МИНОБРНАУКИ РОССИИ

Государственное образовательное учреждение высшего профессионального образования

«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Факультет романно-германской филологии

Кафедра французской филологии

Курсовая работа

Машинный перевод

Студент Онучина А.В.

Руководитель Донина О.В.

Воронеж 2018

Оглавление

[Введение 3](#_Toc535166758)

[Основная часть 4](#_Toc535166759)

[1. История появления машинного перевода 4](#_Toc535166760)

[2. Классификация систем машинного перевода 6](#_Toc535166762)

[3. Классификация электронных словарей 8](#_Toc535166766)

[4. Компьютер на месте переводчика 11](#_Toc535166767)

5. Почему популярны программы-переводчики…………………………...12

[Заключение 13](#_Toc535166768)

[Список литературы 14](#_Toc535166769)

**Введение**

Перевод имеет долгую историю. Своими корнями он восходит к тем далеким временам, когда праязык начал распадаться на отдельные языки и возникла необходимость в людях, знавших несколько языков и способных выступать в роли посредников при общении представителей разных языковых общин.

Переводом называется процесс и результат создания на основе исходного текста на одном языке равноценного ему в коммуникативном отношении текста на другом языке.

В современном мире благодаря развитию кибернетики стал возможен машинный перевод, т.е. перевод, строящийся на использовании машиной определенных и постоянных для данного вида материала соответствий между словами и грамматическими явлениями разных языков.

В настоящее время имеется достаточно широкий выбор пакетов программ, облегчающих труд переводчика, которые условно можно подразделить на две основные группы 3: электронные словари (electronic dictionary) и системы машинного перевода (machine translation system). Целью работы является ознакомление с понятием машинного перевода, а так же с его особенностями.

Основыне задачи ислледования:

1)изучить историю возникновения и развития машинного перевода;

2)изучить классификацию различных систем машинного перевода;

3)выявить достоинства и недостатки машинного перевода.

1. **История развития машинного перевода**

Системы машинного перевода (МП) давно перестали быть диковинкой. Они постепенно выходят из младенческого возраста и вместо бессвязного детского лепета начинают изъясняться на вполне понятном, "человеческом" языке. До последнего времени такие программы были не только очень дороги, уступая в цене разве что мощным графическим и издательским системам, но и весьма сложны и капризны в работе. И вот появились первые переводчики, пригодные для использования на домашнем ПК. Для того, чтобы ознакомиться с ними лучше, следует обратиться к истории развития МП.

Еще выдающийся математик XIX века Чарльз Бэббидж пытался убедить британское правительство в необходимости финансировать его исследования по разработке «вычислительной машины». В числе прочих благ он обещал, что когда-нибудь эта машина сможет автоматически переводить разговорную речь. Однако эта идея так и осталась нереализованной .



Рис. 1.

Датой рождения машинного перевода как исследовательской области обычно считают март 1947 г. Именно тогда специалист по криптографии Уоррен Уивер впервые поставил задачу машинного перевода, сравнив ее с задачей дешифровки. Тот же Уивер после ряда дискуссий составил в 1949 г. меморандум, в котором теоретически обосновал принципиальную возможность создания систем машинного перевода. У. Уивер писал: «У меня перед глазами текст, написанный по-русски, но я собираюсь сделать вид, что на самом деле он написан по-английски и закодирован при помощи довольно странных знаков. Все, что мне нужно, - это взломать код, чтобы извлечь информацию, заключенную в тексте»)



Рис. 2.

Идеи Уивера легли в основу подхода к МП, основанного на концепции interlingva: стадия передачи информации разделена на два этапа. На первом этапе исходное предложение переводится на язык-, а затем результат этого перевода представляется средствами выходного языка.

В 1952 г. состоялась первая конференция по МП в Массачусетском технологическом университете, а в 1954 г. была представлена первая полноценная система машинного перевода - IBM Mark II, разработанная компанией IBM совместно с Джорджтаунским университетом. Очень ограниченная в своих возможностях система великолепно переводила 49 специально подобранных предложений с русского языка на английский с использованием словаря на 250 слов и шести грамматических правил.



Рис. 3.

Одной из новых разработок 70-80-х годов стала технология TM (translation memory), работающая по принципу накопления: в процессе перевода сохраняется исходный сегмент (предложение) и его перевод, в результате чего образуется лингвистическая база данных; если идентичный или подобный исходному сегмент обнаруживается во вновь переводимом тексте, он отображается вместе с переводом и указанием совпадения в процентах. Затем переводчик принимает решение (редактировать, отклонить или принять перевод), результат которого сохраняется системой.

