

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Оренбургский государственный университет»

И. В. Переходько

КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ В ПЕРЕВОДЕ

Учебное пособие

Рекомендовано ученым советом федерального государственного бюджетного образовательного учреждения высшего образования «Оренбургский государственный университет» для обучающихся по образовательной программе высшего образования по направлению подготовки 45.03.02 Лингвистика

Оренбург
2018

УДК 81'25(075.8)

ББК 81.184я73

П 27

Рецензент – профессор, доктор педагогических наук Л.В. Мосиенко

- Переходько, И. В.**
П 27 Компьютерные технологии в переводе: учебное пособие / И.В. Переходько; Оренбургский гос. ун-т. – Оренбург: ОГУ, 2018.
ISBN 978-5-7410-2208-5

Настоящее учебное пособие освещает ключевые теоретические проблемы внедрения компьютерных технологий в переводческую деятельность: лингвистические основы использования информационных технологий в переводе, историю развития машинного перевода, принципы и методы разработки систем машинного перевода, стратегии построения существующих систем, а также практические аспекты использования САТ – программ при переводе. В каждой главе содержатся вопросы и практические задания, направленные на актуализацию теоретических положений и развитие практических навыков работы с системами машинного перевода.

Учебное пособие «Компьютерные технологии в переводе» предназначено для обучающихся по образовательной программе высшего образования очной формы обучения направления подготовки 45.03.02 Лингвистика, изучающих французский язык, и направлено на организацию самостоятельной работы обучающихся.

УДК 81'25(075.8)

ББК 81.184я73

ISBN 978-5-7410-2208-5

© Переходько И.В., 2018

© ОГУ, 2018

Содержание

Введение.....	4
1 Теоретические предпосылки изучения использования компьютерных технологий в переводоведении	6
1.1 Лингвистические основы использования компьютерных технологий в переводе	6
1.2 История развития машинного перевода.....	28
2 Онлайн-словари и переводчики	68
3 САТ-программы	78
4 САТ-программа «Omega T»	87
5 Онлайн-программа «Wordfast Anywhere».....	96
Список использованных источников	101
Приложение А	104
Приложение Б.....	110

Введение

Учебное пособие «Компьютерные технологии в переводе» предназначено для организации самостоятельной работы обучающихся по образовательным программам бакалавриата очной формы обучения при освоении дисциплины «Компьютерные технологии в переводе» в рамках действующего учебного плана по направлению подготовки 45.03.02 Лингвистика.

Учебное пособие освещает ключевые теоретические проблемы внедрения компьютерных технологий в переводческую деятельность, а также практические аспекты использования CAT – программ.

Пособие не претендует на исчерпывающее описание всех направлений внедрения информационных технологий в переводческую деятельность. Задача пособия – познакомить обучающихся с основными тенденциями внедрения компьютерных технологий, а также с принципами работы систем автоматического и автоматизированного перевода.

Пособие предназначено для обучающихся, изучающих французский язык, и в практической части содержит аутентичные материалы на французском языке, предназначенные для формирования навыков письменного перевода с использованием CAT – программ.

Учебное пособие состоит из четырех разделов. В каждом разделе содержатся вопросы и практические задания, направленные на актуализацию теоретических положений и развитие практических навыков работы с системами машинного перевода.

В первом разделе «Теоретические предпосылки изучения использования компьютерных технологий в переводе» раскрываются лингвистические основы использования информационных технологий в переводе, исторические аспекты развития машинного перевода, принципы и методы разработки систем машинного перевода, стратегии построения

существующих систем. Практические задания раздела направлены на актуализацию теоритических знаний.

Во втором разделе «Онлайн-словари и переводчики» представлен обзор онлайн-программ поиска значения или перевода отдельных слов, дана классификация словарей, а также дано описание и алгоритм работы систем машинного перевода «Google Translate», «Яндекс. Переводчик», «PROMT», выявлены особенности работы данных программ. В разделе также представлены задания на развитие практических навыков работы с онлайн-словарями и переводчиками.

В третьем разделе «CAT-программы» описаны особенности работы и классификация программного обеспечения, поддерживающих программы Translation memory (память переводов), а также дан обзор наиболее используемых программ в переводческой деятельности. Практические задания направлены на отработку навыков работы с ТМ-программой «Trados»

В четвертом разделе дано описание программы «Omega T» и представлены задания на отработку алгоритма работы с данной программой.

В пятом разделе отрабатываются навыки работы с онлайн-программой «Wordfast Anywhere (WFA)».

1 Теоретические предпосылки изучения использования компьютерных технологий в переводоведении

1.1 Лингвистические основы использования компьютерных технологий в переводе

Несмотря на распространение знания иностранных языков, изучение их не может полностью обеспечить многосторонние и неуклонно расширяющиеся международные связи, т.к. в международные контакты вовлечено большое количество языков, при этом невозможно обеспечить достаточно высокий уровень владения иностранными языками при их массовом изучении. Исторический опыт свидетельствует о том, что прямое общение с помощью естественных классических языков (типа латыни) и искусственных языков (типа эсперанто) также не является выходом из положения.

Таким образом, возрастающая роль опосредованной коммуникации представляет собой исторически обусловленную закономерность. В связи с этим приобретает развитие письменный перевод информационных сообщений, передаваемых с использованием средств компьютерных коммуникаций, и в особенности его специфической разновидности - компьютерный перевод.

Отличительной чертой процесса опосредованной коммуникации является то, что в нем, помимо двух обычных для всякого коммуникативного процесса фаз (порождение исходного сообщения партнером А и восприятие текста партнером В), есть еще и промежуточная фаза (перекодирование с одного естественного языка на другой). Эта фаза именуется языковым посредничеством.

Следовательно, перевод может быть определен как:

– однонаправленный и двухфазный процесс межъязыковой и межкультурной коммуникации, при котором на основе подвергнутого

целенаправленному (переводческому) анализу первичного текста создается вторичный текст (метатекст), заменяющий первичный в другой языковой и культурной среде;

– процесс, характеризуемый установкой на передачу коммуникативного эффекта первичного текста, частично модифицируемой различиями между двумя языками, двумя культурами и двумя коммуникативными ситуациями [8].

Исследования рынка переводов показали, что объемы этого вида деятельности постоянно увеличиваются, а в составе переводов преобладают специальные, научно-технические переводы – до половины общего объема рынка переводов, затем идут устный, учебный, синхронный, художественный переводы.

Перевод представляет собой сложнейший вид деятельности и, естественно, состоит из множества операций, из которых часть является основными, а часть - вспомогательными. Говоря о письменном переводе, выделим основные операции:

- чтение и понимание текста-оригинала;
- перевод отдельных слов;
- буквальный перевод текста;
- эквивалентный перевод связного текста;
- словесное оформление связного текста адекватно коммуникативной ситуации.

Выделяют также вспомогательные операции:

- письменное фиксирование текста-оригинала и текста-перевода;
- поиск необходимых слов в словаре;
- проверка орфографии и грамматики текста-перевода;
- окончательное оформление текста-перевода: форматирование, стилизация и т.п.;
- сбор статистики о переведенном тексте: количество знаков, слов, строк и т.п.

Определенную часть из названных операций, как основных, так и вспомогательных, можно отнести к разряду рутинных. В то же время именно они отнимают значительную долю времени и сил переводчика. Часть операций при современном развитии информационных технологий достаточно легко автоматизируются, например, все вспомогательные операции: перевод отдельных слов и буквальный перевод текста из основных. Прочие операции автоматизируются в меньшей степени, но все же попытки более или менее успешные предпринимаются.

Соответственно разным уровням автоматизации операций перевода выделяются три основных подхода к толкованию понятия «компьютерный перевод» (узкий, расширенный и широкий) [8].

Узкий (машинный перевод, автоматический перевод) предполагает автоматический перевод текста с одного естественного языка на другой с использованием специального программного обеспечения - программ автоматического перевода текстов. При этом уровень эквивалентности перевода может быть совершенно различным: от компьютерного перевода одного слова, подстрочника, до практически полного смыслового соответствия текста-перевода тексту-оригиналу. Это зависит от следующих факторов:

- программное обеспечение (значительная часть программ-переводчиков не претендует на перевод связного текста, а оперирует только словами);
- текст (лексическая и грамматическая структура, сложность, образность, стиль).

Отсюда различаются уровни машинного перевода:

- перевод корневых морфов (программа исчисляет корневой морф и дает его перевод);
- перевод-подстрочник словоформ - буквальный перевод I уровня;
- перевод-подстрочник связного текста - буквальный перевод II уровня;

– смысловой перевод связного текста (как правило, определенной тематики и определенного стиля).

Узкий подход к компьютерному переводу малопродуктивен, так как из понятия, следовательно, и из процесса, исключаются вспомогательные операции, наиболее поддающиеся компьютеризации.

