

МИНОБРНАУКИ РОССИИ
ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
“ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ”

Факультет романо-германской филологии

Кафедра немецкой филологии

Лингвистика. Теория и методика преподавания иностранных языков и культур

Курсовая работа

12 Способы применения ПК для перевода текстов

Студент _____ Титова Д.С. 02.10.2018

Руководитель _____ Дониная О.В. к.т.н., доцент 03.10.2018

Содержание

1. Введение.....	3
2. Перевод, выполняемый человеком при помощи компьютера	3
2.1. Электронные словари	3
2.2. Плюсы и минусы использования электронных словарей.....	7
3. Перевод, выполняемый компьютером с помощью человека	7
3.1. Системы Translation Memory	7
3.2. Функции Translation Memory	9
4. Машинный перевод.....	9
4.1. Определение и задачи.....	9
4.2. Проблемы машинного перевода.....	10
4.3. Перспективы развития машинного перевода.....	11
4.4. Системы машинного перевода	11
4.5. Преимущества и недостатки машинного перевода в целом	14
5. Джорджтаунский эксперимент	14
6. Заключение	15
7. Используемая литература.....	16
8. Интернет-источники	16

1. Введение

Перевод есть вид человеческой языковой деятельности, в результате которой некоторый текст на одном языке ставится в соответствие тексту на другом языке, при этом обеспечивается их смысловая эквивалентность. Переводом текстов человек начал заниматься ещё в античном мире – 20 веков назад. Изначально появляется пословный перевод, который искажает смысл и стиль текста, впоследствии его сменяет вольный перевод, который заключается в передаче лишь общей идеи текста основного языка.

На данный момент существует множество классификаций переводов, остановимся на одной из них. По степени механизации перевода выделяют:

1. Традиционный (“ручной”) перевод, выполняемый человеком;
2. Перевод, выполняемый человеком при помощи компьютера;
3. Перевод, выполняемый компьютером при помощи человека;
4. Машинный или автоматический перевод (выполняется без участия человека)

Однако перевод, выполняемый полностью без участия человека сегодня ещё не может быть высококачественным, но это как раз та область, в которой на данный момент работает прикладная лингвистика, совершенствуя возможности ПК в сфере перевода.

2. Перевод, выполняемый человеком при помощи компьютера

2.1. Электронные словари

Чаще всего, в процессе перевода специалист использует электронные словари. Человек же (чаще всего ученик), которому требуется не особо качественный перевод может использовать и онлайн-переводчики, они делают переводы, которые сильно искажают текст, ведь программа не вникает в контекст фразы и особенности языка. Поэтому учителя могут легко отличить, кто переводил тексты самостоятельно, а кто при помощи онлайн-переводчика.

Электронный (автоматический) словарь - это словарь в специальном машинном формате, функционирующий как часть программного обеспечения

компьютера. Сегодня широко распространяются электронные версии самых различных словарей. В отличие от традиционных словарей электронный словарь наряду с текстом и графическими изображениями может содержать весь спектр медиаобъектов, включая видео- и анимационные фрагменты, звук, музыку и прочее.

Все электронные словари можно разделить на два типа:

1. автоматические словари конечного пользователя;
2. автоматические словари для программ обработки текста (это информационно-поисковые тезаурусы, частотные словари, рубрикаторы, классификаторы, словари морфологического анализа; словари для машинного перевода), которые включают подробную информацию о морфологических, синтаксических и семантических особенностях функционирования слова. Количество зон словарных статей в таких словарях варьируется от 1 до 100. Каждая зона содержит особый тип словарной информации: лемма, грамматическая информация или стилистические пометы, зоны значения и толкования. В электронных словарях количество зон больше, чем в обычных бумажных.

Электронные словари имеют серьезные преимущества по сравнению со своими бумажными аналогами, что проявляется в быстром росте соответствующего рынка. Электронный словарь принципиально может обойти ключевое противоречие книжной лексикографии: чем больше информации предлагает словарь, чем больше развит его научный аппарат, тем сложнее им пользоваться. Поэтому классические словари разделяются на две категории. Первая - популярные, относительно удобные, но довольно простые. Вторая - обстоятельные академические издания, не всегда позволяющие быстро получить искомую информацию. Современные электронные словари не только значительно превосходят по объему книжные, но и находят искомое слово или словосочетание за несколько секунд.

Фактически многие словари, которые сформировались в языковой атмосфере середины прошлого века, сильно устарели. Появляются новые отрасли производства, науки, бизнеса, культуры. В обычную разговорную речь приходят новые слова, термины, устойчивые словосочетания. В них не указаны современные значения старых слов, а многие новые слова просто отсутствуют, так как бумажные словари слишком долго готовятся. Электронные словари могут оперативно обновляться.

