информации и степени абстрактности понятия, характерное для разных употреблений одного слова, ставит их на грань омонимии и делает несопоставимыми в содержательном и функционально-стилистическом плане. Поэтому следует учитывать эту «двойную» (в нашем случае «тройную»!) детерминированность» слов, которые потенциально могут выступать и в роли термина, и в роли слова общего языка. По мнению А.И. Комаровой, которое мы полностью разделяем, вопрос о принадлежности того или иного слова к разряду терминов или элементов общего языка может быть решен в зависимости от контекста.

Такая ситуация нередка в естественных языках, так как они постоянно меняются, адаптируясь для нужд коммуникации. В том числе, это свойственно подъязыкам ограниченных предметных областей. Формальным способом описания такого явления является теория нечетких множеств[27]. Как указывалось выше, для описания стратификационной структуры лексики предметной области мы ввели лингвистические переменные T (термины), S (общенаучная лексика) и G (общеупотребительная лексика). Исходя из этого, в рамках теории нечетких множеств каждая лексическая единица может принадлежать одному (или более) из этих множеств, причем каждому из них с разной вероятностью (сумма вероятностей для каждой лексемы равна единице), которая определяется по относительным частотам соответствующих значений данной лексической единицы в корпусе текстов, что может быть отражено с помощью следующей записи:

$L = \{p_1/x_1, p_2/x_2, \ldots, p_n/x_n\}$,

где $L \in \{T, S, G\}$ – лингвистическая переменная,

$x_i$ – i-я лексема лексического минимума;

$p_i = m/N$,

где m – частота употреблений лексемы $x_i$ как элемента множества L; N – суммарная абсолютная частота лексемы $x_i$ в корпусе текстов.

Так, фрагмент описаний лингвистических переменных T, S и G, составленный на основе анализа распределений частот словоупотреблений неоднозначных единиц в проанализированном корпусе текстов, имеет вид:

T = {...0,02/modal, 0,78/reference, 0,15/role, ...},
S = {..., 0,98/modal, 0,07/reference, 0,38/role,...},
G = {..., 0/modal, 0,15/reference, 0,47/role, ...}.

Как следует из вышеизложенного, ряд лексических единиц относится к нескольким множествам, причем вероятности их принадлежности распределены различным образом. Имея такую картину, можно определить лексический терминологический минимум, применив операцию сгущения нечетких множеств, которая представляет собой формирование четких множеств из нечетких таким образом, что в случае, когда лексическая единица принадлежит нечеткому множеству с вероятностью, превышающей некоторое пороговое значение, вероятность для данной единицы округляется до 1; если вероятность единицы ниже порогового значения, то она округляется до 0[28]. Результатом является обычное множество, в которое включены только те лексические единицы, которые достаточно часто фигурирует в исследовательском корпусе в соответствующей роли. Так, приняв за пороговое значение величину 0,33, приведенные ранее фрагменты нечетких множеств T, S, G преобразуются в следующие сгущения:

T' = {reference},
S' = {modal, role},
G' = {role}.

Таким образом, для проанализированного корпуса текстов, несмотря на то, что практически все перечисленные слова фигурируют в каждом из трех множеств (за исключением modal, которое в множестве G имеет вероятность 0), наиболее типичным представителем, например, группы терминов является лишь слово reference, остальные лексические единицы как термины употребляются редко.

Уточнение списка терминов в соответствии с правилами, представленными в теории нечетких множеств, требует детальной работы со всем корпусом текстов и привлечения экспертов для маркировки терминов.