Расширенный (технологический) перевод предполагает использование компьютерных технологий для перевода слов и текстов с одного естественного языка на другой. Расширение предыдущего подхода осуществляется за счет принятия следующих положений:

– перевод не только текстов, но и отдельных слов, в таком случае используются не только программы-переводчики текстов, но и компьютерные словари;

– процесс перевода включает в себя множество вспомогательных операций, автоматизировать которые значительно проще, чем непосредственно перевод (набор текста, проверка орфографии, грамматики, оформление текста и т.п.).

Это толкование, действительно, вполне корректно назвать технологическим, так как при таком, исключительно прагматическом, подходе компьютерный перевод превращается в информационную технологию, а именно, совокупность методов, способов, приемов и средств выполнения информационной деятельности.

Широкий перевод предполагает использование компьютерных технологий для перевода слов и текстов с одного языка на другой. Расширение предыдущего понятия состоит в следующем: объект перевода - не только естественные, но и искусственные языки, точнее, языки программирования. При таком подходе под переводом понимается, в том числе и трансляция текста программы, например, с языка программирования высокого уровня на язык машинных кодов.

Широкий подход выходит за пределы коммуникационных систем типа «человек - человек». Это понятие наиболее пригодно специалистам в области информационных технологий, программистам.

Процесс перевода исследуется при помощи разработки различных теоретических моделей, описывающих процесс перевода с большей или меньшей степенью приближенности. Под моделью перевода понимается «условное описание ряда мыслительных операций, выполняя которые можно осуществить процесс перевода всего оригинала или некоторой его части». Модели перевода раскрывают отдельные стороны функционирования лингвистического механизма перевода и знание их может помочь в разработке систем компьютерного перевода.

В прямом значении модель направлена не на описание уже происшедшего и констатацию существующего положения вещей, а на представление существенных характеристик изучаемого объекта и прогнозирование их представленности в будущем тексте перевода. Модель должна включать описание того инварианта содержательности текста оригинала, который должен быть непременно представлен в переводе.

Существует несколько моделей перевода, преимущественно ориентированных на внеязыковую реальность или на некоторые структурно-семантические особенности языковых единиц. Среди них можно назвать такие модели перевода, как ситуативная (или денотативная), трансформационно-семантическая, психолингвистическая, теория закономерных соответствий и другие [7].

Одна из первых попыток создания полноценной теории перевода была предпринята в трудах русских ученых А. В. Федорова и Я. И. Рецкера. Они разработали лингвистическую теорию перевода, получившую название *теории регулярных соответствий*. В основу теории легло два основных понятия – переводческие соответствия и переводческие преобразования. Важнейшим достижением созданной ими теории является определение механизмов перевода, основанных на взаимоотношениях между

логическими понятиями. Это трансформации генерализации и конкретизации с метонимическим и антонимическим видами перевода.

Данная теория учитывает то обстоятельство, что некоторые приемы логико-семантического порядка в процессе перевода повторяются. Так, для перевода отдельных слов и сочетаний необходимо найти существующие в другом языке постоянные эквиваленты. Постоянные эквиваленты имеются во всех языках, например, для имен собственных, терминов, числительных. Для перевода некоторых других слов и словосочетаний можно найти только вариантные контекстуальные соответствия, т.е. временные эквиваленты, годные для данного контекста, для данного употребления слова или словосочетания в речи. И, наконец, для перевода еще одной группы слов или словосочетаний приходится прибегать к приемам логического мышления и связанным с ними лексическими трансформациями.

Таким образом, по мнению Я.И. Рецкера, между единицами текстов оригинала и перевода имеются «три категории соответствий»:

- эквиваленты – однозначные, постоянные соответствия;
- аналоги – соответствия, полученные с помощью выбора одного из синонимов;
- адекватные замены – соответствия, выбранные исходя из целого (текста).

Теория Я.И. Рецкера объясняет основные приемы работы переводчика. Она стимулирует изучение эквивалентов, объема значений слов и словосочетаний, показывает значение контекста и подсказывает наиболее перспективные лексические трансформации, к которым может прибегнуть переводчик для поиска соответствий к единицам исходного текста. В этом большая ценность теории Я.И. Рецкера. Вместе с тем данная теория не объясняет всего процесса перевода, поскольку ее автор ограничивается изучением части объекта науки о переводе — изучением единиц текстов оригинала и перевода [8].

В своем последующем развитии теоретическая мысль не отвергла теорию регулярных соответствий. Сохраняя идею установления соответствия, каждая последующая модель наполняла ее иным содержанием.

Ситуативная (денотативная) модель исходит из положения о том, что любая ситуация может быть в принципе передана средствами любого языка. Предполагается, что процесс перевода происходит следующим образом: поняв содержание оригинала, переводчик определяет, какая ситуация в нем описана, а затем описывает эту ситуацию средствами языка перевода.

Таким образом, процесс перевода осуществляется в два этапа:

- от текста оригинала к действительности;
- от действительности к тексту перевода.

Обращение к действительности в рамках ситуативной модели перевода имеет в виду не только уяснение содержания оригинала, но и сам процесс перевода, т.е. путь, следуя которому, переводчик может создать текст перевода. «Отрезок» реальной действительности, отраженный в совокупном содержании исходного текста, служит внеязыковой основой перевода.

Ситуация в значительной мере определяет выбор варианта перевода и тогда, когда в языке перевода существует и единственный способ описания ситуации, и преобладающий, наиболее распространенный способ ее описания. Решающую роль играет обращение к описываемой ситуации в тех случаях, когда содержащаяся в высказывании информация недостаточна для выбора варианта перевода. В этом случае переводчик обращается к описываемой ситуации, ища необходимые сведения. Они могут быть получены из различных источников, таких как энциклопедии и справочники. Также переводчик может получить необходимую информацию, обратившись к специалисту в данной области.

Таким образом, ситуативная модель перевода правильно отражает ряд важных сторон переводческого процесса и дает возможность объяснить те особенности выбора варианта перевода, которые связаны с обращением

переводчика к действительности и последующим её описанием средствами языка перевода.

Ситуативная модель перевода не работает в тех случаях, когда переводчику необходимо отказаться от описания той же самой ситуации, чтобы обеспечить передачу цели коммуникации оригинала. Если данная ситуация не позволяет адресату перевода сделать необходимые выводы или связана с иными ассоциациями, чем у адресата оригинала, то описание переводчиком той же ситуации средствами языка перевода не обеспечит возможности межъязыковой коммуникации, а значит, перевод не будет адекватным.

Трансформационно-семантическая модель перевода, в отличие от ситуативной, исходит из предположения, что при переводе осуществляется передача значений единиц оригинала. Она рассматривает процесс перевода как ряд преобразований, с помощью которых переводчик переходит от единиц исходного языка к единицам переводящего языка, устанавливая между ними отношения эквивалентности. Таким образом, трансформационно-семантическая модель ориентирована на существование непосредственной связи между структурами и лексическими единицами оригинала и перевода.

Согласно данной модели процесс перевода проходит три этапа:

– I этап: анализ. Осуществляется упрощающая трансформация исходных синтаксических структур в пределах исходного языка, структуры оригинала преобразуются (сводятся) к наиболее простым, легко анализируемым формам. Предполагается, что такие простые «ядерные» (или «около-ядерные») структуры в разных языках достаточно близки и легко заменяют друг друга при переводе. На стадии анализа упрощающим преобразованиям подвергаются и отдельные слова, в значениях которых выявляется набор элементарных смыслов (сем). Подобные элементы смысла выделяются в семантике слова при его сопоставлении со словами близкими по значению и обнаружении различий между ними. У членов такого

семантического ряда можно найти как общие элементы смысла, так и дифференциальные, отличающие значения каждого синонима от других членов ряда. Аналогичным образом выявляется элементарный смысл у любых слов с пересекающимися или смежными значениями или у слов, значения которых связаны отношением «род-вид».

– II этап: «переключение», т.е. переход к ядерным структурам и семантическим компонентам языка перевода. Как было сказано, на уровне таких структур и элементарных сем у разных языков обнаруживается значительное сходство, поэтому в принципе эквивалентные единицы на этом уровне отыскиваются сравнительно легко.

– III этап: «реструктурирование». Осуществляются трансформации на переводящий язык с ядерного («околоядерного») уровня в окончательные структуры и единицы оригинала. При этом в соответствии с нормами языка меняются такие формальные признаки, как порядок слов, структура предложения, число и распределение семантических признаков.

Трансформационно-семантическая модель перевода обладает значительными возможностями и объяснительной силой. Она позволяет описывать многие стороны переводческого процесса, недоступные для непосредственного наблюдения. Особенно важно, что данная модель может показать роль языковых единиц в содержании исходного текста и зависимость от этих единиц средств переводящего языка, используемых в переводе. Тем самым моделируются способы достижения эквивалентности, при которых сохраняется основное значение синтаксических структур и лексических единиц исходного текста [16].

Вместе с тем, очевидно, что и трансформационно-семантическая модель не является универсальной и не претендует на моделирование любого переводческого акта. Она не предусматривает тех случаев, когда между синтаксическими структурами и значениями лексических единиц в оригинале и переводе нет отношений трансформации, и эквивалентность

двух текстов основывается исключительно на общности описываемой ситуации.