На сегодняшний день в России получили широкое распространение электронные словари разных издательств: Lingvo (ABBYY Software House), Мультилекс (МедиаЛингва), Polyglossum (ЭТС - "Электронные и традиционные словари"), Контекст (Ин-форматик), PROMT (ПРОМТ) и многие другие. Эти словари в большой степени универсальны, но вместе с тем каждый из них тяготеет к определённой нише.

Два самых известных электронных словаря - Lingvo компании Abbyu и МультиЛекс, разработанный фирмой МедиаЛингва. Специалисты, создающие эти словари, исповедуют разные взгляды на принципы электронной лексикографии. Компания МедиаЛингва придерживается при разработке словарей МультиЛекс стратегии, которая заключается в создании цифровых копий известных книжных изданий. В основе электронных словарей от этой компании лежат авторитетные современные словари ведущих издательств России. Словари МультиЛекс - это усовершенствованные копии своих печатных оригиналов. В них последовательно воспроизводится содержание печатных словарей, включая предисловия авторов оригинальных изданий и сопутствующие справочные материалы. В качестве объёма словника указывается объём словника оригинала. В распоряжении пользователей имеются мощные и разнообразные средства поиска (распознавание слов и выражений во всех грамматических формах, поиск идиом, выражений, примеров употребления, алфавитный поиск с подсказкой). При работе со словарями МультиЛекс предоставляется точная и полная словарная

информация (транскрипция, отраслевые и стилистические пометы, ударения, толкования, примеры).

Подход МедиаЛингва имеет и недостатки, так как жесткая привязка к бумажному прототипу не дает возможности исправлять и дополнять электронный словарь, а тем более изменять структуру построения словарной статьи. Традиционные словари довольно серьезно отстают от языковой реальности - обычно это не менее десяти лет. А электронные словари можно пополнять чуть ли не ежедневно.

По другому пути пошла компания Abbyy. Основу Lingvo составляет электронный словарь собственной разработки. Каждая новая версия Lingvo дополняется актуальной лексикой, и в ней исправляются найденные ошибки и неточности. Таким образом, благодаря лексикографическим исследованиям англо-русский словарь фирмы Abbyy близок к языковой практике. Словари "Lingvo" не являются точными копиями бумажных аналогов. Так, в качестве источников для подготовки экономического словаря "LingvoEconomics" указаны семь словарей различных авторов. Продукты линейки Abbyy Lingvo давно пользуются заслуженной популярностью у пользователей благодаря полноте и современности лексической базы, удобству в работе, скорости поиска.

Электронные словари не только содержат транскрипцию, но и могут произносить слова. Здесь тоже существует два подхода. В МультиЛекс встроен синтезатор звука и произносятся все слова. Однако полностью доверять такому подходу, не контролируя его по транскрипции, опасно. Синтезатор может неправильно поставить ударение или вообще исказить произношение слова. В Abbyy Lingvo основную лексику озвучивает диктор с оксфордским произношением, считающимся для английского языка классическим.

2.2. Плюсы и минусы использования электронных словарей

Электронные и бумажные словари имеют свои преимущества и недостатки, которые представлены в таблице 1.

Электронные словари		Бумажные словари	
+	-	+	-
Быстрота поиска	Недостаточная фиксация внимания при поиске слова	Тренировка памяти и внимания	Трата большого количества времени на поиск
Большая база данных		Помощь в запоминании слова	
Обновляемость		Возникновение ассоциаций, связанных с поиском слова	
Удобство использования			
Возможность прослушать произношение слова			

Таблица 1. Плюсы и минусы бумажных и электронных словарей

3. Перевод, выполняемый компьютером с помощью человека

3.1. Системы Translation Memory

Говоря о машинном переводе при участии человека, обычно подразумевают редактирование текстов как до, так и после их обработки компьютером. Люди-переводчики изменяют тексты так, чтобы они были

понятны машинам. После того, как компьютер сделал перевод, люди опять-таки редактируют грубый машинный перевод, делая текст на выходном языке правильным.

Среди систем, помогающих переводчику в работе, важнейшее место занимают так называемые системы Translation Memory (TM). Системы TM представляют собой интерактивный инструмент для накопления в базе данных пар эквивалентных сегментов текста на языке оригинала и перевода с возможностью их последующего поиска и редактирования. Эти программные продукты не имеют целью применение высокоинтеллектуальных информационных технологий, а наоборот, основаны на использовании творческого потенциала переводчика. Переводчик в процессе работы сам формирует базу данных (или же получает ее от других переводчиков или от заказчика), и чем больше единиц она содержит, тем больше отдача от ее использования.

