

Эволюция лингво-семантических представлений в интеллектуальных системах на основе расширенных семантических сетей

Evolution of linguistic semantic presentations in the intelligent systems based on the extended semantic networks

Козеренко Е. Б. (kozerenko@mail.ru),
Кузнецов И. П. (igor-kuz@mtu-net.ru)

Институт проблем информатики РАН, Москва

В работе рассматриваются вопросы проектирования и развития семантико-синтаксических и лексико-семантических представлений в лингвистических процессорах ряда систем, основанных на аппарате расширенных семантических сетей (РСС). Системы этого класса создаются для извлечения знаний из текстов на естественных языках, отображения извлеченных сущностей и связей в структуре базы знаний и использования знаний для поддержки экспертных аналитических решений в различных сферах приложения. В фокусе внимания находятся инженерно-лингвистические представления, позволяющие построить целостную работающую лингвистическую модель, которая модифицируется в зависимости от конкретной задачи: от «тяжелой» формы на основе детальных глубинных представлений до фокусных редуцированных оболочек, настроенных на узкую предметную область и ограниченный язык общения.

1. Введение

Данная работа посвящена вопросам создания инженерно-лингвистических моделей естественного языка для построения лингвистических процессоров различных классов информационных систем и описанию опыта создания лингвистических представлений в системах, относящихся к области исследований искусственного интеллекта. В центре нашего внимания находятся интеллектуальные системы, разработанные на основе аппарата *расширенных семантических сетей* (РСС) [1–3, 18–19]. Мы будем их называть *РСС-системы*. Эти системы создавались коллективом разработчиков, включая авторов данной статьи в Институте проблем информатики РАН на протяжении целого ряда лет в рамках исследовательских проектов и прикладных систем, ориентированных на конкретные предметные области заказчиков. Мы выделяем 4 поколения РСС-систем. Лингво-семантические представления, заложенные в основу систем этого класса прошли определенный эволюционный путь.

Интеллектуальные РСС-системы содержат развитые *базы знаний*, при этом знания представлены в виде записей на языке расширенных семантических сетей, называемых *РСС-структурами*. Лингвистические знания, таким образом, являются частным случаем «знаний» и также представлены в виде записей на языке расширенных семантических сетей. Основным конструктивным элементом РСС является именованный N-местный предикат, называемый «*фрагментом*». Все множество языковых объектов задается в виде системы предикатно-актантных структур, при этом поддерживаются механизмы представления вложенных структур, что дает очень мощные изобразительные возможности для описания объектов различных языковых уровней. Очень важным фактором является однородность и единообразие лингвистических представлений.

В процессе анализа и синтеза предложений естественного языка используется формально-грамматический аппарат, сходный с грамматиками зависимостей. При этом подходе опорными элементами являются слова и конструкции, выполняющие

роль предикатов в предложении, и результатом анализа предложения должен стать один предикат, соответствующий сказуемому рассматриваемого предложения (т. е. глаголу в личной форме или другому главному предикатному выражению). Таким образом, в процессе анализа, в первую очередь, происходит выявление «слов-действий» и «слов-отношений», т. е. глаголов и других слов, имеющих синтактико-семантические валентности. Примером «слов-отношений» могут служить, например, слова «отец», «друг», и т. п., то есть в данном случае «отношения» — это слова, которые задают сильные четко выраженные синтактико-семантические ожидания.

Семантический анализ в инженерно-лингвистическом понимании — это процесс перевода естественно-языковых выражений во «внутренние» структуры базы знаний (БЗ), в нашем случае эти «внутренними» структурами являются записи на языке РСС. Таким образом, структуры БЗ — это код смысла в интеллектуальных информационных системах подобного рода.

В работе рассматриваются инженерно-лингвистические решения в системах с «полным» лингвистическим анализом — это системы 1-го и 2-го поколений: ДИЕС1, ДИЕС2, Логос-Д [2–3] и системах с «фактографическим» подходом — интеллектуальных системах поддержки аналитических решений (ИСПАР) [18–19], где целью анализа является выделение сущностей и связей из текстов — это системы 3-го и 4-го поколений.