С начала 1990-х годов на рынок систем ПК выходят российские разработчики. В июле 1990 г. на выставке PC Forum в Москве была представлена первая в России коммерческая система машинного перевода под названием PROMT (PROgrammers Machine Translation). В 1991 г. было создано ЗАО «ПРОект МТ», и уже в 1992 г. компания ПРОМТ выиграла конкурс NASA на поставку систем МП.

1. **Классификация систем машинного перевода**

Что же касается самих систем машинного перевода, то следует отметить, что они прошли несколько стадии своего развития: . «Электронные переводчики» первого поколения - системы прямого перевода (СПП) - представляли собой программно-аппаратные комплексы и анализировали текст «слово за словом» (смысловые связи и нюансы при этом практически не учитывались). Возможности СПП определялись доступными размерами словарей, прямо зависящими от объема памяти компьютера. Именно к категории СПП относилась IBM Mark II, сделавшая принципиально возможным Джорджтаунский эксперимент.

На смену СПП со временем пришли Т-системы (от английского Transfer - «преобразование»), в которых перевод осуществлялся на уровне синтаксических структур (так учат языку в средней школе). Они выполняли набор операций, позволяющих путем анализа переводимой фразы определять ее синтаксическую структуру по правилам грамматики входного языка, а затем преобразовывать ее в синтаксическую структуру выходного предложения и синтезировать новую фразу, подставляя нужные слова из словаря выходного языка. Работы в этом направлении сейчас уже не ведутся: практикой доказано, что реальная система соответствий сложнее и адекватный перевод требует принципиально иного алгоритма действий.

Немногим позднее становящиеся все более многочисленными системы машинного перевода ТМ-комплексы (от Translation Memory - «память переводов»). В качестве реально успешного примера МТ-программы назовем знаменитую канадскую систему METEO, выполняющую перевод метеопрогнозов с французского языка на английский и обратно (эксплуатируется по сей день). Разработчики METEO сделали ставку на то, что действительно автоматизированный машинный перевод возможен только в условиях искусственно ограниченного языка. И добились успеха. Наиболее популярным в мире профессиональным TM-инструментом является пакет Translations Workbench фирмы TRADOS. Подобные программы используют в основном профессиональные переводчики, осознавшие выигрыш от частичной автоматизации своей работы с помощью компьютера при переводе повторяющихся текстов, сходных по тематике и структуре. Основная идея Translation Memory - не переводить один и тот же текст дважды. Эта технология базируется на сравнении документа, который нужно перевести, с данными, хранящимися в предварительно созданной «входной» базе. Когда система находит фрагмент, соответствующий заранее определенным критериям, то его перевод берется из «выходной» базы. Получаемый в итоге текст подлежит интенсивному постредактированию

Табл. 1. «Классификация систем машинного перевода»

|  |  |
| --- | --- |
| **Системы** | **Характиристика** |
| **СПП (системы прямого перевода)**  **(IBM Mark II)** | Они представляли собой программно-аппаратные комплексы и анализировали текст «слово за словом» ). Возможности СПП определялись доступными размерами словарей, прямо зависящими от объема памяти компьютера. |
| **Т-системы** | Они выполняли набор операций, позволяющих путем анализа переводимой фразы преобразовывать ее в синтаксическую структуру выходного предложения и синтезировать новую фразу. |
| **ТМ-комплексы (METEO; Translations Workbench )** | Эта технология базируется на сравнении документа, который нужно перевести, с данными, хранящимися в предварительно созданной «входной» базе. Получаемый в итоге текст подлежит интенсивному постредактированию |

**3. Электронные словари**

Электронные словари (ЭС) начали разрабатываться значительно раньше, чем альтернативные им системы МП, поэтому в настоящее время на рынке программного обеспечения имеется чрезвычайно широкий выбор словарей - от самых простейших (например, DIC) до мощных систем, объединяющих в одной программной оболочке несколько лексических баз данных - специализированных тематических словарей, последовательность подключения которых определяется пользователем (LINGVO). Выбор того или иного программного продукта зависит исключительно от аппаратных возможностей компьютера, которым располагает пользователь, финансовых соображений и конкретных условий, в которых работает переводчик.

Потенциальную аудиторию этих программ можно определить совершенно точно - это учащиеся и студенты, использующие словарь как одно из пособий при изучении иностранного языка, а также домашние и бизнес-пользователи, которым нужен универсальный и простой в применении словарь-справочник для эпизодического перевода деловых бумаг и корреспонденции.