Вот список наиболее известных систем TM:

- § Transit швейцарской фирмы Star,
- § Trados (США),
- § Translation Manager от IBM,
- § Eurolang Optimizer французской фирмы LANT,
- § DeJaVu от ATRIL (США),
- § WordFisher (Венгрия).

Системы TM позволяют исключить повторный перевод идентичных фрагментов текста. Перевод сегмента осуществляется переводчиком только один раз, а затем каждый следующий сегмент проверяется на совпадение (полное или нечеткое) с базой данных, и, если найден идентичный или похожий сегмент, то он предлагается в качестве варианта перевода.

3.2. Функции Translation Memory

1. Функция сопоставления (Alignment). Одно из преимуществ систем ТМ - это возможность использования уже переведенных материалов по данной тематике.
2. Наличие фильтров импорта - экспорта. Это свойство обеспечивает совместимость систем ТМ с множеством текстовых процессоров и издательских систем и дает переводчику относительную независимость от заказчика.
3. Механизм поиска нечетких или полных совпадений. Если при переводе текста система встречает сегмент, идентичный или близкий к переведенному ранее, то уже переведенный сегмент предлагается переводчику как вариант перевода текущего сегмента, который может быть подкорректирован.
4. Поддержка тематических словарей. Эта функция помогает переводчику придерживаться глоссария.
5. Средства поиска фрагментов текста. Этот инструмент очень удобен при редактировании перевода. Если в процессе работы был найден более удачный вариант перевода какого-либо фрагмента текста, то этот фрагмент может быть найден во всех сегментах ТМ, после чего в сегменты ТМ последовательно вносятся необходимые изменения.

4. Машинный перевод

4.1. Определение и задачи

Машинный перевод - технология связного перевода текста компьютерной программой с одного языка на другой.

В настоящее время перед учёными, занимающимися прикладной лингвистикой, стоит задача создать такую машину, которая сможет правильно переводить текст без необходимости дальнейшего редактирования. Наиболее перспективным представляется совершенствование подсистем грамматического анализа и синтеза в направлении развития систем

машинного перевода. Также можно рассмотреть возможность увеличения объёма контекстуального охвата текста и совершенствование семантических цепочек для более корректного подбора значений слов.

4.2. Проблемы машинного перевода

На данный момент существует немало проблем и недостатков машинного перевода, которые представлены в рисунке 1.