2. Концептуально-лингвистическое моделирование в РСС-системах

2.1. Основные аспекты семантического моделирования

Концептуально-лингвистическое моделирование (КЛМ) — это процесс построения естественно-языковой модели предметной области (ПО) (Рис.1), синтезирующий в себе подходы концептуального и лингвистического моделирования [1–3]. Построение концептуально-лингвистической модели некоторой предметной области подразделяется на следующие этапы:

- построение собственно концептуальной модели, т. е. вычленение базовых понятий, организация их в родо-видовые деревья и определение связей между ними;
- разработка идеографического словаря предметной области, т. е. лексическое наполнение концептуальной модели;

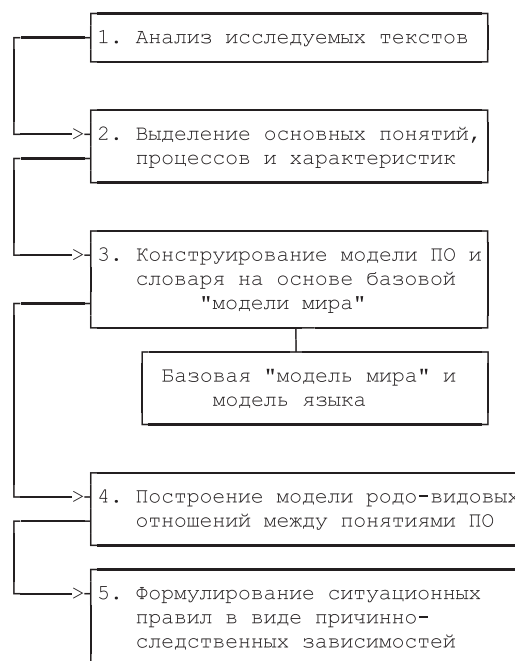


Рис. 1. Процесс концептуально-лингвистического моделирования. — ввод базовых правил, описывающих на естественном языке «модель мира», релевантную для данной ПО

Методика концептуально-лингвистического моделирования на основе аппарата РСС базируется на следующих принципах:

- модель должна быть «открытой», то есть поддерживать эффективный механизм расширения и обновления информации;
- модель представления «смысла» должна учитывать факты экстралингвистической реальности, которые в виде правил и отношений составляют некоторую базовую «модель мира», достраиваемую конкретными моделями предметных областей;
- модель должна быть практичной, то есть не перегруженной детальными описаниями связей и отношений между понятиями, чтобы обеспечить возможность ее реализации, но в то же время, отражать всю релевантную для конкретной задачи информацию.

Реалистичный подход к постановке задачи диктует необходимость ограничения моделируемого подмножества естественного языка. Суть ограничения сводится к следующему:

- во-первых, анализируемые текстовые материалы содержат экспертные знания из конкретных предметных областей (в разработанных авторами системах это были такие предметные области как диагностика брака при изготовлении микросхем, социальное прогнозирование, криминалистика, и другие);

- во-вторых, в целях максимально возможного устранения неоднозначности, словарь строится по модульному принципу: есть некоторая наиболее общая часть (1–2 уровня), которая достраивается специальными словарями для каждой отдельной предметной области.

Предлагаемая модель лексической семантики основана на принципе «ядерного» значения, реализуемого в контексте данной предметной области, с последующим индуктивным наращиванием других значений (если они актуализируются в рассматриваемых контекстах). Также используется таксономия которая реализуется в виде иерархических деревьев классов слов.

Общая «модель мира» системы служит основой для моделей ПО. Элементами этой модели являются классы слов, которые подразделяются на

- понятия / имена,
- отношения,
- действия,
- свойства,
- характеристики действий,
- временные и пространственные характеристики.

Самым общим понятием является *концепт*, или *универсальный класс*, который подразделяется на объект, ситуацию, процесс и др.