Не моделирует трансформационно-семантическая модель процесса перевода и тогда, когда в переводе необходимо передать образные и иные ассоциации, связанные с текстом оригинала, когда в процессе перевода происходит замена ситуации чтобы воспроизвести цель коммуникации.

С позиций компьютерного перевода эти теории наиболее полно реализуются технологиями структурных языков программирования и технологиями баз данных.

Современная теория перевода в качестве отправной точки исходит из того, что перевод, как и язык, является средством общения. Отсюда название этой теории – *коммуникативная модель перевода*. В основе теории было положено понятие коммуникативного акта. Слово коммуникация происходит от лат. «*communico*» - делаю общим, связываю, общаюсь. Под коммуникацией в человеческом обществе подразумевают общение, обмен мыслями, знаниями, чувствами, схемами поведения и т.п. Коммуникацией традиционно принято называть обмен значениями (информацией) между индивидами через посредство общей системы символов (знаков), языковых знаков.

Истоки коммуникативной модели перевода восходят к трудам лейпцигской школы переводоведения, исследованиям О. Каде и А. Нойберга, а также русских ученых В.Н. Комиссарова и А.Д. Швейцера. Коммуникативная модель перевода опирается на понятие модели коммуникации, которое разрабатывали математики [9]. Так, достаточно долгое время в лингвистике пользовались расширенной моделью коммуникативного акта, перекочевавшей из математики и кибернетики (прародителей информатики), моделью коммуникации, предложенной американским математиком Клодом Шенноном. Эта модель сыграла значительную роль в развитии многих наук, связанных с обменом информацией.

Модель включает пять элементов, расположенных в линейной последовательности (линейная модель):

- источник информации;
- передатчик;
- канал передачи;
- приемник;
- и конечную цель.

В дальнейшем модель пересматривалась с тем, чтобы удовлетворить потребности других областей исследования, связанных с другими видами коммуникации.

Пересмотренная модель включала шесть компонентов:

- источник;
- кодирующее устройство;
- сообщение;
- канал;
- декодирующее устройство;
- приемник.

Помимо этих терминов, Клод Шеннон ввел еще понятие шума (энтропия / негэнтропии) и избыточности. Энтропия (шум) в теории коммуникации связана с теми внешними факторами, которые искажают сообщение, нарушают его целостность и возможность восприятия приемником. Негэнтропия (отрицательная энтропия) связана с теми случаями, когда неполное или искаженное сообщение все же получено приемником благодаря его способности распознать сообщение, несмотря на искажения и недостающую информацию.

Понятие избыточности, повторения элементов сообщения для предотвращения коммуникативной неудачи, то есть, средства против энтропии, чаще всего демонстрируют именно на примере естественных человеческих языков. Считается, что все языки приблизительно наполовину избыточны: можно залить кляксами половину слов текста или стереть

половину слов в радиовыступлении, но при этом все же сохраняется возможность понять их. Разумеется, есть предел допустимого шума, за порогом которого возможность понимания резко снижается. В особенности трудно понимать в условиях шума сообщение, использующее малознакомый код.

Статичность модели Клода Шеннона была восполнена понятием обратной связи (feedback). Это понятие позволяло сделать модель более близкой к реальности человеческого взаимодействия в коммуникации. Его введение было связано с проникновением идей кибернетики, в частности одноименной работы Норберта Винера. Модель стала более динамичной. Для того, чтобы модель коммуникации в большей степени соответствовала потребностям в других областях, помимо телеграфной, выдвигались и другие динамические теории коммуникации. Например, психологом Теодором М. Ньюкомом была разработана более подвижная модель коммуникации, отражавшая взаимодействие участников коммуникативного акта, в особенности в отношении их когнитивного, эмоционального и артистического аспекта.

Ряд исследователей большее внимание обращал на канал передачи сообщения, например, известнейший канадский теоретик коммуникации, Маршалл Мак-Люэн, автор работ в области массовой коммуникации, для которого современная массовая коммуникация являлась коммуникацией по преимуществу визуальной. Он утверждает, что канал передачи во многом предопределяет и само сообщение. Его идеи во многом опередили свое время, и сейчас, в эпоху глобального телевидения и компьютерных сетей, находят самый широкий отклик. Слова М.Мак-Люэна: *The medium is the message*, – стали девизом современной цивилизации, в которой визуальный канал коммуникации считается ведущим [2].

В лингвистике идеи Клода Шеннона проявились в интерпретации Р.О. Якобсона. В модели коммуникации или речевого события, по Р.О. Якобсону, участвуют адресант и адресат, от первого ко второму

направляется сообщение, которое написано с помощью кода, контекст в модели Якобсона связан с содержанием сообщения, с информацией, им передаваемой, понятие контакта связано с регулятивным аспектом коммуникации.

Модель Р.О. Якобсона в различных ее вариантах применяется в лингвистике как для анализа функций языка в целом, так и для анализа функционирования отдельных его единиц, производства речи и текста. В духе лингвистики Пражской школы, эта модель телеологична (от греческого слова, обозначающего «цель»), то есть показывает предназначение, функции языка.

Функции коммуникации могут быть настроены над моделью коммуникации и «прикреплены» к ее участникам и элементам. Из модели Р.О. Якобсона вытекают шесть функций: эмотивная, коннотативная, референтивная, поэтическая, фатическая, метаязыковая. Каждая из предложенных им функций связана с тем или иным участником или элементом коммуникации.

Некоторые исследователи (А.А. Леонтьев, Н.Б. Мечковская) добавляют также магическую или заклинательную функцию, этническую (объединяющую народ) функцию, биологическую функцию (для коммуникации животных). Другие исследователи предпочитают минимизировать количество функций, выделяя лишь основные и считая другие разновидностью основных. Так, известный психолог и лингвист Карл Бюлер выделял три функции языка, проявляющиеся в любом акте речи: функцию выражения (экспрессивную), соотносимую с говорящим, функцию обращения (апеллятивную), соотносимую со слушающим, и функцию сообщения (репрезентативную), соотносимую с предметом речи [16].

Современная социолингвистика, теория коммуникации и социология коммуникации также заимствовали модель Р.О. Якобсона для описания коммуникативных процессов. В отличие от швейцарского лингвиста, основателя структурной лингвистики, Фердинанда де Соссюра,

предлагавшего изучать языковую систему «в себе и для себя», эта модель может учитывать не только сам язык, но и пользователя языка, включенного наблюдателя.

Идею наблюдателя ввел в научный обиход Нильс Бор, автор принципа дополнительности в физике и в науке вообще. Этот принцип можно сформулировать следующим образом: нет антагонистических противоречий, описания разных наблюдателей дополняют друг друга. Примечательно, что идею дополнительности Нильс Бор выдвинул первоначально на примере разнообразия языков мира.

В последнее время в науке распространяются идеи философии диалогизма М.М. Бахтина. Две основные идеи М.М. Бахтина весьма существенны и для понимания процесса коммуникации.

Во-первых, необходимым признаком любого высказывания является его обращенность, адресованность, то есть, без слушающего нет и говорящего, без адресата нет и адресанта.

Во вторых, всякое высказывание приобретает смысл только в контексте, в конкретное время и в конкретном месте (идея хронотопа: от греческих слов, обозначающих «время» и «место») [8].

Близкие идеи высказывал и выдающийся французский семиолог Ролан Барт. По Ролану Барту, слово не имеет значения, слово – только возможность значения, получающее его в конкретном тексте. Более того, каждое новое прочтение текста создает новое значение, читающий как бы пишет свой собственный текст заново. Эти взгляды разрушают стройную и четкую, но все же примитивную картину «передачи и восприятия» информации в первоначальной модели коммуникации по Шеннону. Идеи М.М. Бахтина получили весьма широкое распространение в мировой лингвистике, литературоведении, теории коммуникации, и философии.

Французский исследователь болгарского происхождения Юлия Кристева развивая идеи М.М. Бахтина и Ролана Барта, предложила понятие

интертекстуальности: всякий текст создается в виде «цитатной мозаики», прямых или косвенных ссылок на ранее воспринятые чужие тексты.

В кибернетике также стали появляться теории более общего характера. Хайнц фон Фёрстер ввел в обиход выражение «кибернетика второго порядка». В отличие от собственно кибернетики как технической, по преимуществу, науки, кибернетика второго порядка, в первую очередь, обращает свой взгляд на человека, на то, как он мыслит. Фон Фёрстер подчеркивал диалогический характер коммуникации.

Взгляды фон Фёрстера оказали влияние на чилийского исследователя Умберто Матурану, который предложил идею консенсуального взаимодействия самоорганизующихся систем, которые он называет «автопоэтическими», т.е. «самотворящими». Умберто Матурана подвергает сомнению и сам термин «передача информации», ведь в реальном процессе языкового взаимодействия ничего никому не передается в прямом смысле, и «передача» информации – всего лишь неудачная метафора совместной деятельности, в результате которой возникает сходный отклик: более или менее близкое взаимное понимание чего-то иного.