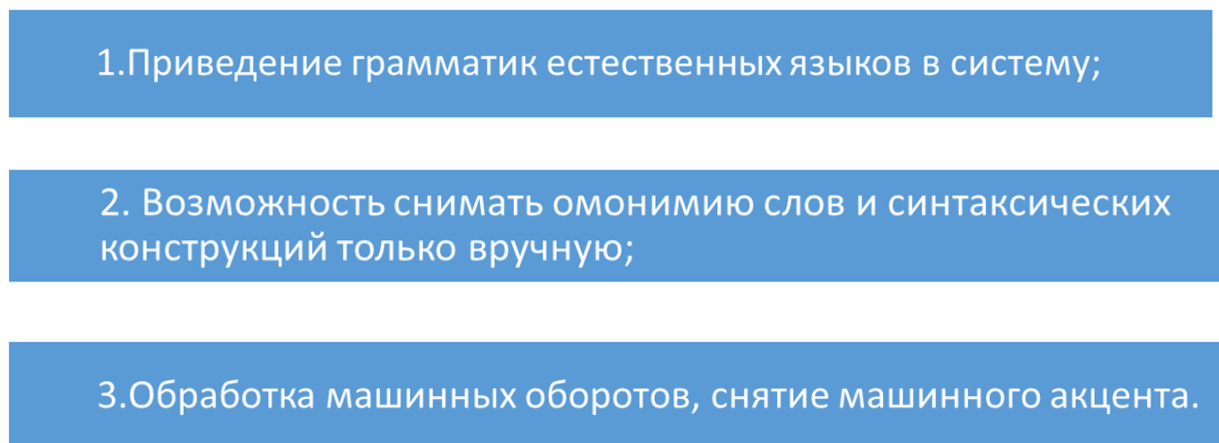


Рис.1. Проблемы машинного перевода

Стоит отметить, что на сегодняшний день результаты работы МП далеки от совершенства, и необходима профессиональная корректировка ряда обнаруженных погрешностей в переведенном тексте. Анализы готового текста на лексическую целостность показали, что вычислительные машины успешно справляются с простыми частями речи и устойчивыми выражениями, но допускают частые ошибки при переводе падежей, речевых оборотов, логико-смысловых акцентов в предложении и построении самих предложений. Это объясняется различной интерпретацией падежей на примере русского и английского языков: в русском — через окончание, в английском — через предлоги. То есть, грамматики естественных языков должны быть более формализованными, чтобы упростить процесс перевода. В таком случае несложная грамматическая структура китайского языка, а именно отсутствие как таковых окончаний у слов, казалось бы, должна упростить задачу МП, но порядок слов в предложении и огромное количество омонимичных единиц в

языке, всё так же оставляет пробелы и недоработки в попытках создания совершенного текста с помощью только одного МП. В таком случае можно заверить, что готовый текст, который был переведен с помощью МП, почти всегда требует дополнительных доработок. Поэтому адекватность и компетентность результатов работы с помощью электронных систем стоит оценивать не только качеством самого МП, но и, конечно, качеством дальнейшего редактирования.

4.3. Перспективы развития машинного перевода

Перспективы машинного перевода тесно связаны с дальнейшим развитием таких научных направлений, как сравнительное языкознание, общая теория перевода, теория закономерных соответствий и способов представления знаний, а также с оптимизацией и совершенствованием лингвистических алгоритмов. Повысить качество перевода лексических единиц помогут строгие теории терминологизации лексики, а формальные грамматики, ориентированные на перевод, дадут возможность улучшить алгоритмы нахождения переводных соответствий в конкретном контексте, который может быть описан в рамках прикладных теорий, представления знаний. И наконец, появление новых технологий программирования и очередных поколений вычислительной техники обязательно будет способствовать дальнейшему развитию теории и практики компьютерного перевода.

4.4. Системы машинного перевода

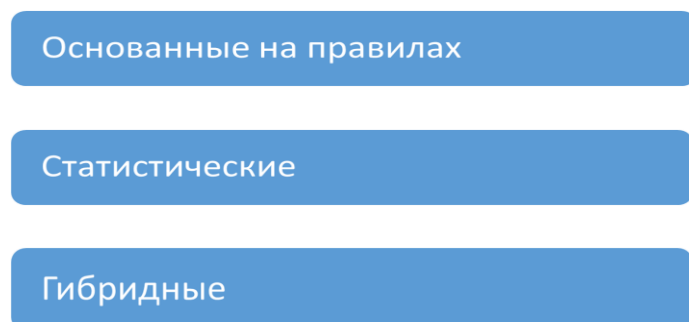


Рис.2. Системы машинного перевода

Существует несколько разновидностей систем машинного перевода, которые представлены на рисунке 3.

1. Системы машинного перевода, основанные на правилах проделывают следующую работу (в большей степени это касается систем пословного перевода):

1. Морфологический анализ каждого слова предложения ИЯ;
2. Синтаксический анализ каждого предложения текста ИЯ;
3. Синтаксический синтез каждого предложения ПЯ;
4. Морфологический синтез каждого слова ПЯ.

ИЯ – исходный язык;

ПЯ – переводный язык.

Всё это должен уметь делать и компьютер, и переводчик. В ходе морфологического анализа каждое слово получает свой набор признаков (род, число, падеж и т.д.). Компьютер может сделать такое, опираясь на формальные признаки слов (суффиксы, окончания и т.д.) либо при помощи использования автоматического словаря. Синтаксический анализ сводится к выявлению составляющих предложения (подлежащего, сказуемого и др.). Синтаксический синтез предложений ПЯ заключается в создании предложений определённой синтаксической структуры, определяемой возможностями ПЯ и синтаксической структурой предложений исходного языка. В памяти ПК должны быть сведения о синтаксических структурах исходного языка и переводного и их соответствиях друг другу. При морфологическом синтезе компьютер ставит все слова в нужном роде, числе и падеже.

Преимущества:

- Синтаксическая и морфологическая точность;
- Стабильность и предсказуемость результатов;

- Возможность настройки на предметную область.

Недостатки:

- Трудоёмкость и длительность разработки;
- Необходимость пополнения словарей;
- «Машинный» акцент.

2. Статистические системы

Статистические системы при своей работе используют статистический анализ. В систему загружается двуязычный корпус текстов (содержащий большое количество текста на исходном языке и его «ручной» перевод на требуемый язык), после чего система анализирует статистику межъязыковых соответствий, синтаксических конструкций и т. д. Система является самообучаемой — при выборе варианта перевода она опирается на полученную ранее статистику. Чем больший словарь внутри языковой пары и чем точнее он составлен, тем лучше результат статистического машинного перевода. С каждым новым переведенным текстом улучшается качество последующих переводов.

