

Зоны словарной статьи	Описание
Лексический вход	
Определение	
Примеры	
Источники	

3.4. Машинный перевод

Понятие перевода и машинного перевода. Классификация систем МП. Системы переводческой памяти. Этапы осуществления полностью автоматизированного МП. Проблемы МП. Примеры систем МП. Параметры оценки систем МП.

Вопросы машинного перевода составляют одну из центральных областей использования информационных технологий в лингвистике. Это обусловлено не только тем, что в машинном переводе как в фокусе концентрируются все проблемы компьютерной лингвистики — от способов анализа содержания до синтеза словоформы, предложения и целого текста [25, 14], но и постоянно возрастающей практической потребностью современного общества в переводе значительного количества текстов различной функциональной направленности.

Так, свыше 5 млрд жителей Земли используют около трех тысяч языков, и все большее их количество включается в мировые информационные потоки. Разноязычная информация, накапливаемая в геометрической прогрессии, становится труднодоступной, так как на поиск и перевод нужных сведений требуются значительные материальные затраты. Было подсчитано, что если синтез нового химического соединения обойдется менее чем в 100 тыс. долларов, выгоднее произвести этот синтез, чем искать описание аналогичной работы на других языках [52, 114].

Другой иллюстрацией возрастания потребности в переводе служат документы международных организаций, которые в обязатель-

ном порядке переводятся на языки стран-участников. Только Европейский союз в настоящее время объединяет 27 государств, в которых используется 23 официальных языка (<http://europa.eu>). Это обеспечивает работой несколько тысяч профессиональных переводчиков, переводящих в год миллионы страниц. Услуги переводчиков обходятся в миллиарды долларов.

Кроме того, что работа переводчика-человека достаточно дорогая, она к тому же весьма медленная. Так, нормой научно-технического перевода считается время 10 дней на авторский лист (24 страницы машинописного текста) [26, 4]. Система машинного перевода позволяет получить перевод сотен авторских листов за 1 час [20, 79].

Кроме того, появляются новые области применения машинного перевода, например, тексты Интернета. По подсчетам исследователей, в Интернете встроенными системами перевода (SYSTRAN, TRADOS и ESTeam Translator) и сетевыми онлайн-словарями ежедневно выполняется 1 млн запросов на перевод текстов в различных форматах [8, 102].

Все вышесказанное свидетельствует об актуальности обращения к проблеме машинного перевода, который хотя и уступает по качеству переводу, осуществляемому человеком, но даже на сегодняшнем этапе развития позволяет преодолевать языковые барьеры, а кроме того, продолжает оставаться интересной научной проблемой компьютерной лингвистики в целом.

Чтобы определить понятие машинного перевода, обратимся сначала к некоторым общим понятиям теории перевода. *Перевод* обычно понимается как деятельность, «в результате которой некоторый текст на одном языке ставится в соответствие тексту на другом языке, при этом обеспечивается их смысловая эквивалентность» [23, 30]. При этом отмечается многозначность понятия перевода: это одновременно и процесс передачи содержания текста на одном языке средствами другого языка, и результат переводческой деятельности [4, 138].

Перевод представляет собой весьма сложный вид интеллектуальной деятельности человека, поскольку это не чисто языковой, а сложный когнитивный феномен: в процессе перевода человек использует лингвистические и экстралингвистические знания, а кроме

того, в этот процесс включаются два принципиально различных этапа: понимание текста на исходном языке (ИЯ) и синтез текста на языке перевода (ПЯ) [4, 138].

Вследствие такой комплексности переводческого процесса наука о нем (*переводоведение*) носит междисциплинарный характер и оказывается связанной с лингвистикой, литературоведением, когнитивными науками и культурной антропологией [4, 138]. В частности, исследователями отмечается, что переводятся не столько слова и их последовательности, сколько мыслительные образы, порождаемые в сознании переводчика под их воздействием [7, 152], т.е. связь перевода и когнитивных, мыслительных процессов человека очевидна.

Системы машинного перевода моделируют работу человека-переводчика. Таким образом, суть машинного перевода та же, что и в случае его выполнения человеком, с той лишь разницей, что в этом процессе используются компьютеры. *Машинный (или автоматический) перевод* (МП) — выполняемое компьютером действие по преобразованию текста на одном естественном языке в эквивалентный по содержанию текст на другом языке, а также результат такого действия [24, 15].

С точки зрения роли человека в процессе выполнения МП различают следующие его виды [15, 54]:

- МАНТ (*Machine-assisted human translation*) — перевод, осуществляемый человеком с использованием компьютера;
- НАМТ (*Human-assisted machine translation*) — машинный перевод при участии человека;
- ФАМТ (*Fully-automated machine translation*) — полностью автоматизированный машинный перевод.