Ниже представлена классификация электронных словарей по их основным техническим и эксплуатационным характеристикам. Всю совокупность ЭС можно подразделить по следующим критериям:

1. По используемой операционной системе. Наиболее простые электронные словари (DIC) работают под управлением ОС MS-DOS, начиная с версий 2.21 и 3.30, что позволяет их использовать практически на любых IBM-совместимых персональных компьютерах, включая XT, АТ-286. Наиболее сложные многооконные и многофункциональные ЭС, позволяющие в одной оболочке подключать различные тематические базы данных, работают под управлением ОС WINDOWS 3.11, WINDOWS NT, WINDOWS 95 и т.д. Естественно, что для их успешного функционирования необходим более мощный компьютер (типа AT-486DX) с оперативной памятью не менее 8 МБ.

2. По способу загрузки. Можно подразделить на нерезидентные и резидентные. К первым относятся простейшие программы (например, подстрочечный словарь DIC), которые работают только в собственной среде и не вызываются из других оболочек, например из текстовых редакторов. В большинстве случаев они функционируют в режиме автоматического ("пакетного") перевода. Вторые загружают свое ядро в оперативную память компьютера (например, "LINGVO for DOS") и могут вызываться в любой момент работы компьютера, например из любого текстового редактора, при помощи нажатия комбинации "горячих клавиш" - клавиш оперативного вызова. Эти словари обеспечивают работу переводчика в интерактивном режиме.

3. По количеству подключаемых словарных баз (словарей). Ранние версии ЭС позволяли подключать только один словарь. Современные программы, например "Система электронных словарей LINGVO", независимо от того в какой ОС они работают, позволяют подключать до нескольких десятков словарных баз и устанавливать приоритет последних.

4. По возможностям расширения словарной базы. Устаревшие ЭС не имели возможности расширения словарных баз пользователем, современные версии, например LINGVO 4.6 и выше, имеют специальные утилиты для создания пользователем собственных и расширения существующих словарей.

5. По режиму перевода. Можно выделить два основных режима перевода: автоматический пакетный (подстрочечный) и интерактивный (режим "запрос - ответ").

В первом случае программа вызывается с указанием имени текстового файла, который необходимо перевести, или маски для целой группы текстовых файлов, предназначенных для перевода. Далее автоматически, без участия переводчика, производится прямой перевод каждого отдельного слова с выводом на экран монитора или на печатающее устройство только переведенных слов или исходного текста, под каждым пере­веденным словом которого будет стоять его значение на русском языке, под теми словами, которых не оказалось в ЭС обычно ставится знак "?". Такой режим называется подстрочечным. Словарные базы подобных ЭС весьма ограниченны, качество перевода оставляет желать лучшего. Подобные программные продукты вряд ли смогут оказать помощь профессиональному переводчику.

Во втором случае программа работает следующим образом. После загрузки резидентной части программы в оперативную память компьютера, пользователь, работая, например, в текстовом редакторе либо непосредственно в оболочке словаря, вводит с клавиатуры неизвестное слово на языке оригинала, а затем, выделив его, чаще при помощи "мыши", нажатием комбинации "горячих клавиш" активизирует ЭС, который заменяет текст оригинала на русскоязычный перевод. Время доступа к переводу составляет примерно 0,2 секунды, что значительно ускоряет работу переводчика. Именно такие электронные словари, с учетом возможности расширения и одновременного подключения нескольких словарных баз,

могут значительно облегчить и ускорить работу переводчика любого уровня .

Табл. 2. «Классификация электронных словарей»

|  |  |
| --- | --- |
| Критерии | ЭС |
| 1. По используемой операционной системе. | 1. Наиболее простые электронные словари (DIC) работают под управлением ОС MS-DOS 2. Наиболее сложные многооконные и многофункциональные ЭС работают под управлением ОС WINDOWS 3.11, WINDOWS NT, WINDOWS 95 |
| 2. По способу загрузки. | 1. Нерезидентные (словарь DIC) 2. Резидентные (словарь LINGVO for DOS) |
| 3. По количеству подключаемых словарных баз (словарей). | 1. Ранние версии ЭС позволяли подключать только один словарь. 2. Современные программы позволяют подключать до нескольких десятков словарных баз и устанавливать приоритет последних. |
| 4.По возможности расширения словарной базы | 1. Устаревшие ЭС не имели возможности расширения словарных баз пользователем 2. Современные версии имеют специальные утилиты для создания пользователем собственных и расширения существующих словарей. |
| 5. По режиму перевода. | 1. Автоматический пакетный (подстрочечный) 2. Интерактивный (режим "запрос - ответ"). |

1. **Компьютер на месте переводчика**

Вычислительная техника, как известно, хорошее подспорье человеку в рутинной работе. Относится ли к таким занятиям перевод текстов? И да, и нет. С одной стороны, труд переводчика во многом формален, а с другой - перевод не может быть выполнен чисто формально. Есть, например, технический перевод, где важно знать принятые за рубежом стандарты обозначений тех или иных понятий. И есть литературный перевод, когда требуется получить текст, по художественной ценности максимально близкий к оригиналу. Возможно ли поручить подобную работу компьютеру?