Дальнейшая обработка полученного словаря–минимума на 1178 лексем для целей обучения профессиональной иноязычной коммуникации проводилась вручную методом «выделения лексических гнезд». Для этих целей были использованы прямой и обратный алфавитно-частотный словари, на основании обработки которых словарь-минимум был представлен в виде словаря лексических гнезд. В данной работе принято широкое толкование гнезда как образования, включающего любые слова, содержащие один корень и близкие по семантике. Все элементы гнезда расположены в алфавитном порядке, при этом головным является слово – производящая основа, а в случае его отсутствия – первое по алфавиту слово. Например, к гнезду *base* отнесены лексемы *base (n,v), basic, basis, database;* гнездо *compute* включает лексемы *computable, computation, computational, compute, computer, pc=personal computer, computepath*; гнездо *use* составлено из лексем *use(n,v), usage, useful, user, reuse(v).*

После такой обработки словарь-минимум стал насчитывать 867 словарных гнезд, что является вполне приемлемым для целей обучения профессиональной иноязычной коммуникации.

Недостатком полученного словаря на данном этапе работы мы считаем то, что он включает однословные лексические единицы. Как известно, научный текст дает слово в типичном для него окружении в составе полилексемных единиц – предельных синтагматических последовательностей[29], являющихся функциональными эквивалентами слова. Выявление таких последовательностей станет следующим этапом работы по составлению лексического минимума по программированию.

Исследование лексического состава научных текстов английского подъязыка «Программирование» привело нас к мысли о возможности создания современной предметной онтологии «Программирование» на базе терминологического слоя полученного частотного словаря-минимума.

Актуальность разработки интеллектуального ресурса, целью которого является моделирование знаний в предметной области «Программирование» не вызывает сомнений. Как известно, в условиях бурного роста информационных технологий рассматриваемая предметная область является одной из наиболее динамично развивающихся, что делает такое исследование весьма своевременным и необходимым.

Современная прикладная лингвистика предъявляет высокие требования к «интеллектуальности» систем автоматической переработки текста. Соответствие системы сформулированному еще на заре развития компьютерной лингвистики и систем искусственного интеллекта тесту Тьюринга (интеллектуальная система – это такая система, в диалоге с которой человек думает, что общается с другим человеком, а не с машиной) остается не до конца решенной задачей. После ряда попыток создания интеллектуальных систем посредством тщательного описания грамматических и лексических элементов некоторого языка, стало очевидно, что лингвистического знания для принятия «разумных» решений недостаточно. Машинам не хватает экстралингвистических знаний, которыми при нормальных условиях обладает человек.

В связи с этим в современные интеллектуальные системы все чаще включают компонент, называемый *онтология,* который является одним из наиболее перспективных способов моделирования знаний. Онтология – это формальная спецификация концептуализации предметов и явлений действительности. Современные онтологии строятся как для представления общей картины мира, так и

ориентированные на определенную предметную область (ПрО). Так, ряд продолжающихся и завершенных проектов в области конструирования онтологий направлен на представление общего знания (Mikrokosmos, OntoQuery и др.). Однако, как правило, современные онтологии ориентированы на определенную ПрО, например, АвиаОнтология[30], GOLD (General Ontology for Linguistic Description)[31], онтология науки и онтология компьютерной лингвистики[32].

Одновременно с развитием общеконцептуальной структуры ПрО ведется развитие лексической базы, ассоциируемой с онтологией, которая выполняется в форме словаря и правил грамматики некоторого языка (например, датская грамматика и лексика в проекте OntoQuery) либо в форме тезаурусов, сходных по структуре с онтологиями, но представляющих лексический профиль ПрО (например, РуТез[33], WordNet[34]).

Построение онтологии ПрО является нетривиальной задачей, так как не существует источника, явно описывающего состав, содержание и структуру концептов ПрО, их взаимосвязи. Моделирование онтологических ресурсов требует извлечения экстралингвистического знания из «косвенных» свидетельств (мнения экспертов, описания понятий в энциклопедических словарях). При этом способ интерпретации полученных знаний и в результате построение онтологии зависят как от субъективной точки зрения исследователя, так и, видимо, частично может определяться структурой ПрО[35].