В первом случае человек использует компьютерные инструменты, направленные на ускорение и упрощение процесса перевода, но собственно перевод текста выполняет сам человек. Вспомогательными системами компьютерной поддержки перевода здесь выступают электронные словари, терминологические базы данных.

Второй тип систем МП является своего рода промежуточным: здесь одинаково важно участие в процессе перевода и человека, и

машины. В машину вводятся электронные словари, морфологические справочники и задается определенный алгоритм выполнения задачи перевода. Роль человека здесь сводится к выбору предлагаемых машиной решений и редактированию текста перевода.

Весьма наглядно такой тип систем МП иллюстрируется системами переводческой памяти (*Translation Memory, ТМ*). Идея таких систем заключается в хранении базы данных переводов, сделанных профессиональным переводчиком, для того чтобы в процессе перевода предлагать человеку уже готовый перевод фразы или куска текста, если он уже был однажды переведен. ТМ-программы значительно повышают эффективность работы переводчика, избавляя его от рутинной, повторяющейся работы. Во многих фирмах, занимающихся переводом, владение одной из таких программ является существенным критерием при приеме на работу.

Третий тип систем МП является наиболее сложным, поэтому остановимся на его характеристике подробнее.

Эффективность полностью автоматизированных систем МП зависит от того, в какой степени в них учитываются объективные законы функционирования языка и мышления. Но эти законы пока еще недостаточно изучены [7, 152; 20, 79], и перед создателями систем МП возникает множество проблем, отражающихся в недостаточном качестве результата МП.

По мере усложнения систем МП и включения в них новых этапов автоматического анализа и синтеза текста выделяют три поколения таких систем [6, 93]:

- 1) П-системы — системы прямого перевода (*direct systems*);
- 2) Т-системы — системы с синтаксическим преобразованием исходного текста (от англ. *transfer* — преобразование);
- 3) И-системы — системы с семантическим и прагматическим анализом (*interlingua* — язык-посредник).

Первый тип систем МП (П-системы) включает лишь этапы морфологического анализа и синтеза, поэтому результат работы таких систем представляет своего рода подстрочный перевод. Во втором типе систем МП (Т-системах) привлекаются методы синтаксического анализа и синтеза, причем в зависимости от их уровня (поверх-

ностный, глубокий или синтактико-семантический) выделяют и разные виды Т-систем. Наиболее сложный тип систем МП — И-системы — включает наряду с лингвистической и экстралингвистическую информацию, т.е. семантику и прагматику предметной области. Поэтому после этапов морфологического и синтаксического анализа фразы исходного текста алгоритм И-системы включает этап семантического анализа. Его результатом служат семантические представления фраз ИЯ и ПЯ, обеспечивающие эквивалентность их смысла [6, 93—94].

В итоге в целом схема машинного перевода включает следующие этапы [20, 80—81; 6, 94]:

- 1) ввод в компьютер текста на ИЯ,
- 2) его морфологический анализ, т.е. определения части речи и морфологических характеристик каждого слова,
- 3) синтаксический анализ каждого предложения текста ИЯ (поиск основных членов предложения и определение типов синтаксических связей между ними, выражаемых в виде дерева зависимостей или дерева непосредственных составляющих),
- 4) семантический анализ каждого предложения ИЯ, в результате которого создается семантическое представление этого предложения, независимое от типа языка (общее и для ИЯ, и для ПЯ),
- 5) синтаксический синтез предложений ПЯ (создание предложений правильной синтаксической структуры, соответствующей правилам ПЯ и типу синтаксической структуры предложения на ИЯ),
- 6) морфологический синтез каждого слова в составе отдельных предложений текста ПЯ (постановка слов ПЯ в нужных морфологических формах);
- 7) вывод текста на ПЯ.

Отдельные трудности процесса МП связаны с необходимостью определения анафорических связей в текстовом целом (*anaphora resolution*) [24, 15], снятия омонимии на разных уровнях, а также с необходимостью привлечения в процесс перевода экстралингвистических знаний [8, 116, 119].

Важность анафорических связей определяется достаточно активным использованием в тексте языковых выражений, которые не могут быть поняты без обращения к предыдущему контексту. Такими выражениями выступают, к примеру, анафорические местоимения *он* или *he*. Установление того, к какому языковому выражению из предыдущего текста относится анафорическое местоимение и к какой сущности реального мира (референту) местоимение и его антецедент отсылает, важно как для понимания всего текста, так и для правильного построения синтаксического и морфологического представления текста. Правильная интерпретация анафорического местоимения требует привлечения данных всех языковых уровней, выхода за рамки одного предложения и привлечения прагматического анализа всего текста [12].

