

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Московский государственный лингвистический университет»
(ФГБОУ ВО МГЛУ)
Federal State Budgetary Educational Institution of Higher Education
«Moscow State Linguistic University»
(MSLU)
ЕВРАЗИЙСКИЙ ЛИНГВИСТИЧЕСКИЙ ИНСТИТУТ-ФИЛИАЛ

Кафедра английской филологии

Колодкина Алёна Анатольевна

**ОСОБЕННОСТИ ВЫРАВНИВАНИЯ ТЕКСТОВ РАЗНЫХ ЖАНРОВ В
ПРОГРАММЕ PARACONC**

Выпускная квалификационная работа студента группы ЕАЛИ9-8-52

Направление подготовки/ специальность: 45.03.02 Лингвистика

Направленность (профиль): Теоретическая и прикладная лингвистика

Научный руководитель доктор филологических наук, доцент, заведующий
кафедрой английской филологии, С.Ю. Богданова

_____ «__» _____ 20__ г.
(подпись)

Заведующий кафедрой доктор филологических наук, доцент, заведующий
кафедрой английской филологии, С.Ю. Богданова

_____ «__» _____ 20__ г.
(подпись)

Заведующий выпускающей кафедрой доктор филологических наук, профессор,
заведующий кафедрой теоретической и прикладной лингвистики Н.Н. Казыдуб

_____ «__» _____ 20__ г.
(подпись)

Иркутск 2016

ОГЛАВЛЕНИЕ

ОГЛАВЛЕНИЕ	2
ВВЕДЕНИЕ	3
ГЛАВА 1. КОРПУСНЫЕ МЕТОДЫ В ПЕРЕВОДЕ ТЕКСТОВ РАЗНЫХ ЖАНРОВ	6
1.1. Теория жанров	6
1.2. Особенности перевода текстов разных жанров	11
1.3 Программные средства осуществления перевода	14
1.4. Корпусная лингвистика и параллельные корпуса текстов	17
Выводы по первой главе	20
ГЛАВА 2. ВЫРАВНИВАНИЕ В ПРОГРАММЕ PARACONC	21
2.1. Средства автоматической обработки параллельных текстов.	21
Программа ParaConc	21
2.2. Автоматическое выравнивание текстов разных жанров.....	23
Выводы по второй главе	48
ЗАКЛЮЧЕНИЕ	49
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ	52
ПРИЛОЖЕНИЯ	58

ВВЕДЕНИЕ

Развитие современной лингвистической науки неразрывно связано с применением информационных технологий, с их помощью процесс исследования языка, его особенностей становится намного эффективнее и быстрее. Особый раздел лингвистики, занимающийся вопросами практического применения языка и его автоматической обработкой, называется «прикладной лингвистикой». В отечественной лингвистике термин «прикладная лингвистика» подразумевает именно компьютерную лингвистику, то есть использование компьютерных методов моделирования языковых процессов. Существует множество различных программ, направленных на работу с различными уровнями и областями языка, его структуризацией, разметкой, переводом.

Говоря о последнем, мы прежде всего подразумеваем машинный перевод, который связан с особой областью прикладной лингвистики – корпусной лингвистикой. Для решения переводческих задач существуют специальные виды корпусов – параллельные корпуса текстов.

Процесс подготовки параллельных корпусов текстов – процедура трудоёмкая, она включает в себя подбор текстов как на языке оригинала, так и на языке перевода, их лингвистическую и экстралингвистическую разметку и выравнивание. В настоящее время все эти процессы в значительной степени автоматизированы посредством использования различного рода программных продуктов, но качество автоматической подготовки текста для создания корпуса по-прежнему остаётся невысоким, поэтому обработанный текст нуждается в ручной правке.

В своей работе мы наиболее подробно рассматриваем завершающий этап в подготовке параллельных корпусов текстов – выравнивание текстового материала.

Выравнивание – установление соответствий на уровне эквивалентных частей текста (слов, предложений) [13].

Процедура выравнивания производится при помощи программы ParaConc на примере текстов различных жанров.

Гипотеза – успешность автоматического выравнивания текстов в значительной степени зависит от их принадлежности к тому или другому речевому жанру, так как каждый жанр имеет свои не только лексические, но и синтаксические особенности, которые должны выявляться в ходе предпереводческого анализа.

Цель – провести исследование результатов автоматического выравнивания текстов различных жанров.

Задачи – определить жанры, подлежащие анализу; произвести отбор текстов на английском языке и их переводов на русский язык в соответствии с определёнными жанрами; произвести выравнивание данных текстов; проанализировать результаты; выявить наиболее удачные и неудачные случаи выравнивания; сделать выводы о зависимости успешности выравнивания от особенностей жанра.

Теоретическая значимость работы состоит в анализе результатов исследований отечественных и зарубежных учёных в данной предметной области, определении понятия «жанр» в рамках данной темы; теоретически обосновать зависимость эффективности выравнивания от принадлежности текстов к различным жанрам.

Практическая значимость работы состоит в описании основных причин ошибочного автоматического выравнивания текстов с учётом их жанровой специфики.

Актуальность. В связи развитием систем машинного перевода, широким распространением использования программ автоматического перевода, а также развитием корпусной лингвистики данная работа является актуальной, так как приводит теоретическое обоснование зависимости качества осуществления машинного перевода от жанра переводимого текста и анализирует практические примеры.

Предмет – Особенности выравнивания текстов.

Объект – Результаты, полученные в процессе выравнивания текстов разных жанров.

Во Введении обозначается общее направление исследования, его теоретические основы и возможность практической реализации.

В Главе I рассматриваются теоретические вопросы, связанные с общей теорией жанров и определением термина «жанр» с точки зрения компьютерной лингвистики, кроме того, вопросы, связанные с автоматизированными средствами перевода и корпусной лингвистикой и, в частности, применением параллельных корпусов как базы для систем машинного перевода.

В Главе II описывается основной функционал программы ParaConc, процесс и результаты выравнивания текстов разных жанров с помощью данной программы.

В Заключении представлен анализ полученных результатов.