Как правило, онтологический ресурс интегрирует знания (концептуальное представление ПрО) и данные (терминологический словарь или тезаурус, описывающий способы репрезентации знаний в языке). Онтология, интегрированная со словарем, может иметь ряд применений:

- использование в целях обучения для знакомства с ПрО, построения учебных планов и программ и т.д.;
- интегрирование в системы поиска профессиональной информации для увеличения его полноты и разрешения лексической многозначности[36];
- средство классификации научных исследований в данной ПрО;
- источник лингвистического знания при описании результатов научных изысканий в данной ПрО.

Таким образом, конструирование онтологий ПрО представляет интерес как для процедур интеллектуальной автоматической обработки текста, так и для использования в качестве энциклопедического ресурса человеком. Онтология при решении указанных задач, актуальных для любой научной сферы деятельности, является источником унифицированного семантического знания.

При проектировании онтологии необходимо, чтобы она удовлетворяла требованиям, вытекающим из возможных областей применения ресурса. Для возможности вывода на знаниях с целью ознакомиться с ПрО, а также применения для расширения полноты поиска онтология должна иметь иерархическую структуру и представлять собой связный граф. Возможность интеллектуального поиска с помощью концептов онтологии предполагает формализацию концептообразующих признаков, являющихся определяющими для сущности понятия. При этом для решения всех задач онтология должна предусматривать оценку значимости одних концептов по отношению к другим.

Модель знаний для ПрО «Программирование», удовлетворяющая этим требованиям, пока не нашла отражения в форме онтологий. Отличительным свойством предлагаемого проекта онтологии является построение структуры понятий, которая позволяет определить не только связи концепта с другими понятиями ПрО, но и связи различной степени значимости между концептами онтологии, не включающие рассматриваемое понятие, но необходимые для его понимания. Для выявления этих связей предполагается привлечение разноплановых источников – консультации экспертов, анализ энциклопедических данных, а также обращение к языковым эквивалентам на разных языках с целью выявления границ понятий.

Концепты онтологии репрезентируются через набор характеристик (отношений заданного типа) и значений этих характеристик. Такая онтология может быть задана в форме реляционной базы данных, каждая запись которой представляет собой бинарное типизированное отношение между концептом и значением некоторой его характеристики. Концепт онтологии задается множеством таких пар.

Предлагаемая нами модель построения онтологии базируется на тех же принципах. Для задания концептов логичным представляется введение пустого концепта верхнего уровня онтологии TOP и его потомков OBJECT, PROPERTY, EVENT[37], а также RELATIONS для определения классов отношений.

Инвентарь характеристик каждого концепта формируется на основе анализа ПрО. Для сохранения критерия иерархичности онтологии характеристики включают отношения меронимии и гипонимии. Для определения связи концептов в типичных ситуациях в качестве характеристик описывается инструментарий семантических ролей (*Agent, Source, Result, Instrument* и т.д.). В ходе исследования планируется определить инвентарь ассоциативных связей (как, например, онтологическая антонимия) с помощью анализа ПрО и выявления отношений между ее объектами. Значения характеристик представляют собой константы (в случае числовых параметров) или другие концепты онтологии.

В отличие от существующих онтологий, с целью увеличения семантической силы содержательного описания концепта предлагается рассматривать иерархические связи не только между

концептами, но и «внутри» описания концепта – иерархию характеристик. Характеристики объединяются в группы по описываемому аспекту понятия. На данный момент нами выделяются группы *Definition* (включает представление отношений с основными понятиями, необходимыми для однозначного понимания концепта), *Hierarchy* (гипогиперонимические отношения, средним звеном которых является данный концепт), *Case-roles* (связь с объектами онтологии, типичными при заполнении некоторых семантических валентностей, задаваемые в форме ссылок на концепты онтологии или их характеристики), *Sem-paradigm* (ассоциативные отношения между концептами).

