

Горизонти



Megaling 2009

прикладної лінгвістики та лінгвістичних технологій

Міжнародна
наукова конференція
Україна, Київ
21 – 26.09.09

Горизонты прикладной
лингвистики
и лингвистических
технологий

Международная
научная конференция

Horizons of Applied
Linguistics
and Linguistic
Technologies

International Scientific
Conference

ПРИМЕНЕНИЕ СТАТИСТИЧЕСКИХ МЕТОДОВ РАСПОЗНАВАНИЯ РЕЧИ ДЛЯ ОПРЕДЕЛЕНИЯ БЛИЗОСТИ ЯЗЫКОВ

Берзинь Акс Улдович (ataols@latnet.lv), Рижский Технический университет, Рига, Латвия

В докладе рассказывается о попытке использования скрытых моделей Маркова, применяемых при распознавании речи, для определения степени близости устных языков. В отличии от распознавания речи, где фонограмма сначала разбивается на единицы речи, мы применяем скрытые марковские модели к целым, продолжительным фонограммам разных информантов.

Таким образом входными данными являются сами фонограммы (которые, конечно, должны соответствовать достаточно высоким требованиям качества), а не их расшифровки в фонетической транскрипции, коими мы пользовались в предидущих исследованиях по определению степени близости языков, что существенно уменьшает трудозатраты по подготовке данных.

Эксперименты были проведены на специально собранном в экспедициях материале латышских и латгальских наречий. Собранные фонограммы содержат произвольную речь на тему своего жизненного пути жителей ареалов соответствующих наречий разных возрастных групп и обоих полов, не утративших способность говорить на своём наречии. Перед использованием в экспериментах из фонограмм удалялись чужие голоса, посторонние звуки и слова и высказывания на других языках.

Результаты экспериментов оказались положительными и даже превзошли наши ожидания.

КОРПУС ПАРАЛЛЕЛЬНЫХ РУССКИХ И БОЛГАРСКИХ ТЕКСТОВ

Гочев Недев Гочев (gochev_g@abv.bg), Великотырновский университет им. Святых Кирилла и Мефодия, Велико Тырново, Болгария

В настоящем сообщении описывается «Корпус параллельных русских и болгарских текстов», который разрабатывается на кафедре русского языка Великотырновского университета.

Корпус представляет собой свод параллельных оригинальных текстов русской литературы и публицистики и их переводов на

болгарский язык в электронном виде. Из этого свода с помощью поисковой системы можно получить информацию о переводящих эквивалентах русского слова.

Корпус составлен с целью служить: а) справочником в процессе перевода русских текстов на болгарский язык; б) материалом для исследований в области теории перевода; в) источником примеров для иллюстрации переводной части русско-болгарских словарей, а также для более точного и адекватного определения переводящих эквивалентов; г) помощником в процессе обучения русскому языку, в частности теории и практике перевода.

Корпус состоит из двух разделов:

— первый раздел (собственно корпус) представляет собой базу данных, включающую русские и болгарские параллельные тексты, выровненные по предложениям;

— второй раздел (корпусный русско-болгарский словарь переводчика) упорядочивает корпус с точки зрения типологии переводящих соответствий лексических единиц: в нем, в виде словарных статей описывается для каждого из зафиксированных слов система эквивалентов в переводе тексте с точки зрения смысловой структуры русского слова, языковых особенностей переводящего языка, правил построения текста на переводе тексте языке и интерферирующего влияния исходного языка на переводящий текст, т.е. с точки зрения словарных соответствий, переводческих трансформаций, неточностей и ошибок в переводе.

ПРОБЛЕМА ВИЗУАЛИЗАЦИИ ЛИНГВИСТИЧЕСКИХ ДАННЫХ

Ищук Павел Александрович (boeg@ua.fm), Украинский языково-информационный фонд, Киев, Украина

Среди отраслей языкоznания, использующих в качестве инструмента компьютерные технологии, особое положение в настоящее время занимает корпусная лингвистика. Во многом благодаря наличию специального инструментария манипулирования языковыми данными эта область становится основой многих лингвистических исследований. Постоянное увеличение объемов естественноязыковой информации, формализуемой в структуре данных корпуса, рост производительности компьютеров и сетей, прогресс технологий программирования — всё это обуславливает необходимость постоянного совершенствования менеджера корпусов, модификации функ-