Говоря о МП, следует прежде всего помнить, что компьютер - создание бездушное. Он не понимает языковых нюансов, намеков в тексте, того, что называется тонкой игрой слов. Да и, собственно, понять содержание текста в полной мере ему не под силу. Мышления как такового при МП не происходит: предложение расчленяется на части речи, в нем выделяются стандартные конструкции, слова и словосочетания переводятся по находящимся в памяти машины словарям. Затем переведенные части речи собираются по правилам другого языка.

Но этого, недостаточно для полноценного перевода. В зависимости от того или иного стиля и назначения текста одно и то же слово нередко имеет разные значения. В какой-то мере эта особенность учитывается в системах МП: предусмотрены сменные словари, иногда для каждого вида текста предусмотрен свой словарь. Если лексики одного машинного словаря не хватает и применяются несколько словарей одновременно, можно указать системе, из какого словаря нужно брать слово, если есть несколько вариантов его перевода. Наконец, программа сама может предлагать на выбор пользователю несколько вариантов перевода, и он выбирает подходящий вариант, так сказать, вручную. Могут возникнуть и проблемы с переводом слов в устойчивых словосочетаниях и фразеологизмах, но это вполне по силам компьютеру.

Наряду с установленными правилами построения предложения в каждом языке существуют и свои неписаные законы, которые иногда называются красотами языка. Например, предложение на английском языке "This is my book" дословно переводится "Это есть моя книга", и формально это будет правильным, но по-русски так не говорят. В данном случае можно сказать, что предложение "написано так, будто его составил иностранец". Конечно, приведенный пример является простейшим, и возможность исключения слова "is" очень просто отражается в программе МП. Но на практике получившийся перевод похож на текст, написанный иностранцем.

Текст также может содержать слова, которые нужно понимать в контексте образа жизни людей в конкретной стране. Например, под словом "демократ" в США подразумеваются политики, выступающие за большее вмешательство государства в экономику, а в России те, кто выступает за большую свободу рынка. Это разные понятия.

Заглавные буквы и сокращения таят в себе и другие подвохи. Когда слово начинается с большой буквы, его перевод будет начинаться тоже с большой буквы. Слово, целиком состоящее из таких букв, также будет в переводе записано заглавными. В англоязычной литературе достаточно часто встречаются внешне эффектные аббревиатуры, которые могут быть прочитаны как одно слово. Такая аббревиатура и будет переведена единым словом.

Таким образом, результаты МП часто требуют редактирования. Насколько адекватными можно считать результаты перевода на компьютере? Это определяется не только качеством системы МП, но и качеством последующего редактирования. Нередко систему МП использует в качестве подспорья специалист, которому нужно быстро перевести, например, техническую документацию. Тогда проблема корректного употребления терминов решается сама собой.

1. **Почему популярны программы-переводчики**

В нашей стране разработка систем МП ведется уже более 35 лет, а специалисты компании ПРОМТ занимаются этой проблемой более 20 лет, однако до сих пор эта сфера компьютерной индустрии вызывает много различных дискуссий. Часто приходится слышать от критиков систем МП: "Программы-переводчики переводят плохо, непонятно кому они вообще нужны. Гораздо проще отдать текст "живому" переводчику или выучить язык самому".

Итак, компьютер пока во многом не может заменить переводчика. Стоит ли тогда вообще применять системы машинного перевода? Конечно, стоит. Если компьютер используется для перевода литературных текстов, то получается черновой вариант текста, так называемый подстрочник, который превращается в произведение искусства человеком, слабо владеющим языком оригинала, но являющимся хорошим литературным редактором. Если же речь идет о переводе технических текстов, то здесь при правильном выборе словаря по специальности, в рамках которой написан текст, получается вполне удовлетворительный результат, иногда не требующий последующего вмешательства. Вообще необходимость редактирования компьютерного перевода очень часто возникает в связи с проблемами, перечисленными ранее. Для этого системы машинного перевода обязательно имеют средства редактирования текстов.

Однако машинный перевод имеет множество достоинств.

1. Высокая скорость. Всего несколько секунд и Вы получаете перевод многостраничного текста. Это позволяет быстро понять смысл текста, а если система настроена на перевод текстов этой тематики, требуется минимальная редакторская правка.