Далее, внутри этих групп, характеристики задаются как бинарные отношения, где вторым участником являются концепты онтологии, а первым участником является либо описываемый концепт, либо концепт, включенный ранее в описание как второй участник одного из отношений. Таким образом, каждый фрейм описания может содержать подфреймы, характеризующие концепт, заполняющий соответствующий слот. Например, фрагмент представления понятия PROGRAMMING в такой онтологии в логической форме имеет вид:

*Name*: PROGRAMMING
*Hierarchy*: [
is_a: EVENT
kind_of: RULE-ORIENTED PROGRAMMING
kind_of: LOW-LEVEL PROGRAMMING
kind_of: PARALLEL PROGRAMMING
kind_of: PROCEDURAL PROGRAMMING
kind_of: FUNCTIONAL PROGRAMMING
kind_of: EVENT-DRIVEN PROGRAMMING
kind_of: GENETIC PROGRAMMING
kind_of: KNOWLEDGE-BASED PROGRAMMING ]
*Definition*: [
is_a: EVENT
purpose: CREATE
dir-obj: PROGRAM ]
*Case-roles*: [
agent: PROGRAMMER
source: ALGORITHM
result: PROGRAM
result: SOFTWARE
instrument: PROGRAMMING LANGUAGE
instrument: PROGRAMMING TOOL ]
*Sem-paradigm*: [
assoc: FLOWCHART
has_part: STEP
has_part: COMPILATION
has_part: DEBUGGING ]

В таком представлении фреймы *Hierarchy, Case-roles, Sem-paradigm, Definition* оперируют другими понятиями онтологии, устанавливая их связь с рассматриваемым концептом. Графически фрагмент концептуальной структуры понятия PROGRAMMING представлен на рис. 1.

Традиционное определение понятия строится как перечисление родового и множества отличительных признаков. Поэтому на схеме в раздел *Definition* также включено отношение is_a, что приводит к наложению двух зон. Однако при представлении в логической форме (как можно видеть выше для данного концепта) родовой признак однозначно относится к разделу *Hierarchy*. Такое несоответствие легко преодолеть путем определения способа обхода графа при формировании определения концепта, задав правило *Определение = Hierarchy:is_a(1)* $\cup$ *Definition:ALL* (определение значения понятия задается как объединение первого из упомянутых отношений is_a в группе *Hierarchy* и всех признаков группы *Definition*).