Ойген Розеншток-Хюсси, философ, провозгласил тождество «грамматики языка и грамматики общества», т.к. считал, что каждое высказывание политическим актом и видел в языке, в общении путь к миру в обществе.

Увеличение количества средств расширяет и сферу коммуникации. Два противоположных взгляда на эти процессы принадлежат М. Мак-Люэну и Э. Сепиру. М. Мак-Люэн считал, что средства во многом определяют и само содержание сообщения. Он во многом предвидел развитие коммуникации в современном обществе и еще в 60-70-е годы XX века отдавал приоритет визуальной коммуникации. Он считал, что современная культура – визуальная по своей сущности, в противовес, например, культуре XIX – начала XX века, преимущественно письменной (печатной).

Э. Сепир, напротив, высказывал опасение «быть понятым слишком многими». С его точки зрения, это ставит под угрозу психологическую реальность образа расширенного «Я», противопоставленного «не-Я». Невозможность удержать сообщение в тех границах, на которые оно рассчитано, признавалось и платой за облегчение коммуникации. В то же время он признавал, что разнообразие языков и необходимость перевода является сдерживающим фактором коммуникации.

С целью более полного отражения психических процессов деятельности переводчика разрабатывается *психолингвистическая модель перевода*, которая рассматривает перевод как двухэтапный процесс [8]:

- преобразование переводчиком своего понимания содержания оригинала в свою внутреннюю программу;
- развертывание этой программы в текст перевода.

Однако данная модель не раскрывает подробно каждый из указанных этапов процесса перевода, не выявляет элементы содержания, которые сохраняются во внутренней программе, а также не объясняет механизм выбора необходимого пути реализации этой программы в тексте перевода[8].

Таким образом, можно сделать вывод, что современное понимание коммуникативного акта, в том числе посредством переводчика (человека или машины) выходит далеко за пределы установления соответствия языковых конструкций. Человеческий мозг по своей сложности и мощности на несколько порядков превышает современные информационные технологии. Что касается работы мозга по переводу с одного языка на другой - это сложнейшая интеллектуальная работа, весьма слабо реализованная в существующих программах. Проблемы автоматизированного перевода прежде всего связаны со сложностью задачи (многовариантность перевода, для правильности которого мозг человека учитывает весь свой предыдущий жизненный опыт). Только создание программ, обладающих зачатками искусственного интеллекта, позволит окончательно решить эту задачу. Для описания коммуникативного акта средствами компьютерных технологий

наиболее полно подходят системы искусственного интеллекта, но и этой теории недостаточно, поскольку многое остается за гранью формализации в области чувственного (аффективная или даже магическая функции коммуникации).

Говоря об информационном обмене на основе компьютерных технологий, можно выделить следующие особенности коммуникации в системе «человек – человек»:

- опосредованность процесса коммуникации (образуется система «человек – компьютер (машина) – человек»);
- новое состояние письменности (использование компьютерных технологий при передаче информации);
- стремление к стиранию языковых барьеров, которые на сегодняшний день являются едва ли не основным нетехническим препятствием расширения и интенсификации информационного обмена.

Вопросы:

1. Дайте определение машинному переводу в узком и широком смыслах данного понятия.
2. Каковы основные и вспомогательные операции при письменном переводе?
3. Что понимается под моделью перевода?
4. Какие модели перевода существуют? Приведите примеры.
5. Кто разработал одну из первых моделей перевода и в чем ее сущность?
6. В чем заключается процесс перевода согласно ситуативной модели перевода?
7. Перечислите этапы процесса перевода согласно трансформационно-семантической модели перевода.
8. На исследования какого ученого опирается коммуникативная модель перевода и в чем ее сущность?

9. Каковы сильные и слабые стороны описанных моделей перевода?

10. Охарактеризуйте особенности коммуникации в системе «человек – человек» исходя из специфики информационного обмена на основе компьютерных технологий.

Тест № 1

1. Продукт коммуникативной деятельности и её предметно-знакового продолжения:
 - a. язык
 - b. текст
 - c. речь
2. Традиционные жанры письменного перевода:
 - a. Библейские тексты, литературный перевод, специальный перевод
 - b. Библейские тексты, литературный перевод, политический перевод, военный перевод
 - c. литературный перевод, специальный перевод
3. Фактор, ограничивающий возможности машинного перевода:
 - a. стилистические условия
 - b. синтаксические структуры
 - c. грамматика иностранного языка
4. Условно принятое количество этапов процесса перевода:
 - a. 2 - де- и перекодирование
 - b. 3 - декодирование, кодирование и перекодирование
 - c. 4 - анализ, декодирование, кодирование, перекодирование
5. Этапы машинного перевода:
 - a. анализ, выделение маркеров, подбор эквивалентов и их построение в одну цепочку
 - b. дробление на семантические составляющие, подбор эквивалентов, перевод
 - c. анализ, трансфер, синтез

6. Модель перевода, наиболее релевантная процессу перевода:
 - a. ситуационная
 - b. многоуровневая
 - c. коммуникативная
7. Автор коммуникативной модели перевода:
 - a. В.Н. Комиссаров
 - b. Т. Ягер
 - c. Ю. Найда
8. Исходный и конечный пункт акта коммуникации:
 - a. иностранный язык и переводной язык
 - b. оригинал и перевод
 - c. адресант и адресат
9. Результат аналитико-синтетической обработки текста:
 - a. первичная информация
 - b. вторичная информация
 - c. акцентирование на ключевых моментах текста
10. На примере какой предметной области проводил исследования Г.П. Мельников?
 - a. кибернетика
 - b. информатика
 - c. лингвистика
11. Система «человек-машина» определяет значение:
 - a. информационно-технологических проблем
 - b. прагматических проблем
 - c. экстралингвистических проблем
12. Аналитико-синтетическая обработка текста включает:
 - a. получение информации
 - b. обмен её
 - c. оба варианта
13. Практические аспекты переводы изменяются в диапазоне:

- a. видов отношений между адресатом и адресантом
 - b. видов отношений коммуникации и среды
 - c. видов отношений среды и смысловой составляющей сообщения
14. Методика обеспечения стабильного качества перевода включает в себя:
- a. сборка финальной версии перевода, состоящей из сегментов нескольких разных переводов одного и того же оригинала
 - b. подбор переводчика по его билингвальной компетенции для выполнения конкретной работы и последующем контрольном редактировании
 - c. постредактирование
15. Неофициальный критерий оценки компетенции переводчика, вводимый многими лингвистами:
- a. опыт работы
 - b. кругозор
 - c. специализация
16. Наиболее проблемный жанр перевода:
- a. библейский
 - b. литературный
 - c. специальный
17. Особенность ситуационной модели перевода заключается в:
- a. подборе ситуационного эквивалентного параметра в другом языке при переводе
 - b. выдерживании стилистики оригинала
 - c. подборе наиболее точных лексических компонентов
18. Коммуникативная модель перевода включает в себя реалии дискурса и его анализ в терминах:
- a. лингвистики
 - b. семиотики
 - c. обеих наук
19. Среда функционирования текстовой информации - это:

- a. сложноструктурная системную среду, состоящую из ряда отдельных подсистем
 - b. аудитория
 - c. ситуация
20. Перевод с точки зрения лингвистической теории это:
- a. процесс преобразования речевого произведения на одном языке в текст на другом языке, а также результат этого процесса
 - б. процесс преобразования речевого произведения на одном языке в текст на другом языке
 - с. результат процесса, т.е. сам переведенный текст
21. Проблемами перевода с одного данного языка на другой данный язык занимается:
- a. общая теория перевода
 - б. частная теория перевода
 - с. процессуальная транслатология
22. Проблемами перевода текстов разных типов и жанров занимается:
- a. история перевода
 - б. частная теория перевода
 - с. специальная теория перевода
23. Теории принципиальной непереводаемости придерживались такие ученые как:
- a. В. Гумбольдт, Э. Сепир, Б. Уорф
 - б. Декарт, Лейбниц
 - в. Л.С. Бархударов, Я.И. Рецкер, В.Н. Комиссаров

Практическая работа № 1

1. Представьте схему последовательности формальных операций, обеспечивающих анализ и синтез в системе машинного перевода.

2. Заполните таблицу: «Основные теоретические модели перевода».

Модель перевода	Автор	Основные принципы	Достоинства	Недостатки

1.2 История развития машинного перевода

Впервые мысль о возможности автоматического перевода в начале 40-х годов XIX века высказал британский математик Чарльз Бэббидж. Он пытался убедить правительство в необходимости финансировать исследования по разработке механического прототипа ЭВМ и обещал, что его машина сможет переводить разговорную речь. Проекту Ч. Бэббиджа суждено было остановиться на стадии прототипа, и его идеи не нашли практического воплощения.