Слова, относящиеся к классам действий и отношений, представлены как семантико-синтаксические

фреймы, задающие предикатно-актантные структуры (модель управления). Однако, в описываемом подходе (назовем его РСС-подход) существенно расширена область значений актантов. Суть расширения состоит, во-первых, в том, что в роли актантов могут выступать не только простые объекты, соответствующие отдельным словам, но и структурные объекты, представляющие словосочетания и фразы, а во-вторых, в том, что понятие «падежа» включает в себя не только семантические, но и синтаксические признаки.

Подход, основанный на РСС, позволяет отражать произвольный уровень вложенности структур за счет пропозициональных вершин семантической сети, что обеспечивает представление сложных синтаксических конструкций фраз ЕЯ, а также позволяет отразить структурный характер лексической семантики, которая в предлагаемой модели имеет иерархически-сетевую структуру. Лингвистические знания представлены в системном словаре и декларативных модулях лингвистического процессора. В РСС-системах также реализована функция динамически формируемого семантического словаря, который на основе исходной лингвистической информации достраивается системой автоматически в процессе обработки конкретных текстов. На Рис. 2 представлено такое «внутреннее» описание глагола в семантическом словаре. Этот словарь автоматически генерируется РСС-системами ДИЕС2, ЛОГОС-Д, ИКС в процессе обработки естественно-языковых текстов.

```
{(ВЫРАБАТЫВА895_)(DICSEM)
COORD(PROGNOZ1,RUS,ВЫРАБАТЫВА895_,S50_31_51_20,%)
SUB(UNIV,0+) SUB(UNIV,1+) SUB(UNIV,2+)
ВЫРАБАТЫВ(0-,1-,2-/3+) INFI(3-) ПРИДЕТСЯ(3-)
ПРИДЕТСЯ(3-/4+) FUT1(4-) SUB(СРЕД,5+)
```

Рис. 2. Пример записи представления глагола «вырабатывать» в семантическом словаре

2.2. Аппарат РСС — основа концептуально-лингвистического моделирования

Дадим краткое описание аппарата расширенных семантических сетей и дадим обоснование выбора именно этого метода представления для моделирования естественного языка. Классическое понятие семантической сети сводится к следующему: задаются некоторые вершины, соответствующие объектам. Вершины связываются дугами, которые помечаются именами отношений. Однако с помощью подобных сетей оказывается трудно представлять сложные виды информации, например, когда объекты, связанные отношениями, образуют агрегаты, и когда отношения связываются между собой отношениями и др. Поэтому в сети вводятся

вершины, соответствующие именам отношений, а также специальный композиционный элемент, называемый вершиной связи. Вершина связи как бы «разрывает» дугу и подсоединяется одним ребром к вершине-отношению, а другими ребрами — к вершинам-объектам. РСС является развитием такого сорта сетей в направлении повышения изобразительных возможностей при сохранении свойства однородности.

Основой РСС является множество вершин (V), из которых составляются элементарные фрагменты (ЭФ) следующего вида:

$$V_0(V_p V_{2p} \dots V_k/V_{k+1}),$$

где $V_0, V_1, V_2, \dots, V_k, V_{k+1}, V, k > 0$.

Такой фрагмент представляет k -местное отношение. Позиции вершин в элементарных фрагментах (ЭФ) определяют их роли. Вершина V_0 ставится в соответствие имени отношения, вершины V_1, V_2, \dots, V_k — объектам, участвующим в отношении, а вершина V_{k+1} , отделенная косой линией (/), — всей совокупности упомянутых объектов с учетом их отношения. В дальнейшем будем V_{k+1} называть S -вершиной ЭФ. Множество ЭФ образуют расширенную семантическую сеть (РСС). С помощью РСС представляются наборы отношений, различные ситуации, сценарии. Сильной стороной РСС-подхода является возможность однородного представления как предметной (концептуальной), так и лингвистической информации, что обеспечивает эффективную обработку знаний и поддержание непротиворечивости базы знаний.

