

*Б.Н. СУДАКОВ*, канд. техн. наук, проф., НТУ "ХПИ", Харьков,  
*А.Н. ФИЛИПЕНКО*, магистр, НТУ "ХПИ", Харьков

## **СИНТАКСИЧЕСКИЙ СИНТЕЗ ЕСТЕСТВЕННО-ЯЗЫКОВЫХ ТЕКСТОВ В ЭКСПЕРТНЫХ СИСТЕМАХ**

В статье рассмотрены методы синтеза текстов, способы формирования синтаксической структуры ответов. Рассмотрены причины использования ссылок и деления диалогов на составные части. А также рассмотрены проблемы, которые встречаются при таких методах синтеза, и способы их устранения. Библиогр.: 10 назв.

**Ключевые слова:** метод синтеза текстов, ссылка, диалог.

**Постановка проблемы и анализ литературы.** Формализация интеллектуальной деятельности человека приводит к возникновению лингвистической задачи. Суть задачи состоит в построении формальной модели естественного языка. Компонента, реализующая формальную лингвистическую модель – лингвистический процессор – имеет функцию синтеза текстов. Синтез текстов сопровождается проблемами, связанными с моделированием текстов, построением синтаксической структуры ответов, определением структуры диалогов и пр.

В [1] производится исследование механизмов работы лингвистического процессора, как реализованной на ЭВМ формальной лингвистической модели, способной понимать и производить тексты на неограниченном языке. В [2, 3] приводится классификация интеллектуальных систем по их структуре и решаемым задачам. Описываются особенности систем общения и экспертных систем, получивших наиболее широкое распространение на практике. Рассматривается состояние развития систем общения (естественно-языковых, систем речевого общения, обработки визуальной информации, машинного перевода), экспертных систем и инструментальных систем для их создания. В [4] излагается комплексный подход к проблеме общения, состоящий в исследовании данного процесса, разработке описывающих его моделей и реализации на их основе действующих систем. В [5] изложены основные понятия систем общения на естественном языке, рассматриваются архитектура и основные классы ЕЯ-систем, вопросы создания основных компонент ЕЯ-систем и их методы организации. В [6] рассматриваются отдельные направления теории интеллектуальных систем: представление знаний, логический вывод на знаниях, языки общения, понимание текстов, модели обучения, планирование действий и обработка зрительной информации. В [7] рассмотрены экспертные системы и примеры их реализации. В [8] затрагивается разработка систем, основанных на знаниях. В [9]

включаются сведения, относящиеся к теории экспертных систем, где показано, какое место занимают экспертные системы во всем объеме компьютерных наук. В [10] рассматриваются универсальные модели и методы, применяемые при создании компьютерных систем, способных выполнять функции, традиционно считающиеся интеллектуальными.

**Цель статьи** – определение методов синтеза текстов, обоснование необходимости использования ссылок, решение проблем, возникающих при их использовании.

**Построение синтаксической структуры ответа.** Данный этап является основным модулем синтаксического синтеза. Он выполняет перевод семантического графа ответа в синтаксическую структуру, однозначно соответствующую некоторой фразе естественного языка. Для решения указанной задачи будут использованы  $\Delta$ -грамматики.

Работа данного этапа состоит в последовательном применении к семантическому графу ответа правил. Процесс выбора правил при построении синтаксической структуры направляется методом blacktrack, обеспечивающим исправление, а не отбрасывание неверной синтаксической структуры.

**Выполнение анафорических ссылок.** Существуют следующие основные виды анафорических ссылок: местоименные и именные. Введение анафорических ссылок необходимо для придания синтезируемому тексту ясности и компактности, что способствует его лучшему пониманию пользователем. Выделим следующие случаи использования анафорических ссылок при синтезе текста. Для именных ссылок это использование родовых понятий, замена конкретных участников ситуации на их обобщенное название, синонимическая замена понятия на более лаконичное понятие. Выполнение указанных замен предназначено в основном для придания тексту компактности.

Осуществление местоименных ссылок, кроме придания тексту компактности, выполняет также функции точной идентификации используемых понятий.

**Приписывание морфологической информации лексемам синтаксической структуры фразы.** Указанное действие выполняется для того, чтобы обеспечить синтаксическую сочетаемость лексем в предложении. Для решения данной задачи используются грамматические правила русского языка, записанные в виде  $\Delta$ -грамматики.

Алгоритм работы данного этапа сводится к последовательному применению к синтаксической структуре ответа правил из  $P^{\text{транс}}$  (правил трансформаций синтаксической структуры), а затем правил  $P^{\text{пр}}$  (правил

приписывания морфологической информации). Рассмотрим каждый тип правил из  $P_2$  (множество правил данной грамматики).

Левой частью правила из  $P^{\text{транс}}$  является некоторое дерево зависимостей, в вершинах которого стоят частично характеризованные лексемы. В правой части правила стоит преобразованное дерево зависимостей. Преобразование сводится либо к переименованию дуг, либо к введению дополнительных вершин, возможно сопровождаемому приписыванием вершинам определенной морфологической информации.

