

## СИНТАКСИЧЕСКИЙ АНАЛИЗАТОР ПРЕДЛОЖЕНИЙ РУССКОГО ЯЗЫКА НА ОСНОВЕ МОДЕЛЕЙ УПРАВЛЕНИЯ

*Шамаева Елена Денисовна*

*Студент*

*Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия*

*E-mail: Derinhelm@yandex.ru*

*Научный руководитель — Волкова Ирина Анатольевна*

В работе предложен алгоритм многовариантного синтаксического анализа по моделям управления для простого предложения русского языка (без однородных членов, частиц, союзов, обращений).

Современные синтаксические анализаторы делятся на две группы: основанные на правилах и использующие машинное обучение и статистику. Первые требуют полного описания грамматики, что весьма затруднительно для естественного языка. Они также часто не справляются с непроективными конструкциями. Вторые напрямую зависят от качества и полноты датасетов, по которым собирается статистика или на которых проводится машинное обучение. К минусам машинного обучения можно также отнести слабую интерпретируемость его результатов.

Предлагаемый синтаксический анализатор основан на обобщенных моделях управления, идея которых предложена в [1]. Модель управления показывает, могут ли два слова быть связаны отношением «главное–зависимое». Например, слова «осень» и «золотая» можно связать моделью управления, а слова «к» и «очень» — нет.

Всего различаются обобщенные модели управления трех уровней. Модель управления 3 уровня связывает отношением «главное–зависимое» две конкретные словоформы, например, *синего+дома*; *работают+с*. Модель управления 2 уровня — конкретную словоформу и грамматическое слово (заданное только своими грамматическими характеристиками), например, *(прил., Р.п., ...)+дома*; *работают+(предлог)*. Модель управления 1 уровня — два грамматических слова, например, *(прил., Р.п., ...)+(сущ., Р.п., ...)*; *(глагол, наст. время, ...)+(предлог)*.

Модели управления выделены из синтаксически размеченного корпуса СинТагРус. Для их хранения создана база данных, которая содержит 39486 слов, 22429 моделей 1 уровня, 272876 моделей 2 уровня, 494662 модели 3 уровня. Модели управления соответствует весовая оценка, пропорциональная встречаемости в корпусе.

При синтаксическом разборе предложения последовательно выделяются пары слов, которые можно связать моделями управления (разобранное — в качестве главного, а неразобранное — в качестве зависимого). При каждом таком связывании в дерево разбора добавляется одно слово. Таким образом, синтаксический анализ предложения осуществляется путем последовательных переходов от состояния, в котором разобрано  $n$  слов, в состояние, в котором разобрано  $(n+1)$  слово. На каждом шаге алгоритма используется модель управления с наибольшей оценкой.

Например, при анализе предложения «*Я помню чудное мгновение*» сначала разбирается глагол «*помню*», далее происходит поиск слов для связывания с уже разобранным. В базе есть модель управления 3 уровня *я+помню*, поэтому эти два слова можно связать моделью управления. В базе нет модели управления 3 уровня *помню+мгновение*, но есть модель управления 2 уровня *помню+(сущ., В.п.)*, поэтому слова «*помню*» и «*мгновение*» связываются моделью управления. По аналогичным соображениям связываются слова «*чудное*» и «*мгновение*». В результате будет получен синтаксический разбор заданного предложения.

С помощью предложенного алгоритма можно получить несколько синтаксически допустимых вариантов разбора предложения, что является одним из преимуществ данного анализатора.

Анализатор по моделям управления учитывает и статистику встречаемости слов, и их грамматическую сочетаемость; полученные результаты интерпретируемы. Отличительной особенностью данного анализатора является также его устойчивость к непроективным и неграмматичным предложениям. Так, анализаторы Link Grammar Parser и MaltParser не смогли найти правильный разбор предложения «*Книги нашлись старинные*», а анализатор по моделям управления справился с этим примером.

Работоспособность данного алгоритма проверена на предложениях из корпуса СинТагРус, отличных от тех, на основе которых построена база. В дальнейшем планируется разработать синтаксический анализатор, способный разбирать сложные предложения с однородными членами.

### Литература

1. Одинцев Н. В. Обобщенные модели управления. Синтаксический анализатор на основе моделей управления // Диалог'2002, Москва, 2002, Т. 2, С. 401–406.