

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ФИЛОЛОГИЧЕСКИЙ ФАКУЛЬТЕТ

образовательная программа магистратуры "Прикладная, экспериментальная и математическая лингвистика"

РЕЦЕНЗИЯ

На выпускную квалификационную работу магистранта
Дмитрия Юрьевича Баня
на тему **«Исследование параметров сегментации при статистическом машинном переводе с арабского на русский язык»**

Актуальность ВКР, новизна:

Актуальность и новизна данной работы сомнений не вызывают: автоматический статистический перевод с русского на арабский язык является широко востребованной и вызывающей надежды областью компьютерной лингвистики. В то же время количество работ по данному направлению перевода достаточно мало.

Оценка содержания работы, положительные стороны ВКР:

Теоретическая часть работы написана хорошим научным языком и подробно освещает главные аспекты статистического машинного перевода: основные этапы и применяемые методики – выравнивание по словам параллельного корпуса, построение фразовой таблицы, создание модели перестановок, обучение языковой модели, работа декодера, тюнинг.

В то же время, практическая часть работы очень краткая, автор не приводит достаточного количества информации о методиках и принципах реализации экспериментов по улучшению качества сегментации арабского текста, проведённых в рамках работы.

Практическое значение работы:

Работа практическое значение имеет: в ходе работы были получены важные отрицательные результаты, публикация которых позволит улучшить качество автоматического перевода с арабского на русский язык в дальнейших исследованиях.

Недостатки и замечания по ВКР:

Основным недостатком данной работы является фрагментарное представление основных этапов и инструментов, с помощью которых были проведены

эксперименты. Главными инструментами при проведении экспериментов были выбраны сегментатор Stanford Word Segmenter, система перевода Moses и система оценки BLEUScore, однако в работе не раскрываются принципы работы программы Stanford Word Segmenter, а также то, каким образом была обучена её модель. Информация по BLEUScore также является крайне скудной: из работы неясно, чему соответствует 1 балл в данной системы оценки, и насколько приемлемыми являются оценки, полученные в ходе экспериментов. Описание системы Moses нареканий не вызывает, принципы работы системы раскрыты в достаточном объёме.

Автор сформулировал пять принципов, которые могут улучшить сегментацию арабского текста, из них один оказался неприменим к текстам, на базе которых строятся эксперименты, два уже были реализованы в использованных инструментах. Для двух подходов, которые автор реализовал самостоятельно, автор не раскрывает подробности реализации, в том числе не приводится исходный код скрипта Perl, который разработал автор. В связи с этим вызывает сомнения исследовательская составляющая данной работы.

Работа изобилует опечатками, орфографическими и грамматическими ошибками.

Рекомендуемая оценка ВКР удовлетворительно

Рецензент Алексеев Василий Анатольевич, магистр лингвистики, руководитель отдела разработок, ООО «ГРАНДКАПИТАЛ»

«6» июня 2016 г.



(подпись)