Посредством РСС в базе знаний представлены лингвистические (ЛЗ) и предметные знания (ПЗ). Обработка этих знаний осуществляется продукциями языка ДЕKL, на котором реализованы следующие шесть блоков: морфологического анализа (МА), семантического анализа слов (САС), синтаксико-семантического анализа форм (ССА), прагматических функций (ПФ), организации системной активности (БА) и обратный лингвистический процессор (ОЛП). С помощью продукций осуществляется последовательное преобразование сети — РСС. При этом проходятся фазы, соответствующие уровню понимания входного текста. Рассмотрим их.

1. На первом шаге анализа происходит построение пространственной структуры предложения с морфологической информацией для каждого слова. Каждый член предложения представляется вершиной семантической сети. Вместо слова — генерируется код (если слово многозначно, т. е. принадлежит к нескольким классам, — то более одного кода). Основой кода служит корень слова. На этом этапе предложение представляется в виде набора фрагментов типа LRR (специальные метки результатов 1-го этапа анализа), объединяемых в целостную структуру посредством вершины связи. Результат 1-го этапа постоянно обращается к словарю: «Что значит данное слово?»
2. На втором этапе каждой вершине сопоставляется семантический класс и присваивается новый код. За словами (т. е. конкретными вершинами РСС) система видит объекты, действия, свойства — то есть, строит классификации. Производится семантико-синтаксический анализ без выявления глагольных словоформ, при этом предложение представляется в виде совокупности фрагментов типа SEM и SEMD (специальные метки результатов 2-го этапа анализа) (Рис. 3).
3. На третьем этапе происходит частичное «сворачивание» синтаксических структур

в более компактные (например, свойство объекта и сам объект) с присваиванием нового кода, и строится фрагмент для объекта, обладающего этим свойством.

4. На четвертом этапе выявляются отношения и действия и производится анализ непосредственного контекста на соответствие заданным семантическим падежам. Система смотрит, подходят ли объекты (концепты, понятия) на аргументные места данного действия или отношения. При этом отглагольные существительные («делатель» — т. е. агент действия, или «делание» — процесс, анализируются как слова с двойной природой — вначале как действия, а затем как объекты). Результатом этого этапа является целостная семантическая структура предложения, которая представляется фрагментом типа SEMSTR (метка результата 4-го этапа анализа) (Рис. 4).
5. На пятом этапе происходит анализ прагматики: установление кореференциальных отношений, частичное восстановление эллиптических конструкций, система производит дальнейшие действия с построенными фрагментами.

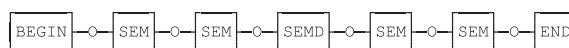


Рис. 3. Семантико-синтаксический анализ без выявления глагольных словоформ

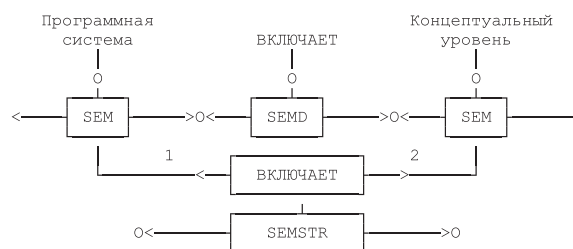


Рис. 4. Целостная семантическая структура предложения

ДИЕС допускает ввод полисемичных форм глаголов. Для этого следует воспользоваться формальной записью лингвистических знаний. Например, можно ввести запись: ВЗЯЛ/-ДЕЙСТВИЕ, КОГО-ЧЕЛОВЕКА ЗА ЧТО-ПРЕСТУПЛЕНИЕ.

Тогда ДИЕС будет понимать предложения типа ИВАНА ВЗЯЛИ ЗА КРАЖУ и другие предложения такого типа. Но ДИЕС будет отличать это действие от других значений глагола ВЗЯТЬ, например, ВЗЯТЬ КНИГУ. Итак, в системах, основанных на РСС, все функции реализованы на единой основе — в рамках языков РСС и ДЕKL, которые были разработаны с ориентацией на задачи обработки естественного языка.