Левая часть правила  $P^{\text{пр}}$  является некоторым деревом зависимости с частично характеризованными лексемами, а правая часть – тоже дерево зависимости, вершинам которого приписана некоторая новая морфологическая информация.

**Определение порядка слов в синтезируемом предложении.** Необходимость выполнения данной задачи вытекает из того факта, что словоформы в предложениях русского языка линейно упорядочены, а вершины в дереве зависимостей не имеют линейного порядка. Указанная задача может быть выполнена в три этапа:

1. Сначала из дерева зависимостей формируются исходные группы – последовательности слов, связанных "локальными" синтаксическими отношениями (СИО).

2. Затем исходные группы объединяются в приведенные группы – последовательности исходных групп, связанных "квазилокальными" СИО. Приведенные группы более или менее соответствуют тому, что традиционно называется "группами основных членов предложения".

3. Последним этапом является расстановка приведенных групп, связанных "глобальными" СИО.

Этап определения порядка словоформ синтезируемой фразы завершается расстановкой приведенных групп. Взаимное расположение этих групп определяется в русском языке большим количеством различных по силе факторов, сложно взаимодействующих между собой.

**Морфологический синтез словоформ синтезируемого предложения.** Задачей данного этапа является построение по основе и приписанной ей морфологической информации конкретной словоформы русского языка. Метод морфологического синтеза (МС) может быть реализован декларативным и процедурным способами. При декларативном способе в словаре системы каждой основе сопоставлены различные возможные для нее наборы морфологической информации (МИ) и соответствующие им словоформы. Работа декларативного МС состоит в считывании из словаря по основе и набору МИ соответствующей словоформы этой основы. При процедурном МС

формирование словоформ осуществляется с помощью таблиц аффиксов, расклассифицированных по частям речи. Основными частями речи для синтеза являются существительное, прилагательное и глагол (глагольные формы). Наиболее сложно осуществляется синтез глагольных форм.

**Синтез связного текста (дискурса). Проблема ссылок.** Основным способом лингвистически выражаемой взаимосвязи предложений дискурса является использование ссылок.

Проблема установления ссылок может быть разбита на две подпроблемы: где в контексте искать сущность (референт), обозначаемую данной ссылкой и как определить, что референт и ссылка соответствуют друг другу.

Простейшим решением первой проблемы является поиск референта в предшествующих (на заданное расстояние) утверждениях. Основным недостатком данного метода являлось отсутствие средств для определения глубины возврата назад в поисках исходного понятия. Указанный недостаток на практике приводит как к увеличению времени поиска, так и к ошибкам в установлении ссылок. Решение проблемы ссылок требует как знания структуры диалога, так и учета фактора близости. Решение второй проблемы является тривиальным для простейших видов ссылок и весьма трудным для случаев несовпадения референта и ссылки. Сложность данной задачи вызвана тем, что люди при ссылке на некоторую введенную ранее сущность не только не указывают всех ее свойств, но, наоборот, имеют тенденцию уменьшать количество информации, используемой для ссылки.

Исследования диалогов показывают, что при выборе формы, в которой осуществляется ссылка на референт, люди учитывают два фактора: минимизацию длины сообщения и минимизацию времени поиска слушающим референта по ссылке, использованной говорящим.

**Структура диалога.** Структура диалога повторяет структуру решения задачи, обсуждаемой в данном диалоге. Другими словами, диалог разбивается на поддиалоги, соответствующие подзадачам, выделяемым в процессе решения исходной задачи. Выделяются два основных подкласса диалога-помощи: диалог, ориентированный на решение задачи (ДОР), и диалог, ориентированный на поиск (ДОП) в базе данных. На основании анализа реальных диалогов видно, что влияние структуры задачи особенно ощутимо в диалогах типа ДОР и меньше заметно в ДОП.

Тематическая структура диалога определяется на основе знания структуры решаемой задачи, с одной стороны, и выделения в диалоге лингвистических конструкций, указывающих на сегментацию диалога.

**Контекст и его роль в обработке дискурса.** Общение с помощью отдельных изолированных (несвязанных) предложений даже в ограниченной предметной области весьма неудобно. В то время как общение, допускающее связный текст, не имеет серьезных недостатков. Однако обработка связного текста предъявляет системе естественного языка ряд новых задач. Задача интерпретации предложения связного текста (дискурса) состоит в соотнесении отдельного предложения с контекстом, в котором оно встречается. Без соотношения с контекстом отдельные предложения, образующие дискурс, зачастую являются многозначными или вообще лишенными смысла. На интерпретацию конкретного предложения связного текста оказывают влияние как предыдущие предложения, так и ситуационное окружение, в котором это предложение встречается. Ситуационный контекст в связи с концепцией умолчания, свойственной естественным языкам, обычно в явном виде не присутствует в тексте. Он воссоздается участниками в процессе понимания ими диалога. Можно сказать, что разделение на текстовой и ситуационный контексты подчеркивает способ формирования контекста – явный (текстовой) и неявный (ситуационный). С точки зрения задач, решаемых при анализе дискурса, удобно выделять глобальный и локальный контексты. Глобальный контекст включает общую направленность текста и его ситуационное окружение. С помощью глобального контекста решается проблема установления именных ссылок. Глобальный контекст выделяет из общих знаний системы те, которые уместны при понимании текущей фазы диалога. Локальный контекст содержит предложения, непосредственно предшествующие обрабатываемому предложению. Локальный контекст используется для раскрытия "умолчаний" в предложениях, содержащих эллипсис (намеренный пропуск несущественных слов в предложении без искажения его смысла).