В 1933 году изобретатель П.П. Смирнов-Троянский получил в СССР патент на механическую «машину для подбора и печатания слов при переводе с одного языка на другой». Огромная заслуга этого человека в том, что он предложил и автоматический двуязычный словарь и схему кодирования межъязыковых грамматических соответствий, но только для «синтетического» языка эсперанто. «Лингвистический арифмометр» П.П. Смирнова-Троянского опередил время, но дошел до наших дней лишь в списке научных курьезов: расширить его функциональность для работы с естественными языками так и не удалось.

В 1939 году на Всемирной ярмарке в Нью-Йорке «Bell Labs» (бывшая американская, а ныне франко-американская, корпорация, крупный исследовательский центр в области телекоммуникаций) продемонстрировала первое электронное устройство синтеза речи, «прапредок» современных систем перевода «на лету». Практического значения «железный болтун» не получил, но вызвал огромный интерес.

Вплоть до конца 40-х годов XX века машинный перевод был скорее объектом увлекательных исследований, чем важной сферой использования вычислительной техники. Его история как научно-прикладного направления началась в 1947 году.

Теоретической основой начального периода работ по машинному переводу (конец 40-х – начало 50-х годов XX века) была идея о том, что язык – это кодовая система. Пионерами машинного перевода были математики и инженеры. Описания их первых опытов, связанных с использованием только что появившихся ЭВМ для решения криптографических задач, были опубликованы в США в конце 40-х годов.

Датой рождения машинного перевода как научного направления принято считать 1946 год, когда Уоррен Уивер, директор отделения естественных наук Рокфеллеровского фонда, впервые сформулировал концепцию машинного перевода, которую в 1949 году развил в своем меморандуме «Translation», адресованном фонду. 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». Аналогия между переводом и дешифрованием была естественной в контексте послевоенной эпохи, если учитывать успехи, которых достигла криптография в годы Второй мировой войны [5].

Идеи У. Уивера легли в основу подхода к машинному переводу, основанного на концепции «Interlingva», согласно которой стадия передачи информации разделена на два этапа. На первом входное предложение следует перевести на искусственный язык-посредник (например, созданный на базе упрощенного английского языка), а затем результат этого перевода представить средствами выходного языка. Несмотря на то, что существует данная научная концепция уже полвека, ни одной реально работающей системы на ее основе пока не создано. Главная проблема - в разработке «посредника» и формальном описании его в терминах естественного языка.

Меморандум У. Уивера вызвал самый живой интерес к проблеме машинного перевода. В 1948 году А. Бут и Р. Риченс произвели некоторые предварительные эксперименты. Так, Р. Риченс разработал правила

разбиения словоформ на основы и окончания. Вскоре началось финансирование исследований. На ранних этапах разработка машинного перевода активно поддерживалась военными, при этом в США основное внимание уделялось русско-английскому направлению, а в СССР - англо-русскому направлению [10].

Помимо очевидных практических нужд важную роль в становлении машинного перевода сыграло то обстоятельство, что предложенный в 1950 году английским математиком А. Тьюрингом знаменитый тест на разумность («тест Тьюринга») фактически заменил вопрос о том, может ли машина мыслить на вопрос о том, может ли машина общаться с человеком на естественном языке таким образом, что тот не в состоянии будет отличить ее от собеседника-человека. Тем самым вопросы компьютерной обработки естественно-языковых сообщений на десятилетия оказались в центре исследований по кибернетике, а впоследствии по искусственному интеллекту, а между математиками, программистами и инженерами-компьютерщиками с одной стороны и лингвистами - с другой установилось продуктивное сотрудничество.

В 1952 году состоялась первая конференция по машинному переводу в Массачусетском технологическом университете, а в 1954 году в Нью-Йорке была представлена первая система машинного перевода - IBM Mark II, разработанная компанией IBM совместно с Джоржтаунским университетом. Это событие вошло в историю как Джоржтаунский эксперимент. Была представлена очень ограниченная в своих возможностях программ, она имела словарь в 250 единиц и 6 грамматических правил, осуществлявшая перевод с русского языка на английский.

В том же 1954 году был осуществлен первый эксперимент по машинному переводу в СССР И.К. Бельской (лингвистическая часть) и Д.Ю. Пановым (программная часть) в Институте точной механики и вычислительной техники Академии наук СССР. В январе 1954 года состоялась первая публичная демонстрация машинного перевода с русского

языка на английский, осуществленного на машине ИБМ-701. Первый опыт перевода с английского языка на русский с помощью машины БЭСМ был получен уже к концу 1955 года. Программы для БЭСМ составляли Н. П. Трифонов и Л. Н. Королев.

Первый промышленно пригодный алгоритм машинного перевода и система машинного перевода с английского языка на русский на универсальной вычислительной машине были разработаны коллективом под руководством Ю.А. Моторина. После этого работы начались во многих информационных институтах, научных и учебных организациях страны. Казалось, что создание систем качественного автоматического перевода вполне достижимо в пределах нескольких лет, при этом акцент делался на развитии полностью автоматических систем, обеспечивающих высококачественные переводы. Участие человека на этапе постредактирования расценивалось как временный компромисс.

К началу 50-х годов целый ряд исследовательских групп в США и в Европе работали в области машинного перевода. В эти исследования были вложены значительные средства, однако результаты очень скоро разочаровали инвесторов. Одной из главных причин невысокого качества машинного перевода в те годы были ограниченные возможности аппаратных средств: малый объем памяти при медленном доступе к содержащейся в ней информации, невозможность полноценного использования языков программирования высокого уровня. Другой причиной было отсутствие теоретической базы, необходимой для решения лингвистических проблем, в результате чего первые системы машинного перевода сводились к пословному переводу текстов без какой-либо синтаксической и смысловой целостности.

В 1959 году философ Й. Бар-Хиллел выступил с утверждением, что высококачественный полностью автоматический машинный перевод не может быть достигнут в принципе, так как для выполнения качественного перевода необходимо обладать знанием внеязыковой действительности, а это

знание слишком обширно и разнообразно, чтобы вводить его в компьютер. Однако Й. Бар-Хиллел не отрицал идею машинного перевода как таковую, считая перспективным направлением разработку машинных систем, ориентированных на использование их человеком-переводчиком [8].

Это выступление самым неблагоприятным образом отразилось на развитии машинного перевода в США. В 1966 году специально созданная Национальной Академией наук комиссия ALPAC (Automatic Language Processing Advisory Committee), основываясь, в том числе и на выводах Й. Бар-Хиллела, пришла к заключению, что машинный перевод нерентабелен, т.к. соотношение стоимости и качества машинного перевода было явно не в пользу последнего, а для нужд перевода технических и научных текстов было достаточно человеческих ресурсов. За докладом ALPAC последовало сокращение финансирования исследований в области машинного перевода со стороны правительства США - и это несмотря на то, что в то время как минимум три различные системы машинного перевода регулярно использовались рядом военных и научных организаций, в числе которых ВВС США, Комиссия США по ядерной энергии, Центр Евратома в Италии.

Следующие десять лет разработка систем машинного перевода осуществлялась в США университетом Brigham Young University в Прово, штат Юта и финансировалась Мормонской церковью, заинтересованной в переводе Библии. В Канаде разработкой систем машинного перевода занималась группа исследователей, в том числе группа «TAUM» в Монреале, которая добилась наилучшего результата, разработав систему «МЕТЕО», выполняющую задачу англо-французского перевода сводок погоды; в Европе - группами «GENA» (Гренобль) и «SUSY» (Саарбрюкен) [10].

Особого упоминания заслуживает работа в этой области отечественных лингвистов, таких, как И.А. Мельчук и Ю.Д. Апресян (Москва), результатом которой стал лингвистический процессор «ЭТАП». В 1960 г. в составе Научно-исследовательского института математики и механики в Ленинграде была организована экспериментальная лаборатория машинного перевода,

преобразованная затем в лабораторию математической лингвистики Ленинградского государственного университета.

Новый подъем исследований в области машинного перевода начался в 70-х годах и был связан с серьезными достижениями в области компьютерного моделирования интеллектуальной деятельности. Соответствующая область исследований, возникшая несколько позже машинного перевода, получила название искусственного интеллекта, а создание систем машинного перевода было осмыслено в 70-е годы как одна из частных задач этого нового исследовательского направления.

При этом несколько сместились акценты. Исследователи теперь ставили целью развитие «реалистических» систем машинного перевода, предполагавших участие человека на различных стадиях процесса перевода. Системы машинного перевода из «врага» и «конкурента» профессионального переводчика превращаются в незаменимого помощника, способствующего экономии времени и человеческих ресурсов.