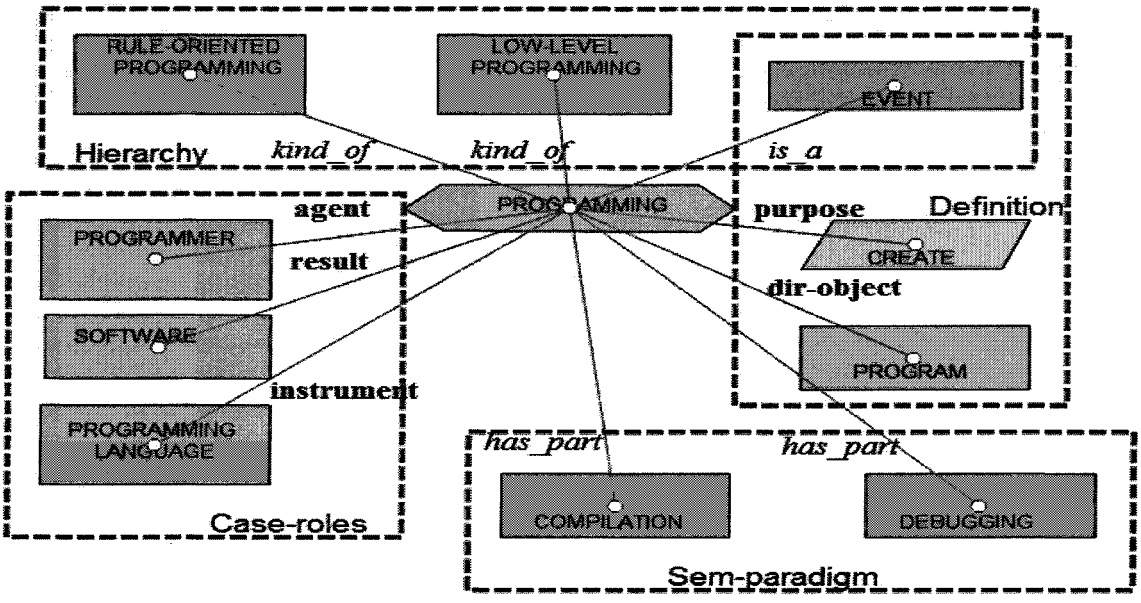


**Рис. 1. Графическая репрезентация структуры концепта PROGRAMMING**

Концепт, заполняющий слот описания, определяет потенциальный список отношений, посредством которых данный концепт далее может детализироваться. Так, например, для потомков класса EVENT в качестве потенциальных полей подфрейма будет выступать список семантических ролей, описанных в онтологии в статье соответствующего события.

Каждый объект, наряду с онтологическими отношениями, должен задаваться посредством набора лексических отношений (синонимия – задает список терминов, соответствующих концепту), описанных в зоне *Lexicon*. Зона *Lexicon* включает лингвистические единицы, обозначающие данное понятие в некотором языке (языках). В качестве значений даются ссылки (по заглавному слову) на статью в словарной части онтологии, хранящую лингвистическое описание соответствующих лексических единиц. Например, для описания понятия PROGRAMMING зона Lexicon имеет следующий вид:

*Lexicon*: [
programming
coding
soft-writing ]

Построение онтологии в рамках описанной концепции предполагает выполнение работ в несколько этапов.

1. Формирование списка справочной литературы (источника энциклопедических знаний о ПрО) в сотрудничестве с экспертами. Список должен включать учебники, одноязычные и двуязычные терминологические словари. Двуязычные словари необходимы как контролирующее средство при выявлении содержания и объема понятий (логично предположить, что важные для ПрО концепты должны быть представлены содержательно эквивалентными языковыми терминологическими выражениями в разных языках). Использование терминологических словарей представляется целесообразным для описания концептов предметной области, так как термины лишены коннотативного компонента значения, следовательно, значение термина приближается к соответствующему понятию, отражающему наиболее общие и существенные признаки предмета или явления[38], что соответствует онтологической трактовке концепта.

2. Подбор корпуса научных текстов в ПрО «Программирование» (первоначально на английском языке) для выявления значимых на современном этапе терминов (путем построения частотных списков терминов) и, совместно с изучением источников энциклопедического знания, определения набора концептов, требующих описания в онтологии. Очевидно, что на данном этапе необходимо будет также прибегнуть к алгоритмам извлечения словосочетаний из корпуса.

3. Анализ семантической структуры концептов (с помощью учебников, многоязычных словарей и энциклопедий, подобранных на первом этапе) с целью формирования инвентаря онтологических отношений между понятиями в ПрО (потенциальных полей фрейма концепта) и выявления типов ограничений на заполнение слотов фрейма.

4. Исследование свойств выявленных отношений (транзитивности, правил наследования и т.д.) для формулирования правил вывода на знаниях.

5. Определение спецификации языка для представления онтологии на основе анализа существующих языков (OWL, DAML+OIL). Язык должен интегрировать в себе возможность описания выявленных отношений между концептами, а также обеспечивать реализацию правил вывода.

6. Создание онторедактора для построения онтологии в заданном формализме.

7. Описание концептуальной структуры ПрО с помощью онторедактора.

8. Предъявление макета онтологии экспертам для оценивания соответствия построенной модели представляемой предметной области и определения значимости отношений между концептами в статьях онтологии (путем приписывания весов связям между концептами).

9. Адаптация правил вывода на знаниях с учетом значимости отношений между концептами.

10. Сбор разноязычных корпусов в ПрО «Программирование» для расширения спектра используемых языков, к которым применима онтология. Разработка многоязычной словарной поддержки онтологии.

11. Уточнение структуры представления знаний ПрО.

12. Расширение функций онторедактора для визуализации структуры ПрО и осуществления навигации по ее концептам.

