**Доклад ‘Оценивание автоматического парсинга’**

О.Н. Ляшевская, И.М. Пантелеева

27.11.2017

Мы представляем оценку синтаксического парсера UDPipe, проверенного на текстах из обчающего корпуса REALEC. Наша задача – создать инструмент, который бы оценивал синтаксическую сложность текста. В данной работе мы исследовали, какие ошибки студентов вызывают неправильную работу парсера. Мы сравнивали выдачу парсера с ручной разметкой, обращая внимание на аннотирование, предложенное в корпусе.

Ошибки, влияющие на работу парсера следующие:

* Синтаксическая омонимия
* Ошибки связанные с порядком слов
* Грамматические и спеллинговые ошибки студентов
* Конструкции с причастиями
* Конструкции с вводными словами

Предложения по улучшению работы парсера:

* Сперва следует провести проверку spellcheckerом, а затем применять UDPipe
* Увеличение выборки для более точных показателей
* Избежать ошибок парсера с помощью использования текстов из корпуса для тренировки модели