**Интерпретация входного текста и вывод.** Поиск и вывод выполняются как реакция системы понимания на входной текст. Диапазон возможных способностей реагировать чрезвычайно широк и определяется "интеллектуальной" мощностью системы. В минимальном варианте система способна буквально понимать вход, определять требуемые факты как с помощью поиска в знаниях системы, так и с помощью вывода и вычислений. Для извлечения из системы знаний, уместных для входного текста, в общем случае необходимо использовать дедуктивный вывод.

**Выводы.** В результате проделанной работы было показано, каким образом определяется структура диалога и разбиение его на составные

части, а также необходимость использования ссылок и проблемы, которые возникают при их использовании, а также способы их решения. Также показана рациональность разбиения диалога на поддиалоги и зависимость такого разбиения от вида диалога. Предложено учитывать влияние контекста на обработку дискурса, а также последовательность и структуру реакции системы в ответ на полученный ею текст.

**Список литературы:** 1. *Апресян Ю.Д.* Лингвистический процессор для сложных информационных систем / *Ю.Д. Апресян, И.М. Богуславский, Л.Л. Иомдин и др.* – М.: Наука, 1992. – 416 с. 2. Искусственный интеллект: В 3 кн. Кн.1. Системы общения и экспертные системы: Справочник / Под ред. *Э.В. Попова.* – М.: Радио и связь, 1990. – 464 с. 3. Искусственный интеллект: В 3 кн. Кн.2. Модели и методы: Справочник / Под ред. *Д.А. Поспелова.* М.: – Радио и связь, 1990. – 304 с. 4. *Попов Э.В.* Общение с ЭВМ на естественном языке / *Э.В. Попов.* – М.: Наука. Гл. ред. физ.-мат. лит., 1982. – 360 с. 5. *Евдокимова И.С.* Естественно-языковые системы: курс лекций / *И.С. Евдокимова.* – Улан-Уде: Изд-во ВСГТУ, 2006. – 92 с. 6. Искусственный интеллект: Справочник. Книга 2. Модели и методы / *Д.А. Поспелов.* – М.: Радио и связь, 1990. – 304 с. 7. *Балтрашевич В.Э.* Реализация инструментальной экспертной системы / *В.Э. Балтрашевич.* – СПб.: Политехника, 1993. – 238 с. 8. *Гаврилова Т.А.* Базы знаний интеллектуальных систем / *Т.А. Гаврилова, В.Ф. Хорошевский.* – СПб.: Питер, 2000. – 384 с. 9. *Джарратано Дж.* Экспертные системы: принципы разработки и программирование / *Дж. Джарратано, Г. Райли* – М.: "Вильямс". 2007. – 1152 с. 10. *Бондарев В.Н.* Искусственный интеллект: Учебное пособие для вузов / *В.Н. Бондарев, Ф.Г. Аде.* – Севастополь: СевНТУ, 2002. – 615 с.

*Статью представил д.т.н., проф. НТУ "ХПИ" Серков А.А.*

УДК 651.326

**Синтаксичний синтез природно-мовних текстів в експертних системах / Судаков Б.М., Філіпенко А.М.** // Вісник НТУ "ХПІ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2012. – № 38. – С. 184 – 189.

У статті розглянуто методи синтезу текстів, способи формування синтаксичної структури відповідей. Розглянуто причини використання посилань і поділу діалогів на складові частини. А також розібрано проблеми, які зустрічаються при таких методах синтезу і способи їх усунення. Бібліогр.: 10 назв.

**Ключові слова:** метод синтезу текстів, посилання, діалог.

UDC 651.326

**Synthesis syntax of natural language texts in expert systems / Sudakov B.N., Philipenko A.N.** // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modelling. – Kharkov: NTU "KhPI". – 2012. – №. 38. – P. 184 – 189.

This article describes methods for the synthesis of texts, methods of forming the syntactic structure of answers. The reasons of references usage and dividing the dialogues into its constituent parts. And also examined the problems encountered in such methods of synthesis and their solutions. Refs.: 10 titles.

**Keywords:** methods for the synthesis of text, links, dialogue.

*Поступила в редакцію 10.05.2012*