# Магистерская диссертация

Попов Андрей Михайлович

«Проблемы автоматического выравнивания в параллельных корпусах флективных языков»

Профиль: «Компьютерная лингвистика и интеллектуальные технологии»

Программа: «Прикладная и экспериментальная лингвистика»

Научный руководитель: к.ф.н., доц. Захаров В. П.

Работа представляет собой исследование, посвящённое проблеме выравнивания в параллельных корпусах. Основной упор сделан на автоматическое выравнивание параллельных текстов по предложениям и по документам. В рамках исследования разработан собственный алгоритм выравнивания текстов по документам. Помимо этого, была проведена серия экспериментов по определению качества выравнивания по предложениям одного из популярных общедоступных инструментов по автоматическому выравниванию. Также, в работе предложен ряд методов по улучшению качества выравнивания параллельных текстов на флективных языках, эффективность этих методов была подтверждена экспериментально.

Актуальность исследования обусловлена отсутствием в свободном доступе инструментов для выравнивания параллельных текстов по документом и чрезвычайно небольшим числом работ по выравниванию параллельных текстов на русском языке.

Материалом для проведения исследования послужили два источника: для выравнивания текстов по документам были выбраны тексты с новостного портала Euronews, для выравнивания по предложениям был использован тестовый корпус WMT13 с сайта, посвящённого статистическому машинному переводу www.statmt.org.

Работа состоит из введения, трёх глав и заключения. Список использованной литературы содержит 36 источников, из них 23 на иностранном языке. Общий объём диссертации 68 страниц (включая приложения).