Можно выделить два основных стимула к развитию работ по машинному переводу в современном мире. Первый - собственно научный. Он определяется комплексностью и сложностью компьютерного моделирования перевода. Как вид языковой деятельности перевод затрагивает все уровни языка - от распознавания графем (и фонем при переводе устной речи) до передачи смысла высказывания и текста. Кроме того, для перевода характерна обратная связь и возможность сразу проверить теоретическую гипотезу об устройстве тех или иных языковых уровней и эффективности предлагаемых алгоритмов. Эта характерная черта перевода вообще и машинного перевода в частности привлекает внимание теоретиков, в результате чего продолжают возникать все новые теории автоматизации перевода и формализации языковых данных и процессов.

Вместе с тем разработки в области машинного перевода стимулировали развитие не только лингвистики. Результаты работ по машинному переводу способствовали началу и развитию исследований и разработок в области

автоматизации информационного поиска, логического анализа естественно-языковых текстов, экспертных систем, способов представления знаний в вычислительных системах и т.д.

Второй стимул - социальный, и обусловлен он возрастающей ролью самой практики перевода в современном мире как необходимого условия обеспечения межъязыковой коммуникации, объем которой возрастает с каждым годом. Другие способы преодоления языковых барьеров на пути коммуникации - разработка или принятие единого языка, а также изучение иностранных языков - не могут сравниться с переводом по эффективности. С этой точки зрения можно утверждать, что альтернативы переводу нет, так что разработка качественных и высокопроизводительных систем машинного перевода способствует разрешению важнейших социально-коммуникативных задач.

Одной из новых разработок этого периода стала технология ТМ (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. А в конечном итоге не нужно дважды переводить одно и то же предложение. В настоящее время разработчиком известной коммерческой системы, основанной на технологии ТМ, является система «TRADOS», основанная в 1984 году [17].

В СССР с середины 70-х годов были созданы промышленные системы машинного перевода с английского языка на русский «АМПАР» на основе исследований и разработок коллектива Ю.А. Моторина, с немецкого языка на русский «НЕРПА», с французского языка на русский «ФРАП»,

автоматические терминологические словари в помощь человеку-переводчику.

Система «АМΠΑР» длительное время находилась в промышленной эксплуатации. Впоследствии на ее базе были созданы более эффективные системы машинного перевода для персональных компьютеров семейства «СПРИНТ». Была также разработана система машинного перевода с русского языка на английский «АСПЕРА». На этих разработках основываются такие системы машинного перевода, как «Stylus», «Socrat» и другие.

В 90-х годах происходило бурное развитие рынка персональных компьютеров и информационных технологий. Сеть Интернет становится все более интернациональной и многоязыкой. Все это сделало возможным, а главное востребованным, дальнейшее развитие систем машинного перевода. Появляются новые технологии, основанные на использовании нейронных сетей, концепции коннекционизма, статистических методах.

С начала 90-х годов на рынок компьютерных технологий выходят отечественные разработчики. В июле 1990 года на выставке «PC Forum» в Москве была представлена первая в России коммерческая система машинного перевода под названием «ПРОМТ» (PROgrammer's Machine Translation). В 1991 году было создано ЗАО «ПРОект МТ», и уже в 1992 году компания «ПРОМТ» выиграла конкурс NASA на поставку систем машинного перевода. «ПРОМТ» была единственной неамериканской фирмой на этом конкурсе [12].

Сегодня специалисты связывают построение адекватных систем машинного перевода с развитием искусственного интеллекта: машина сможет переводить с одного языка на другой, когда научится думать, как человек. Другой путь совершенствования машинного перевода, более доступный на современном этапе, - составить корпус соответствий на двух языках.

Большинство средств компьютерного перевода построены на основе концепции памяти перевода (translation memory) — простой базы данных,

каждая запись которой представляет собой единицу (предложение или абзац) параллельных текстов (как правило, на двух языках). Такая база данных хранит предыдущие переводы с целью их возможного повторного использования и решения задач быстрого поиска по содержимому.

Несмотря на то, что программы, оснащенные памятью перевода, называются системами автоматизированного перевода (CAT, computer-aided/assisted translation), их не следует путать с программами машинного перевода (machine translation) — память перевода ничего не переводит сама по себе, в то время как машинный перевод основан на генерации переводов по результатам анализа исходного текста.

Вопросы.

1. Когда впервые была высказана мысль о возможности автоматического перевода и нашла ли она свое практическое воплощение?
2. Когда были предприняты первые попытки перевода с помощью ЭВМ в России?
3. Какой год принято считать датой рождения машинного перевода?
4. Какое событие вошло в истории как Джорджтаунский эксперимент?
5. Когда состоялась первая публичная демонстрация машинного перевода в России?
6. Что поспособствовало сокращению исследований в области машинного перевода?
7. В чем причина неудач разработок систем машинного перевода на начальном этапе?
8. Когда начался новый подъем исследований в области машинного перевода и чем он обусловлен?
9. Каковы стимулы к развитию работ по машинному переводу?
10. В чем особенности развития исследований в области машинного перевода на втором этапе?

11. Какие положения лежат в основе развития систем машинного перевода на современном этапе?

12. Исследователи считают, что причины появления и развития идеи машинного перевода лежат в технической, политической и социальной областях. Поясните каждую из причин.

13. Как вы можете объяснить связь процесса машинного перевода и дешифровки текстов?

14. Охарактеризуйте этапы развития машинного превода. Какую роль в развитии идеи машинного превода сыграл американский ученый У. Уивер?

15. Какую роль человек может играть в процессе машинного перевода? Что такое предредактирование и постредактирование?

16. В чем, на ваш взгляд, заключается будущее машинного перевода?

Тест № 1

1. Ученые из каких стран были первопроходцами в вопросах машинного перевода?

- а. США, СССР
- б. США, Германия
- в. Германия, СССР

2. Как назывался первый опыт практики машинного перевода?

- а. Кейптаунский эксперимент
- б. Джорджтаунский эксперимент
- в. Стэнфордский эксперимент

3. Первый в мире машинный перевод с английского языка на русский был выполнен на базе:

- а. Академии наук СССР
- б. Национальной академии наук США
- в. Мичиганского университета

4. Ниже даны две причины, по которым работы по использованию ЭВМ в переводе начались с проектирования и создания систем МП, а не автоматических словарей. Выберите лишнее:

- а. Существовала общая недооценка уровня сложности перевода как интеллектуальной деятельности.
- б. Технические возможности вычислительной техники в то время не соответствовали грандиозности поставленной задачи.
- в. Была уже разработана теория создания систем МП, и ее необходимо было применить на практике.

5. Чем завершился второй период развития машинного перевода (после 1957)?

- а. С. Н. Разумовский, Н. П. Трифонов, Л. Н. Королев составили систему правил перевода, вошедших в основу современных программ машинного перевода.
- б. СССР пришел к выводу об экономической невыгодности машинного перевода и о недостижимости удовлетворительного результата.
- в. Национальной академией наук США был опубликован отчет по поводу экономической целесообразности машинного перевода и перспектив научных исследований в данной области.

6. Чем характеризуется современный этап развития машинного перевода?

- а. определением задач и путей развития машинного перевода;
- б. разработкой теоретических проблем, созданием компонентов систем МП и переходом к разработке и внедрению их промышленных образцов;
- в. прекращением противопоставления ЭВМ человеку.

7. Быстрый рост объема научно-технической и деловой документации на разных языках и необходимость оперативного перевода значительной ее части обусловил:

- а. переход к диалогу человека и ЭВМ, в котором последняя играла подчиненную роль;

- б. попытку замены человека-переводчика машиной, работающей в автоматическом режиме;
 - в. попытку замены человека-переводчика машиной, работающей совместно с человеком-редактором.
8. Влияние какого из факторов на развитие машинного перевода и вычислительной лексикографии недооценивается некоторыми лингвистами?
- а. быстрый рост объема научно-технической и деловой документации на разных языках и необходимость оперативного перевода значительной ее части;
 - б. невозможность выполнить эту работу традиционными способами;
 - в. возросшие возможности вычислительной техники.
9. Что из нижеперечисленного не является атрибутом компьютерных словарей?
- а. автоматическое устранение опечаток, распознавание текста, набранного транслитом;
 - б. возможность поиска слова в его текстовой форме;
 - в. поиск словосочетаний и устойчивых выражений.
10. Использование автоматических переводных может ускорить процесс перевода на:
- а. 60%
 - б. 40%
 - в. 15%
11. Создание доступных компьютеров привело к появлению нового направления в разработке автоматических переводных словарей в помощь человеку:
- а. словарей фразеологизмов;
 - б. словарей индивидуального пользования;
 - в. мультязычных словарей.