Результатом выполнения работы станет:

- модель представления знаний с использованием мультииерархического описания связей между концептами в форме спецификации языка;
- база данных (БД), репрезентирующая онтологию ПрО «Программирование», включающая онтологическое и лексическое (на материале английского языка) описание концептов по построенной модели;
- онторедактор как инструмент доступа к БД, автоматизации построения и пополнения онтологии. Онторедактор является необходимым компонентом при разработке онтологии, который служит интерфейсом между пользователем, разработчиком и экспертом ПрО, с одной стороны, и базой данных, хранящей онтологию, с другой. Использование онторедакторов (например, InTez[39], OntoGrid[40] и др.) дает возможность автоматизировать ввод знаний в онтологию, а также контролировать согласование вновь вводимых знаний с уже имеющимися. Онторедактор является необходимым средством моделирования (создания и редактирования) онтологии, так как при увеличении числа концептов неминуемы ошибки при ручном построении базы данных. Построение редактора

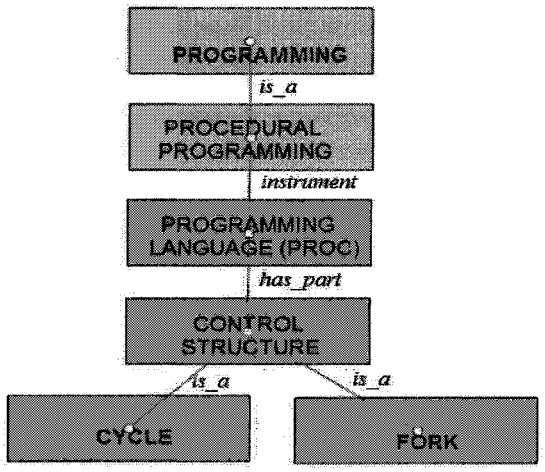


**Рис. 2. Опосредованная связь между концептами**

даст возможность автоматически отслеживать соответствие вводимых данных общей концепции онтологии.

Для построения предметной онтологии терминологический слой лексики английского подъязыка «Программирование», выделенный и минимизированный нами для прикладных нужд до 568 единиц, был подвергнут морфолого-семантическому анализу.

В соответствии с предлагаемой моделью построения онтологии все термины были распределены по их принадлежности к концептам OBJECT, PROPERTY, EVENT, RELATIONS.

Для части терминов с целью выявления концептуальной структуры обозначаемых ими понятий были рассмотрены словарные определения (из англоязычного словаря компьютерной лексики McMillan, англо-русского словаря по программированию и информатике А.Б. Борковского[41], а также англо-русских специализированных компьютерных словарей Lingvo и Multitran) и выведен предварительный инвентарь семантических отношений, репрезентирующих поля фреймов концептов. Примеры таких отношений (*is_a, has_part, agent, result* и т.д.) показаны в зонах *Definition, Hierarchy, Sem-paradigm* и *Case-roles* для концепта PROGRAMMING (см. рис. 1).

Представления, построенные согласно описанной модели на выборке терминов, показали ее применимость для глубокого описания семантической структуры и, как следствие, содержания понятий предметной области «Программирование». Иерархичность описания концепта дает возможность выявить не только отношения концепта, напрямую связывающие его с другими понятиями онтологии, но и выявить наиболее существенные опосредованные связи. Например, через концепт ALGORITHM концепт PROGRAMMING опосредованно связан с такими понятиями как GENERALITY и EFFICIENCY, которые вводятся посредством отношения *property* концепта ALGORITHM. Связь с данными понятиями определяет цель программирования как достижение этих свойств конечным продуктом. Аналогично, опосредованная связь концепта PROGRAMMING с концептами FORK и CYCLE (рис. 2) позволяет детализировать средства программирования и включить в них также соответствующие этим концептам управляющие конструкции.

Построение предметной онтологии на базе частотного терминологического словаря-минимума даст возможность не только глубокого описания семантической структуры и, как следствие, содержания понятий предметной области «Программирование», но и позволит представить такую онтологию в динамическом аспекте, с точки зрения распространения концептов, а вместе с ними и терминов, от ядра концептуальной сферы к ее периферии.

