



## ЗАДАЧА СИНТАКСИЧЕСКОЙ СЕГМЕНТАЦИИ КАК ЭТАП СИНТАКСИЧЕСКОГО АНАЛИЗА

**Гаевская Я. К.**

*Национальный технический университет  
"Харьковский политехнический институт",  
г. Харьков, ул. Пушкинская, 79/2, тел. 707–63–60,  
e-mail: yanna.rey@gmail.com*

В связи с быстрым ростом количества информации в разных сферах человеческой деятельности возникает острая необходимость автоматического решения разнообразных задач, связанных с обработкой текстов на естественном языке. Одним из таких средств является синтаксический анализ. Его цель – это автоматическое построение функционального дерева фразы, то есть нахождение взаимосвязей между разноуровневыми элементами предложения. Синтаксическая структура отображает связи, которые существуют между словами (словосочетаниями) в предложении.

Актуальность работы определяется тем, что системы автоматической обработки текстов на естественном языке, которые используют методы статистического анализа (например, те, которые основываются на поиске ключевых слов) для решения задач, связанных с обработкой текстов, в большинстве случаев не позволяют достичь результата, качество которого достаточно для практического приложения. Причина кроется в необходимости учета не только слов, которые представляют текст на естественном языке, но и взаимосвязей между ними, не показанных в тексте в явном виде. Для этого и используется синтаксический анализ.

Выполняет синтаксический анализ инструкция, которая состоит из стандартных элементов, которые осуществляют определенную последовательность операций над словоформами. Результатом анализа является синтаксическая структура предложения, представленная как совокупность данных о синтаксических связях между его единицами.

Однако еще до проведения синтаксического анализа есть возможность выдвинуть некоторые предположения о структуре разбираемого предложения, выделить его фрагменты (сегменты), которые можно разобрать независимым образом. Дальнейший синтаксический анализ будет опираться на эти предположения и получит возможность сразу отбросить часть вариантов. Для использования этих возможностей вводится этап синтаксической сегментации [1].

Построение сегментов, облегчает последующие этапы синтаксического анализа и является базой для формирования нового объекта – графа сегментов предложения, который может служить основой для следующего этапа моделирования структуры многосегментного предложения – исследования семантики связей его сегментов. Сегментация текста на слова и предложения является необходимым этапом автоматической обработки текста. Последующие



этапы анализа существенным образом зависят от принятых соглашений о делении на слова и предложения.

Задачами синтаксической сегментации являются: 1) уменьшение количества омонимов, соответствующих каждой словоформе; 2) выделение синтаксических конструкций предложения. Правило синтаксической сегментации состоит из шаблона поиска фрагмента, шаблона формирования фрагмента и списка исключений.

Функциональные омонимы – это этимологически родственные слова, совпадающие по звучанию, но относящиеся к разным частям речи. Они могут составлять двучленные и многочленные ряды, в которых есть исходное слово и слово (слова), образованные от исходного морфолого-синтаксическим способом. Такие двусмысленности нежелательны, поскольку мысль можно понять неверно. Для борьбы с неоднозначностью используется несколько основных методов [3]:

- а) совершенствуются и уточняются детерминированных правил, которые работают на основе лексических и грамматических данных;
- б) создаются базы знаний об окружающем мире и онтологии, которая дает возможность учитывать экстралингвистические данные;
- в) разрабатываются вероятностные анализаторы, которые учитывают статистические данные языки.

Практически все существующие алгоритмы снятия омонимии включаются в состав синтаксического анализа, который создает трудно разрешимое противоречие, поскольку для успешного снятия омонимии нужны точные результаты синтаксического анализа, для получения которых, в свою очередь, надо вначале снять омонимию. Кроме того, значительный объем начального числа связей существенно замедляет обработку.

Хотя задача синтаксического анализа, необходимого для всех систем машинного перевода и автоматического понимания текстов, была поставлена более полувека назад, устойчивой и общепринятой системы синтаксического анализа до сих пор не создано. Чтобы продвинуться в ее решении, сегодня необходимо не только совершенствовать имеющиеся наработки, но и находить новые подходы – как в теоретическом плане, так и в практическом.

Очевидно, что задача синтаксического анализа, сначала казавшаяся, как и задача машинного перевода, решаемой в обозримое время, является – в силу бесконечной вариативности естественного языка – сложнейшей задачей, допускающей лишь приближенные решения.

#### **Список литературы**

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие / Большакова Е. И., Клышинский Э. С., Ландэ Д. В., Носков А. А., Пескова О. В., Ягунова Е. В. – М.: МИЭМ, 2011. – 272 с.
2. *Гладкий А. В.* Синтаксические структуры естественного языка в автоматизированных системах общения. – М.: Наука, 2007. – 152 с.
3. *Иорданская Л. Н.* Автоматический синтаксический анализ. Т.2. Межсегментный синтаксический анализ. – Новосибирск: Наука, 1967. – 231 с.