3. Основанные на примерах или гибридные системы

Гибридные системы сочетают в себе другие два подхода. Ожидается, что гибридные системы машинного перевода позволят объединить все преимущества, которыми обладают статистические системы и системы, основанные на правилах.

1. Преимущества:

- Легко построить при наличии двуязычного корпуса;
- Перенос технологии на любые пары языков
- Лексическая гладкость.

2. Недостатки:

- Ограниченность параллельных корпусов и их качество;
- Плохая справляемость с морфологией и синтаксисом;
- Искажение информации.

4.5. Преимущества и недостатки машинного перевода в целом

Преимущества и недостатки машинного перевода представлены в таблице 2.

Преимущества	Недостатки
Высокая скорость перевода	Качество перевода, необходимость его корректировки
Низкая стоимость перевода	
Конфиденциальность	
Универсальность	
Перевод в режиме онлайн (перевод интернет-страниц)	

Таблица 2. Преимущества и недостатки машинного перевода

5. Джорджтаунский эксперимент

Джорджтаунский эксперимент — демонстрация возможностей машинного перевода, состоявшаяся 7 января 1954 года. Подготовлена Джорджтаунским университетом совместно с IBM. В ходе неё был продемонстрирован полностью автоматический перевод более 60 предложений с русского языка на английский. Презентация положительно повлияла на развитие машинного перевода в последующие 12 лет.

Эксперимент был задуман и подготовлен с целью привлечения общественного и правительственного внимания. Парадоксально, но в его основе лежала довольно простая система: она была основана всего на 6 грамматических правилах, а словарь включал 250 записей. Система была специализированной: в качестве предметной области для перевода была выбрана органическая химия. Программа выполнялась на мэйнфрейме IBM 701.

Демонстрация была широко освещена в СМИ и воспринята как успех. Она повлияла на решение правительств некоторых государств, в первую очередь США, направить инвестиции в область вычислительной лингвистики. Организаторы эксперимента уверяли, что в течение трёх—пяти лет проблема машинного перевода будет решена. Однако в действительности всё оказалось сложнее, и в 1966 году, на основании отчёта комитета ALPAC — который подытожил, что более чем 10 лет исследований не дали законченного результата, — финансирование было значительно урезано.

6. Заключение

Таким образом, на данный момент переводчик может в большей или меньшей степени использовать ПК для облегчения процесса перевода, пользуясь электронными словарями, системами ТМ, а также почти полностью доверяя перевод ПК. Стоит также отметить, что уже сегодня машинный перевод в сочетании с дальнейшим редактированием текста становится достаточно конкурентно способной переводческой работой на основании стратегических направленностей и рекомендаций, разработанных специалистами на данном поприще. Нельзя не отметить постоянный, инновационный прогресс в сфере компьютерных технологий и растущее осознание необходимости их использования. Таким образом, профессиональные переводчики осваивают полезные и продуктивные плоды постоянного развития информационных систем, и эффективность от такого взаимодействия человека и машины в быстро меняющемся мире перевода будет только возрастать.

7.Используемая литература

- 1.Зубов А.И., Зубова И.И. Информационные технологии в лингвистике: Учеб. пособие для студ. лингв. фак-тов высш. учеб. заведений – Москва: Издательский центр “Академия”, 2004. - с.75-77
- 2.Скворцова О. В., Тихонова Е. В. Проблемы и преимущества автоматизированного и машинного переводов // Молодой ученый. — 2016. — №9. — С. 1287-1289. — URL <https://moluch.ru/archive/113/29477/>
- 3.Андреева А. Д., Меньшиков И. Л., Мокрушин А. А. Обзор систем машинного перевода // Молодой ученый. — 2013. — №12. — С. 64-66. — URL <https://moluch.ru/archive/59/8581/>
- 4.Мироненко А.Б., Омельченко П.А. Проблемы машинного перевода //Актуальные научные проблемы в мире(глазами молодых исследователей) – Красноярск, 2016, - с.1194

8.Интернет-источники

- 1.<https://studfiles.net/preview/3104805/page:3/>
- 2.<https://studfiles.net/preview/4534538/page:2/>
- 3.<https://studopedia.info/2-44741.html>
- 4.https://itc.ua/articles/mashinnyj_perevod_preodolevaya_yazykovye_barery_2002_9/