Работа по построению предметной онтологии «Программирование» будет продолжена в соответствии с указанным выше стратегическим направлением исследований с целью способствовать развитию когнитивной лингвистики и дальнейшей «интеллектуализации» систем автоматической переработки текста.

В заключение отметим, что предложенный нами лингвокогнитивный подход к анализу научного текста является интегральным, поскольку позволяет интегрировать не только лингвоцентрический, дискурсивный и когнитивный подходы, но и использовать полученные данные в лингводидактике для нужд профессиональной коммуникации.

---

[1] Тураева З.Я. Лингвистика текста на исходе второго тысячелетия. Віснік Київ лінгв. ун-ту. Серия. «Філологія». 1999. Т. 2. С.17–25; Дроздова Т.В. Научный текст и проблемы его понимания (на материале англоязычных научных экономических текстов): дис. … д-ра филол. наук. М., 2003. С. 13.

[2] Бабенко Л.Г. Васильев И.Е., Казарин Ю.В. Лингвистический анализ художественного текста: учебник для вузов по специальности «Филология». Екатеринбург, 2000. С. 16–30.

[3] Кубрякова Е.С. Начальные этапы становления когнитивизма: лингвистика – психология – когнитивная наука // Вопросы языкознания. 1994. № 4. С. 3.
[4] Дроздова Т.В. Проблемы понимания научного текста (англоязычные экономические тексты). М.; Астрахань: Изд-во АГТУ, 2003. С.16–17.
[5] Кубрякова Е.С. Цит. соч. С. 3.
[6] Дроздова Т.В. Цит. соч. С. 18.
[7] Хомутова Т.Н. Язык для специальных целей (LSP): лингвистический аспект // Известия Российского государственного педагогического университета им. А.И. Герцена. Общественные и гуманитарные науки. № 11(71) СПб., 2008. С. 99.
[8] Общеевропейские компетенции владения языком. Департамент современных языков. Страсбург – Оксфорд. М.: МГЛУ, 2003. С. 109–112.
[9] Общеевропейские компетенции владения языком. Там же. С.112.
[10] Алексеев П.М. Частотные словари и приемы их составления. Статистика речи. Л.: Наука, 1968. С. 61–63; Алексеев П.М. Частотный словарь английского подъязыка электроники. Статистика речи. Л.: Наука, 1968. С. 151–166.
[11] Анализ научного текста: сборник научных трудов / под ред. Т.Н. Хомутовой. Челябинск: ЧГТУ, 1993. С. 152; Хомутова, Т.Н. Составление учебных материалов по иностранным языкам: методические рекомендации для преподавателей вузов неязыковых специальностей. Челябинск: ЧГТУ, 1993. С. 71.
[12] Под словоупотреблением понимается одна из всех словоформ текста или любая последовательность букв, ограниченная двумя пробелами.
[13] Кузнецова Е.Л., Кокорина Е.Л. Составление учебного словаря для студентов специальности ЭВМ // Анализ научного текста: сб. науч. тр./ под ред. Т.Н. Хомутовой. Челябинск: ЧГТУ, 1993. С. 17–25.
[14] Глушко М.М. Отбор и организация слов в учебном терминологическом тезаурусе. Теория и практика английской научной речи. М.: МГУ, 1987. С. 226–234.
[15] Петрушевская Н.Н. Опыт лингвостатистического отбора лексики для обучения чтению в техническом вузе. Анализ содержания курса иностранного языка. Томск: ТГУ, 1976. Вып. 3.
[16] Залевская А.А. Введение в психолингвистику. М.: РГГУ, 2000. С. 262.
[17] Там же. С. 245–247.
[18] http://www.wikipedia.org
[19] http://www.lanaconsult.com
[20] Бабина О.И. Построение модели извлечения информации из технических текстов: дис. … канд. филол. наук. Челябинск, 2006 С. 235.
[21] Глушко М.М. Лингвистические особенности современного общенаучного языка: дис. … канд. филол. наук. М., 1970.