12. Создание в 60-х автоматических переводных словарей объемом в сотни тысяч лексических единиц было бы невозможным в 50-х по причине
- а. недостаточного спроса на их использование;
 - б. недостаточного уровня развития вычислительной техники;
 - в. неудобства их использования.
13. Как называется разработанная в 80-х концепция перевода, основанная на использовании большого массива текстов для поиска в нем аналогов фрагментов, которые могут быть использованы для формирования выходного текста?
- а. Example based translation
 - б. Translation Memory
 - в. Corpus linguistics translation
14. Новые информационные технологии в переводе предполагают использование компьютера как:
- а. замену печатной машинке;
 - б. средства реализации специализированного программного обеспечения;
 - в. средства обеспечения доступа к информации.
15. К основным компонентам современной информационной технологии в переводческом бизнесе следует прежде всего отнести:
- а. информационно-справочную базу, а также средства накопления, архивирования, поиска и восстановления фрагментированных переводов;
 - б. использование текстовых редакторов и программ по распознаванию текстов;
 - в. возможность осуществления чернового перевода онлайн-переводчиками.
16. Борьба за конкурентное преимущество в современном переводческом бизнесе может осуществляться за счет:
- а. условий и сроков предоставления заказчику готового продукта;
 - б. специфики трудовых ресурсов;

- в. использования новых специфических технологий переводческого процесса.
17. Что недооценивали первые ученые, выдвинувшие идею реализации машинного перевода на практике?
- а. объем памяти, которую будут занимать электронные базы данных;
 - б. сложность и огромную трудоемкость поставленной задачи;
 - в. невозможность применить машинный перевод на практике.
18. После 1957 г. внимание исследователей переключилось с лексики на:
- а. синтаксис;
 - б. семантику;
 - в. прагматику.
19. Промышленная система машинного перевода с английского языка на русский, созданная в СССР, называлась
- а. АМПАР
 - б. НЕРПА
 - в. ФРАП
20. В настоящее время разработчиком известной коммерческой системы, основанной на технологии ТМ, является
- а. Google
 - б. IBM
 - в. Trados

Практическая работа № 1

1. Заполните таблицу: «История развития систем машинного перевода».

<i>Этап</i>	<i>Период</i>	<i>Характеристика этапа</i>	<i>Основные события</i>	<i>Итоги</i>
<i>Первый этап развития систем машинного перевода</i>				
<i>Второй этап</i>				

<i>развития машинного перевода.</i>				
<i>Третий этапа развития машинного перевода.</i>				

1.3 Системы машинного перевода

Компьютерный перевод, также называемый машинный или автоматический – это, согласно оксфордскому словарю - перевод, осуществляемый при помощи компьютера». Иными словами – это процесс, который использует двуязычные данные, включающие в себя лексику и грамматику обоих языков, а также модели фраз, используемые в данных языках. Результатом этого процесса является перекодирование текста одного естественного языка на другой.

Лингвистическая обработка данных с использованием компьютера за последние десятилетия продвинулась далеко вперед, и сегодня качество и стоимость машинного перевода пришли в соответствие со спросом и ожиданиями по срокам вывода продукта на рынок.

Машинный перевод дает возможность компаниям перевести информационные материалы, которые не были переведены прежде из-за финансовых и временных ограничений. Технологические решения по машинному переводу могут быть адаптированы к любому объему работ и любым требованиям заказчика, включая особые стили письма и грамматические правила, специфическую терминологию и словари, используемые компанией, особенные правила для формата чисел и дат и множество других деталей, содержащихся в материалах.

Примеры использования машинного перевода:

- предварительный перевод больших проектов и последующее их редактирование лингвистами для ускорения вывода продукта на рынок;
- переводы в режиме реального времени для служб клиентской поддержки или чат-приложений, часто задаваемые вопросы и интернет-форумы;
- поиск в большом объеме данных в учебных и юридических целях;
- внутренняя переписка;
- каталоги и другая организованная информация.

Преимущества машинного перевода:

- высокая скорость перевода (использование системы машинного перевода позволяет значительно сократить время, требуемое для перевода текстов);
- низкая стоимость перевода (прибегая к услугам профессиональных переводчиков, мы вынуждены платить деньги за каждую страницу перевода, однако часто необходимости в получении идеального перевода текста нет, а нужно быстро уловить смысл присланного письма или содержания страницы в Интернете);
- конфиденциальность (многие пользователи регулярно используют системы машинного перевода для перевода личных писем, ведь далеко не каждый человек готов отдать постороннему переводчику личную переписку или доверить перевод финансовых документов);
- универсальность (профессиональный переводчик, как правило, имеет специализацию по переводу текстов определенной тематики, тогда как программа-переводчик справится с переводом текстов из самых разных областей: для правильного перевода специализированных терминов достаточно подключить необходимые настройки) [3].

В настоящее время различают автоматизированный перевод (Computer Assisted Translation, CAT) и автоматический (машинный перевод) (Machine Translation, MT).

Если автоматизированный перевод (CAT) – это программное обеспечение, используемое человеком-переводчиком в процессе перевода для повышения производительности труда, то автоматический (машинный) перевод (MT) – это такая компьютерная технология, когда перевод с одного языка на другой выполняется компьютерной программой без участия человека.

Концепция **автоматизированного перевода** была предложена японским исследователем М. Нагао в 80-х годах XX века. Идея заключалась в следующем – при накоплении достаточно большой коллекции ранее переведенных фраз для конкретных типов текстов и в узких областях велика вероятность того, что большая часть последующих текстов будет аналогична уже переведенным вручную. Ярким представителем этого подхода являются инструменты CAT (ComputerAided-Translation tools), реализующие технологии автоматизированной поддержки перевода. К системам CAT относят те системы, которые обеспечивают работу на базе «памяти переводов» (Translation Memory), поэтому их часто называют ТМ-инструментами. Функция обращения к базам «памяти перевода» дает переводчику и компании несомненные преимущества:

- одинаковая лингвистическая единица либо предложение не переводится дважды, необходимо лишь откорректировать предыдущий перевод;

- перевод текста получается более терминологически и стилистически однородным, в случае если выполняет группа переводчиков [6] .

Изначально CAT-система пустая. База данных ТМ состоит из пар параллельно сопоставленных сегментов (чаще всего это предложение), базы формируются программно по ходу перевода и хранятся отдельно. CAT-программа при переводе производит обращение к базе, и если новый сегмент имеет точное или частичное соответствие в имеющейся накопленной базе переводов, то предлагается вариант перевода, если соответствия нет, то перевод идет вручную и вариант записывается в базу как новый сегмент.

Многими пользователями отмечается, что при использовании САТ-систем значительно повышается эффективность работы переводчика, работающего с типовыми специальными текстами, для которых уже накоплена богатая практика переводческих примеров, т. е. базы переводов.

Автоматизированный перевод — это широкое понятие, охватывающее широкий спектр простых и сложных инструментов, которые могут включать:

- программы для проверки правописания, которые могут быть встроены в текстовые редакторы или дополнительные программы;
- программы для проверки грамматики, которые также встраиваются в текстовые редакторы или дополнительные программы;
- программы для управления терминологией, которые позволяют переводчикам управлять своей собственной терминологической базой в электронной форме (FileMaker LogiTerm, MultiTerm, Termex, TermStar и т.п.);
- словари на компакт-дисках, одноязычные или многоязычные;
- терминологические базы данных, хранимые на компакт-дисках или подключаемые по Интернету (The Open Terminology Forum или TERMIUM);
- программы для полнотекстового поиска (или индексаторы), которые позволяют пользователю обращаться с запросами к ранее переведённым текстам или разного рода справочным документам (Naturel, ISYS Search Software и dtSearch);
- программы конкорданса, позволяющие находить примеры слов или выражений в употребляемом контексте в одноязычном, двуязычном или многоязычном корпусе, как битекст или память переводов (Transit NXT);
- программное обеспечение для управления проектами, которое позволяет лингвистам структурировать сложные переводческие проекты, передавать выполнение различных задач разным сотрудникам и наблюдать за процессом выполнения этих задач;
- программы управления памятью переводов (ТММ), состоящие из базы данных сегментов текста на исходном языке и их переводов на один или более целевых языков (Transit NXT).

При этом надо иметь в виду, что, хотя при использовании программы автоматизированного перевода уменьшается время перевода и увеличивается его качество, основную работу по переводу (анализ текста, выбор той или иной переводческой стратегии) выполняет человек. Система автоматизации перевода – это, прежде всего, инструмент, облегчающий работу переводчика.

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. Помимо такого порядка работы, существуют системы машинного перевода, требующие постоянного присутствия человека-переводчика, помогающего компьютеру делать перевод особенно сложных или неоднозначных конструкций. Машинный перевод с помощью человека применим в большей степени к текстам с ограниченным вокабуляром узкоограниченной тематики [8].

Автоматизированный перевод предполагает следующие формы взаимодействия:

- частично автоматизированный перевод: например, использование переводчиком-человеком компьютерных словарей.

- системы с разделением труда: компьютер обучен переводить только фразы жёстко заданной структуры, а всё, не уложившееся в схему, отдаёт человеку.

Экономичность использования машинного перевода с помощью человека - вопрос все еще спорный. Сами программы обычно достаточно дорогостоящи, а для работы некоторых из них требуется специальное оборудование. Предварительному и последующему редактированию необходимо обучаться. Создание и поддержание в рабочем состоянии баз данных - процесс трудоемкий и зачастую требует специальных навыков. Однако для организации, переводящей большие объемы текстов в четко-определенной тематической сфере, машинный перевод с помощью человека

может оказаться достаточно экономичной альтернативой традиционному человеческому переводу.