3. Представление семантики глаголов, глубинные и поверхностные структуры

В процессе анализа выявляются семантические вершины предложения — происходит выявление «слов-действий», т. е. глаголов, и «слов-отношений». Что же является конструктивной основой задания семантических представлений предикатных слов и выражений? Как убедительно показано в работе Ю.Д. Апресяна «Экспериментальные исследования семантики русского глагола» [4], семантика глагола определяется его дистрибутивно-трансформационными свойствами. Поэтому смысл предикатных выражений должен кодироваться с учетом их дистрибутивных и трансформационных признаков.

Выдвинутая рядом лингвистов гипотеза (Хомский, Филлмор) [5–8] о том, что все предложения имеют глубинные и поверхностные структуры, явилась очень продуктивным источником проектных решений при создании первых РСС-систем и развивалась в дальнейшем. В теоретико-лингвистическом понимании глубинная структура — это абстракция, содержащая все элементы, необходимые для образования поверхностных структур предложений со сходной семантикой. В инженерно-лингвистическом понимании глубинная структура — это запись на языке БЗ, например, на РСС, которая может быть представлена в «поверхностном» виде на одном из естественных языков в результате конечного числа определенных преобразований. Например, предложения

- (1) The dog chases the cat.
- (2) The cat is chased by the dog.

имеют истоком одну глубинную структуру:



хотя и отличаются своими поверхностными структурами. В каждом из них имеется агент (the dog), объект (the cat), и действие (chase). Согласно концепции *надежной грамматики* Филлмора [5], глубинная структура для обоих предложений инвариантна. Эту структуру можно представить в виде скобочной записи V(AGENT, OBJECT). В графическом виде глубинная структура предложения также может быть представлена диаграммой в виде дерева, где отражены инвариантные отношения зависимости между предикатной вершиной и актантами (Рис. 5), при этом в таком представлении явным образом разграничиваются *модальность* (MOD) и *пропозиция* (PROP).

В исходном виде [5] теория признавала шесть падежей: агентив, инструменталис, датив, объектив, локатив и фактив. По мере развития теории [8] происходило увеличение числа падежей, однако «умножение» количества падежей утяжеляет перво-

начальную конфигурацию, поэтому при построении инженерных семантических представлений требуется некоторый «компромиссный» вариант, сочетающий в себе необходимую полноту, с одной стороны, и простоту и гибкость, с другой.

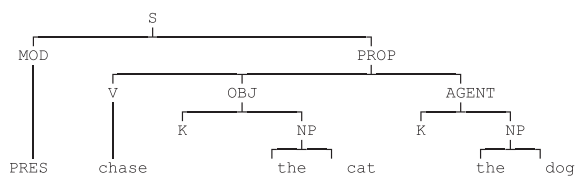


Рис. 5. Глубинная структура предложений

4. Многоязычные системы

Одним из приоритетных направлений развития РСС-систем было обеспечение обработки текстов на нескольких языках, прежде всего, для русско-английской языковой пары. В системах 2-го поколения — ДИЕС2, ИКС, ЛОГОС-Д были реализованы лингвистические процессоры и словари для русского и английского языков, позволявшие обрабатывать тексты для ряда предметных областей, также поддерживались режим ввода лингвистических знаний лингвистом-аналитиком и автоматический режим самообучения системы по вводимым текстам. Проводились также эксперименты для итальянского и французского языков. При создании многоязычных систем мы обращались к европейским языкам. Очевидно, что европейские языки обладают большим количеством общих правил, чем любой из них с языками других групп. Но при этом все естественные языки обладают общей структурой на самом глубинном уровне. На этом уровне располагаются главные элементы естественного языка: Предложение, Модальность, Пропозиция.