[22] Кокорина С.Б., Чернышева Е.Л. Учебный лексический минимум для студентов специальности ЭВМ (английский язык): учебное пособие/ под ред. Т.Н. Хомутовой. Челябинск: ЧГТУ, 1991. С. 67.
[23] Современные словари рассматривают в качестве терминов большее количество слов, чем это было в 1991 году, что можно объяснить а) адаптивностью терминологической системы, в результате чего ряд общеупотребительных слов перешли в слой терминологической лексики; б) развитием лексикографических ресурсов, которые в настоящее время представляют более полные списки терминов в данной области.
[24] Top-Level Categories for the ACM Taxonomy / URL: http://www.computer.org/portal/pages/ieeecs/publications/author/ACMtaxonomy.html/
[25] Гнаткевич Ю.В. Обучение иноязычной лексике в неязыковом вузе. Киев: Высшая школа, 1989. С. 56–110.
[26] Будагов Р.А. Литературные языки и языковые стили. М., 1967. С. 194; Ахманова О.С. Очерки по общей и русской лексикологии. М.: УРСС, 2004. С. 29–30; Комарова А.И. Функциональная стилистика: научная речь. Язык для специальных целей (LSP). М.: Едиториал УРСС, 2004. С. 49–50.
[27] Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений / пер с англ. М.: Мир, 1976. С. 165.
[28] Пиотровский Р.Г. Лингвистический автомат (в исследовании и непрерывном обучении): учебное пособие. СПб.: Изд-во РГПУ, 1999. С. 256.
[29] Тер-Минасова С.Г. Словосочетание в научно-лингвистическом и дидактическом аспектах. М., 1981.
[30] Лукашевич Н.В., Невзорова О.А. АвиаОнтология: анализ современного состояния ресурса // Компьютерная лингвистика и интеллектуальные технологии: Материалы международной конференции «Диалог-2004». М., 2004.
[31] URL: http://www.linguistics-ontology.org
[32] Загорулько Ю.А. Подход к построению предметной онтологии для портала знаний по компьютерной лингвистике // Компьютерная лингвистика и интеллектуальные технологии: материалы международной конференции «Диалог-2006». М., 2006.
[33] Добров Б.В. Онтологии для автоматической обработки текстов: описание понятий и лексических значений // Компьютерная лингвистика и интеллектуальные технологии: Материалы международной конференции «Диалог-2006». М., 2006.
[34] URL: http://wordnet.princeton.edu
[35] Ср., например, АвиаОнтология (Б.В. Добров, Н.В. Лукашевич, О.А. Невзорова и др.), онтология портала компьютерной лингвистики (Ю.А. Загорулько и др.), онтология организации (А.Ф. Тузовский и др.)
[36] Примером такого использования может служить интеллектуальная поисковая машина (В.Н. Поляков).
[37] Аналогичные концепты верхнего уровня были введены в проекте Mikrokosmos для построения иерархии понятий, где каждый концепт при продвижении по его иерархии определяется как потомок одного из представленных концептов верхнего уровня. Эти концепты задают наиболее общие свойства, наследуемые своими потомками, и необходимы для построения связного графа онтологии.
[38] Степанов Ю.С. Основы общего языкознания. М.: Просвещение, 1975. С. 12.
[39] Рубашкин В.Ш. Онторедактор как комплексный инструмент онтологической инженерии // Компьютерная лингвистика и интеллектуальные технологии: материалы международной конференции «Диалог-2008». М., 2008.
[40] Гусев В.Д. Система OntoGRID для построения онтологий // Компьютерная лингвистика и интеллектуальные технологии: материалы международной конференции «Диалог-2005». М., 2005.
[41] Борковский А.Б. Англо-русский словарь по программированию и информатике (с толкованиями). М.: Русский язык, 1987. С. 333.