О снятии омонимии говорилось ранее, необходимость же включения экстралингвистической информации в процесс МП иллюстрируется, к примеру, следующими фразами [цит. по: 8, 120]:

Председатель Центральной избирательной комиссии назначается президентом Российской Федерации.

Согласно задумкам американских ученых, сразу после старта вражеские ракеты будут уничтожать авиационные лазеры и мобильные комплексы малых противоракет.

Лишь знания о соответствующих предметных областях позволяют в данном случае определить типы глубинных синтаксических отношений *председатель — президент* ('председатель становится президентом' или 'президент назначает председателя') и *лазеры — ракеты* ('лазеры уничтожают ракеты' или наоборот).

В итоге для функционирования систем МП требуется лингвистическое, программное и информационное обеспечение систем МП. Лингвистическим обеспечением таких систем выступают словари слов и словосочетаний с соответствующими признаками для ИЯ и ПЯ; морфологические таблицы суффиксов и окончаний для ИЯ и ПЯ; базы грамматических правил и др. К программному обеспечению относятся программы выполнения перевода, ведения словарей, формирования базы правил и т.д. Информационное обеспечение представляет база экстралингвистических знаний о предметной области [6, 94—95].

К числу наиболее распространенных в России систем МП относятся [6, 95]:

- *Stylus* — система МП, включающая множество словарей по разным предметным областям;
- *Universal Translator* — многоязычная система МП;
- *Socrate* — система, позволяющая сканировать документы, переводить их содержимое и проверять орфографию;
- *Polyglossum* — многоязычная система МП с широким набором предметных словарей;
- *Prompt* — многоязычная система МП, содержащая множество словарей по разным предметным областям;
- *WebTranSite* — система для перевода веб-страниц (сам процесс перевода веб-страниц и сообщений компьютерных программ называется локализацией).

Сравнение и оценка систем МП осуществляется по следующим параметрам (*Framework for the Evaluation of Machine Translation, FEMTI*) [8, 106—107]:

- характеристики программного обеспечения: надежность системы, удобство использования, скорость работы, возможность обновлений, эффективность, мобильность и т.п.;
- характеристики пользователя и задач перевода: особенности пользователя, автора и текста, а также назначение перевода;
- особенности системы МП: стратегия построения системы, лингвистические ресурсы и т.п.;
- специфика выходного текста: точность, целостность, стиль и т.п., а также наличие ошибок любого характера.

В частности, системы МП письменных текстов в значительной степени отличаются от систем перевода устной речи как по программному обеспечению (в последнем случае обязательно включение в процесс МП этапов автоматического анализа и синтеза устной речи), так и по тематике. Системы для перевода устного диалога обычно ориентированы на узкую тематику: резервирование мест в

гостинице, определение маршрута проезда по городу и т.д. [6, 91]. Соответственно, и оценку каждой из систем МП нужно производить с учетом их названных особенностей.

Итак, машинный перевод, представляющий собой процесс передачи содержания текста на одном языке средствами другого языка с использованием компьютеров, является одним из первых и не теряющих своей актуальности направлений компьютерной лингвистики. Процесс машинного перевода может предполагать разную степень активности человека в его выполнении, что обуславливает многообразие его форм, выбор которых зависит от целей перевода и его условий.

Вопросы для обсуждения

1. Исследователи считают, что причины появления и развития идеи МП лежат в технической, политической и социальной областях. Поясните каждую из причин.
2. Как вы можете объяснить связь процесса машинного перевода и дешифровки текстов?
3. Охарактеризуйте этапы развития МП. Какую роль в развитии идеи МП сыграл американский ученый У. Уивер?
4. Какую роль человек может играть в процессе машинного перевода? Что такое предредактирование и постредактирование?
5. В чем, на ваш взгляд, заключается будущее МП?

Рекомендуемая литература

1. Баранов А.Н. Введение в прикладную лингвистику: учеб. пособие. 3-е изд. М.: ЛКИ, 2007. С. 168—173.
2. Беляева Л.Н. Лингвистические автоматы в современных гуманитарных технологиях: учеб. пособие. СПб.: Книжный Дом, 2007. С. 102—132.
3. Всеволодова А.В. Компьютерная обработка лингвистических данных: учеб. пособие. 2-е изд., испр. М.: Флинта: Наука, 2007. С. 53—63.
4. Зубов А.В., Зубова И.И. Информационные технологии в лингвистике: учеб. пособие. М.: Академия, 2004. С. 75—81, 108—109.
5. Овчинникова И.Г., Угланова И.А. Компьютерное моделирование вербальной коммуникации: учебно-метод. пособие. М.: Флинта: Наука, 2009. С. 80—91.