Моделирование смысловых представлений — это процесс, развивающийся в направлении от поверхностных семантических структур — к глубинным. Поиск такого внутреннего представления смысла в условиях многоязычной ситуации является развитием методов концептуально-лингвистического моделирования на базе расширенных семантических сетей.

5. Интеллектуальные системы поддержки аналитических решений

РСС-системы 3-го и 4-го поколений направлены на извлечение знаний в виде *объектов*, или *сущностей*, и связей между ними из предметно-ориентированных текстов на русском и английском языках [18–19].

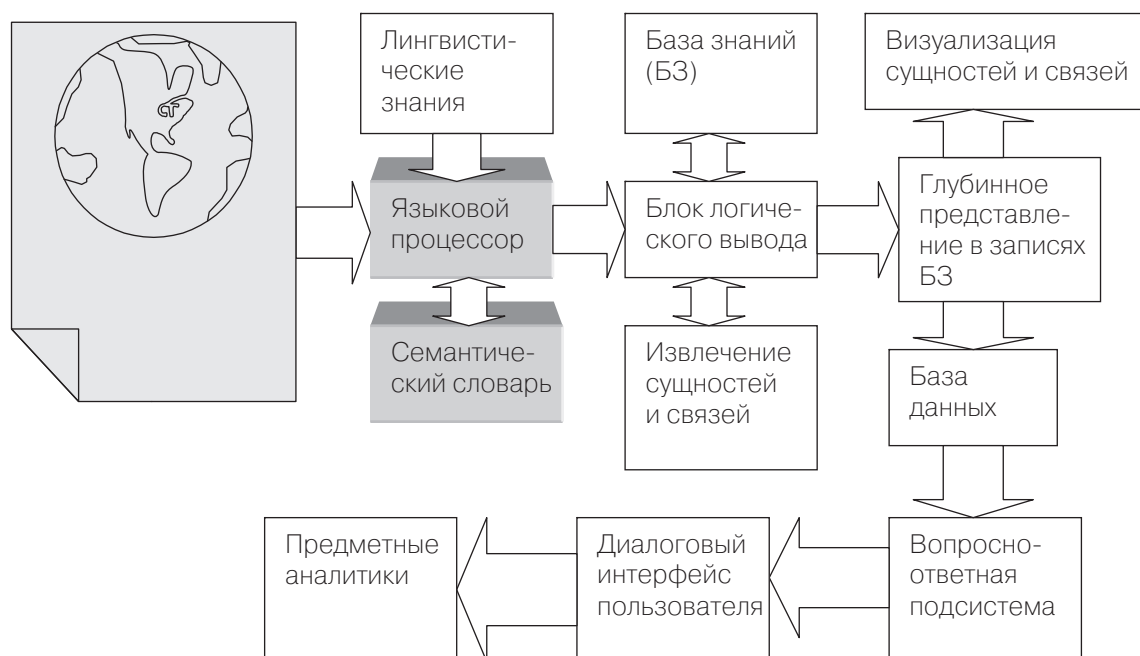


Рис. 6. Обобщенное функциональное представление систем ИСПАР

В настоящее время в мире активно ведутся работы по созданию систем извлечения фактов из текстов на естественных языках [13–16], создаются развитые тезаурусы и онтологии [17]. РСС-системы функционально шире, поскольку помимо возможностей извлечения фактов поддерживают механизмы логического анализа и экспертного вывода на основе извлеченных знаний. Системы такого рода являются интеллектуальными системами поддержки аналитических решений (ИСПАР). В целом это направление исследований требует дальнейшей проработки лексико-семантических представлений, создания предметно-ориентированных семантических словарей. Обобщенное функциональное представление систем ИСПАР дано на Рис. 6.

В рамках ИСПАР на основе расширенных семантических сетей (ИСПАР-РСС) были реализованы полномасштабные и пилотные проекты для ряда предметных областей: криминалистики, управления кадрами, мониторинга финансово-экономического кризиса, и других [18–19].