2. Низкая стоимость. Это очень легко оценить: если вы обращатесь к профессиональным переводчикам, приходиться платить за каждую страницу переведенного текста, либо вы нанимаете штатного переводчика, которому приходится платить зарплату. В случае с системой машинного перевода, вы платите деньги только один раз - при покупке программы.

3. Доступ к услуге. Немаловажный фактор, который многие критики систем МП не принимают в расчет. Программа-переводчик всегда под рукой, а обращаться в переводческое бюро во многих случаях связано с дополнительными затратами времени и сил.

4. Конфиденциальность. Системе МП вы можете доверить любую информацию. Приятно ли вам отдавать на перевод личную переписку? Готовы ли Вы к тому, чтобы посвятить в свои финансовые дела постороннего переводчика. Если да, то действительно вопросов нет. Но даже в этом случае, согласитесь, немногие пользователи Вас в этом поддержат.

5. Универсальность. Любой переводчик всегда имеет специализацию, т.е. переводит тексты по той теме, которой он хорошо владеет. Система МП выгодно отличается тем, что она абсолютно универсальна. Нужно только грамотно подключить специализированный словарь по соответствующей тематике. В ряде случаев также рекомендуется вести свой собственный словарь новых терминов или новых значений. В этом случае вы гарантированно получаете необходимое качество перевода.

**Заключение**

История машинного перевода насчитывает немногим более 50 лет. За это время сменилось несколько поколений систем машинного перевода - от первых программ, использовавших ограниченные ресурсы универсальных компьютеров первого поколения до современных коммерческих продуктов, использующих мощные ресурсы серверов и персональных компьютеров. До уровня полной автоматизации перевода человечество еще не дошло. Причиной этому является, недостаточный уровень развития наук, работающих над созданием подобных систем. Рассмотрев проблемы машинного перевода, можно сказать, что на данном этапе развития машинного перевода пока еще остались не разрешенными его основные проблемы. Наибольшее число трудностей при переводе и соответственно ошибок при переводе относится к уровню грамматики. Средства машинного перевода пока не могут улавливать все смысловые нюансы оригинального текста. Как видно из всего сказанного выше, практические проблемы, возникающие при применении автоматического оборудования для перевода с одного языка на другой разнообразны и сложны. Однако прогресс не стоит на месте, и с увеличением быстродействия процессоров и ростом объема оперативной памяти постепенно повышается и качество машинного перевода. Пока еще машинный перевод нуждается в обязательном редактировании. В заключение подчеркнем, что программа-переводчик - это, прежде всего, инструмент, который позволяет решить проблемы перевода или повысить эффективность труда переводчика только в том случае, если он используется грамотно. 

Список литературы

1. О. С. Кулагина. О современном состоянии машинного перевода. Математические вопросы кибернетики, вып. 3, М.: Наука, 1991, стр. 5—50. Библиография из 140 названий.
2. Гращенко Л. А., Клышинский Э. С., Тумковский С. Р., Усманов З. Д. Концептуальная модель системы русско-таджикского машинного перевода. Докл
3. ады Академии наук Республики Таджикистан. — 2011. — том 54, № 4. — С. 279—285.
4. Будущее машинного перевода. Компьютерра № 21, 5 июня 2002 – (<https://web.archive.org/web/20140623234325/http://old.computerra.ru/offline/2002/446/18251/>)
5. Бабина О. И. Языковая личность переводчика и машинный перевод. Вестник Челябинского государственного университета. 2011. № 24 (239). Филология. Искусствоведение, вып. 57. С. 191–193.
6. Особенности машинного перевода – (<http://ikit.edu.sfu-kras.ru/files/12/ch01s01s03.pdf>).
7. Васильев А. "Подводная лодка" 6/98. Компьютер на месте переводчика. – (<http://www.ets.ru/arc07-r.htm>).
8. И.М. Богуславский. Автоматический перевод. Большая российская энциклопедия. – (<https://bigenc.ru/linguistics/text/1799594>).
9. Андреева А. Д., Меньшиков И. Л., Мокрушин А. А. Обзор систем машинного перевода. Молодой ученый. – 2013. – №12. – С. 64–66.
10. Гудий К.А. От оригинала к переводу: проблема взаимодействия автора и переводчика // Филология и лингвистка в современном обществе: материалы Международной научной конференции (Москва, май 2012 г.) – Москва, 2012. – С. 99–103.
11. Машинный перевод. Энциклопедия Кругосвет – (<http://www.krugosvet.ru/enc/gumanitarnye_nauki/lingvistika/MASHINNI_PEREVOD.html>).