6. РСС-подход в лингвистических исследованиях

В настоящее время в рамках проектов, направленных на создание открытых лингвистических ресурсов [20] для научно-практических целей ведутся работы по выравниванию параллельных текстов научных статей, патентов и финансово-экономических текстов. В качестве одного из методов выравнивания используется РСС-подход, поскольку он позволяет

отразить глубинно-семантический уровень языковых структур. На рисунке 7 представлен фрагмент первого этапа лингвистического анализа в многоязычных системах — для «идеальной» ситуации, когда структуры исходного текста и текста перевода практически совпадают, такая ситуация имеет место в меньшинстве случаев. Основные трудности возникают при наличии переводческих трансформаций в параллельных текстах. Особое внимание мы уделяем глагольно-именным трансформациям, например, явлению *номинализации*, поскольку она очень продуктивна для всех исследуемых нами языков.

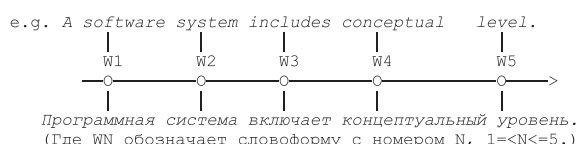


Рис. 7. Первый этап анализа параллельных текстов

Ключевой задачей при разработке методов сопоставления параллельных текстов является выявление и детальное описание тех языковых трансформаций, которые имеют место при переводе естественноязыковых конструкций с одного языка на другой [9], потому что далеко не всегда некоторое содержание передается структурно-подобными средствами в текстах на разных языках. Сравнительное исследование употребления различных частей речи в параллельных текстах на разных языках дает основу для выявления и описания языковых трансформаций, при этом центральной трансформацией является *номинализация*. Явление номинализации было исследова-

но в ряде работ отечественных и зарубежных лингвистов [9–12]. Ближе всего к нашему пониманию этого явления следующие определения номинализации: «конструкции... называются номинализованными — в том смысле, что их естественно рассматривать как результат номинализации конструкций с предикативным употреблением глаголов и прилагательных»; «номинализация — это синтаксический процесс, который соотносит предложения с именными группами». Выявление номинализованных конструкций в параллельных научных и патентных текстах на русском, английском, французском и немецком языках в научных и патентных текстах и сопоставительное описание глагольно-именных межъязыковых трансформаций — одна из центральных задач наших инженерно-лингвистических исследований.

7. Заключение

В данной работе представлен опыт создания и развития лингво-семантических представлений в интеллектуальных информационных системах, разработанных на основе аппарата расширенных семантических сетей (РСС). Аппарат РСС обеспечивает мощные изобразительные возможности для описания всех уровней естественного языка, включая уровень глубинно-семантических представлений, и межъязыковых соответствий. Конкретные лингвистические процессоры, которые были соз-

даны на основе этого подхода, прошли определенный эволюционный путь и позволили выработать проектные решения для основных задач текущего этапа — извлечения и обработки содержательных знаний из текстов на естественных языках и сопоставления языковых структур в текстах на различных языках с учетом базовых трансформаций.

Проблема извлечения и обработки знаний открывает перспективы развития интеллектуальных направлений компьютерной лингвистики, поскольку ее основной акцент смещен в сторону глубинных представлений языка, в которых используются как грамматические (морфологические и синтаксические), так и семантические атрибуты для описания языковых объектов. Проводимые нами исследования параллельных текстов направлены также на рассмотрение этой проблемы [20]. Центральное место в наших лингвистических исследованиях занимает изучение и формализация процессов трансформации языковых структур, особенно все варианты глагольно-номинативных трансформаций, создание развитых дистрибутивно-трансформационных описаний предикатных структур для рассматриваемых языков.

Для задач извлечения знаний и создания систем ИСПАР дистрибутивно-трансформационные описания имеют также особое значение, поскольку таким образом задаются все возможные способы перевода языковых структур в предикатно-аргументные представления, которые затем используются в процедурах обработки знаний.

Литература

1. Кузнецов И. П. Семантические представления // Москва: «Наука», 1986. 290 с.
2. Козеренко Е. Б. Концептуально-лингвистическое моделирование в среде интеллектуального редактора знаний ИКС // «Проблемы проектирования и использования баз знаний». Ин-т кибернетики им. В.М. Глушкова, Киев, 1992. С. 73–79.
3. Kozerenko E. B. Multilingual Processors: a Unified Approach to Semantic and Syntactic Knowledge Presentation // Proceedings of the International Conference on Artificial Intelligence IC-AI'2001. H. R. Arabnia (ed.), Las Vegas, Nevada, USA, June 25–28, 2001. CSREA Press, 2001. P. 1277–1282.
4. Апресян Ю. Д. Экспериментальное исследование семантики русского глагола // Москва: Наука, 1967. 252 с.
5. Филлмор Ч. Дело о падеже // «Новое в зарубежной лингвистике». Вып. X. М.: Прогресс, 1968. С. 369–495.
6. Хомский Н. Аспекты теории синтаксиса // Москва: Изд-во МГУ, 1972.
7. Хомский Н. Язык и мышление // Москва: Изд-во МГУ, 1972.
8. Fillmore C. The case for case reopened // P. Cole & J. Sadok, Eds. Syntax and Semantics. New York: Academic Press. 1977. Vol. 8.
9. Жолковский А. К., Мельчук И. А. О семантическом синтезе // «Проблемы кибернетики», вып. 19, М, 1967.
10. Падучева Е. В. О семантике синтаксиса. Материалы к трансформационной грамматике русского языка. Изд. 2-е. // Москва: КомКнига, 2007. 296 с.
11. Jacobs R. A. and Rosenbaum P. S. English Transformational Grammar. // Blaisdell, 1968.
12. Балли Ш. Общая лингвистика и вопросы французского языка. Изд. 2-е, // Москва: УРСС, 2001.
13. Cunningham H. Automatic Information Extraction // Encyclopedia of Language and Linguistics, 2nd ed. Elsevier, 2005.
14. Han J. and Kamber M. Data Mining: Concepts and Techniques // Morgan Kaufmann, 2006.
15. FASTUS: a Cascaded Finite-State Transducer for Extracting Information from Natural-Language Text. // AIC, SRI International. Menlo Park. California, 1996.
16. Han J., Pei Y. Yin and Mao R. Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach. // Data Mining and Knowledge Discovery, 8(1), 2004. P. 53–87.
17. Добров Б. В., Лукашевич Н. В. Онтологии для автоматической обработки текстов: Описание понятий и лексических значений // Компьютерная лингвистика и интеллектуальные технологии: Тр. междунар. конференции Диалог'06, Бекасово, 31 мая — 4 июня 2006 г., 2006. С. 138–142.
18. Kuznetsov I. P., Efimov D. A., Kozerenko E. B. Tools for Tuning the Semantix Processor to Application Areas // Proceedings of ICAI'09, Vol. I. WORLD-COMP'09, July 13–16, 2009, Las Vegas, Nevada, USA. — CRSEA Press, USA, 2009. P. 467–472.
19. Kuznetsov I. P., Kozerenko E. B., Kuznetsov K. I., Timonina N. O. Intelligent System for Entities Extraction (ISEE) from Natural Language Texts // Proceedings of the International Workshop on Conceptual Structures for Extracting Natural Language Semantics — Sense'09, Uta Priss, Galia Angelova (Eds.), at the 17 International Conference on Conceptual Structures (ICCS'09), University Higher School of Economics, Moscow, Russia, 2009. P. 17–25.
20. Kozerenko E. B. INTERTEXT: A Multilingual Knowledge Base for Machine Translation // Proceedings of the International Conference on Machine Learning, Models, Technologies and Applications, June, 25–28, 2007, Las Vegas, USA. — Las Vegas: CSREA Press, 2007. P. 238–243.