Наибольшей популярностью среди переводчиков пользуются такие программные решения, как SDL Trados, Wordfast, Déjà Vu, STAR Transit, Omega T. Кроме них существует большое количество как платных, так и бесплатных версий программ автоматизированного перевода, функционально программы похожи. Выходят демонстрационные бесплатные версии программ. Рынок CAT-программ постоянно развивается, параллельно накапливается множество специальных баз переводов [13].

Например, используя Dragon Naturally Speaking или VoiceReader, переводчик может произносить текст, а программа будет преобразовывать озвученные фрагменты в электронный текст в текстовом редакторе. Учитывая тот факт, что люди проговаривают текст гораздо быстрее, чем набирают его, использование подобных программ способно существенно ускорить работу переводчика, которому уже нет необходимости набирать все фрагменты перевода с помощью клавиатуры. По такой технологии функционируют, к примеру, лингвистические редакторы Wordfast Anywhere, Google Translator Toolkit, Logoport.

Автоматический (машинный) перевод – это вид машинного перевода, при котором в компьютер вводится текст на одном языке, этот текст обрабатывается и компьютер выводит этот же текст на другом языке. К сожалению, реализация такого вида автоматического перевода сталкивается с определенными препятствиями, которые еще предстоит преодолеть и на сегодняшний день, полностью автоматизированный машинный перевод текстов с обширной тематикой все еще является невыполнимой задачей.

При этом выделяют три вида систем машинного перевода:

- системы на основе грамматических правил (Rule-Based Machine Translation, RBMT);
- статистические системы (Statistical Machine Translation, SMT);
- гибридные системы.

Системы на основе грамматических правил (RBMT) – общий термин, который обозначает системы машинного перевода на основе лингвистической информации об исходном и переводном языках. Глобальной задачей такой технологии машинного перевода является формальное моделирование естественных языков как объектов, реализованных в определенной программной и аппаратной компьютерной среде, а также моделирование различных процедур анализа, сопоставления (трансфера) и синтеза лингвистических единиц и текстов для этих естественных языков. Такая задача требует реализации процедур:

- графематического (символьного) анализа;
- морфологического и лексического анализа и синтеза;
- синтаксического анализа и синтеза;
- семантических преобразований.

На основе всех этих данных исходный текст последовательно, предложение за предложением, преобразуется в текст на требуемом языке. Основным принцип работы таких систем — связь структур исходного и конечного текстов. Системы на основе грамматических правил производят анализ текста, который используется в процессе перевода. Такой подход к машинному переводу еще называют классическим.

Подобные системы состоят из двуязычных словарей и грамматик, охватывающих основные семантические, морфологические, синтаксические закономерности каждого языка. К компонентам типичной системы на основе грамматических правил следует отнести следующие:

- лингвистические базы данных: двуязычные словари, файлы имен, транслитерации, морфологические таблицы.
- модуль перевода: грамматические правила, алгоритмы перевода.

За историю машинного перевода как научного направления выделились три подхода к моделированию систем такого рода:

- системы пословного перевода;

– трансферные системы (Transfer) - преобразуют структуры входного языка в грамматические конструкции выходного языка;

– интерлингвистические системы (Interlingua) – используют промежуточный язык описания смысла.

Первые системы машинного перевода характеризуются стратегией прямого (пословного) перевода. Сущность этого подхода к построению машинного перевода заключается в том, что исходный текст на входном языке постепенно через ряд этапов преобразуется в текст выходного языка. Преобразования сводятся к тому, что слово (словосочетание) на входном языке заменяется на его словарный эквивалент на выходном языке. Понятно, что в системах первого поколения, использующих стратегию прямого перевода, нет необходимости моделировать функционирование языковой системы в целом. Для работы таких систем оказывается вполне достаточно правил словарных соответствий. В редких случаях проводится анализ контекста для перевода однословных выражений, опять-таки представленных в словаре системы [7].

Важно иметь в виду, что стратегия прямого перевода не делает различий между пониманием (анализом) и синтезом (порождением), поскольку они фактически исключены из преобразований по правилам словарных соответствий. Прямой перевод всегда привязан к конкретной паре языков.

Например, неоднозначность выражений входного языка разрешается только в той степени, в которой это оказывается необходимым для выходного языка. По временным рамкам системы первого поколения в основном создавались в период с конца 40-х до середины 60-х гг.

Существенная модификация стратегии прямого перевода обнаруживается в системах с трансфером — этапом межъязыковых операций, не сводимых только к замене лексем входного языка на словарные соответствия выходного языка. Наличие этапа трансфера предполагает построение «промежуточного» или «внутреннего» представления, которое

далее «приспосабливается» к структуре предложения выходного языка. В отличие от первой стратегии, в архитектуре систем машинного перевода с трансфером анализ (понимание) и синтез существуют как особые процедуры и обслуживаются различными алгоритмами.

Основным достоинством систем на основе трансфера является высокая полнота охвата текстов при приемлемом уровне качества перевода, а также низкий уровень затрат на первичную разработку и модернизацию.

Развитие идеи трансфера привело к появлению перевода, основанного на глубинном лингвистическом анализе. Данная стратегия подразумевает анализ входного текста на всех языковых уровнях (морфологическом, синтаксическом, семантическом, прагматическом), а также многоуровневый синтез выходного текста.

Критика стратегии прямого перевода привела к созданию стратегии языка-посредника (интерлингвы). Главная особенность этой стратегии заключается в том, что между структурами входного языка и структурами выходного языка находится один или несколько промежуточных языков, на которые по соответствующим правилам последовательно «переписываются» выражения входного языка. Анализ и синтез при использовании языка посредника принципиально разделяются. Анализ ведется в категориях входного языка, а синтез — в категориях выходного. В качестве языка (языков) - посредников могут выступать языки представления синтаксической и семантико-синтаксической структуры, чисто семантические языки, языки глубинной семантики, приближающиеся к концептуальному представлению в категориях теории знаний (фреймов, сценариев, планов). Интерлингвы требуют очень долгой разработки и создания огромных баз знаний о языке. Примером таких систем является система «ZARDOZ», ориентированная на ряд языков, в частности, американский, ирландский, японский языки.

Преимуществами систем на основе грамматических правил являются грамматическая и синтаксическая точность, стабильность результата,

возможность настройки на специфическую предметную область. К недостаткам систем на основе грамматических правил относят необходимость создания, поддержки и обновления лингвистических баз данных, трудоемкость создания такой системы, а так же ее высокая стоимость.

Статистические системы (SMT) – системы, которые строятся на использовании больших объемов параллельных корпусов текстов, с помощью которых система учится, выявляя закономерности при переводе.

Таким образом, статистический машинный перевод использует свойство «самообучения» языку (machine learning), т.е. чем больше накоплено параллельных текстов и чем точнее они соответствуют друг другу, тем лучше результат статистического машинного перевода. Статистический машинный перевод - разновидность машинного перевода, где перевод генерируется на основе статистических моделей, параметры которых являются производными от анализа двуязычных корпусов текста (text corpora).

Считается, что интерес к использованию лингвостатистических методов в машинном переводе возник еще в конце 40-х годов XX века параллельно с развитием теории информации К. Шеннона. В 60-х годах XX века исследователи успешно применяли лингвостатистику к решению задач дешифровки древних текстов. Но активно это направление стало развиваться только в начале 90-х годов XX века во время технологического и информационного бума, накоплением данных в корпусной лингвистике, машинном обучении и информационном поиске. Поэтому глобальные поисковые системы стали внедрять сервис статистического перевода в свои интерфейсы. Например, поисковик Google перешел на эту технологию в 2007 году и предлагает сервис Google Translate.

В начале 2011 года Яндекс, как ранее Google, внедрил собственную подобную систему машинного перевода. Статистические системы при своей работе используют статистический анализ. В систему загружается

двухязычный 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. [1].

В процессе перевода статистическая система подбирает наиболее вероятный вариант, используемый с точки зрения частотности. При статистическом методе компьютер делит текст языка оригинала на N-граммы, которые могут быть, как отдельными словами, так и словосочетаниями, и предложениями, после чего из огромного количества проанализированных параллельных текстов, компьютер выбирает варианты перевода данных N-грамм, составляет из них множество вариантов перевода исходного текста и выбирает один наиболее употребительный вариант, основываясь на проанализированных параллельных текстах [Молчанов А.]. С каждым новым переведенным текстом улучшается качество последующих переводов.

Таким образом, первые идеи статистического машинного перевода были опубликованы Уорреном Уивером в 1949 году. «Вторая волна» развития систем статистического машинного перевода приходится на начало 1990-х в компании IBM, «третья волна» (Google, Microsoft, Language Weaver, Яндекс).

Выделяют следующие статистические модели перевода:

- по словам (Word-based translation - WBT);
- по фразам (Phrase-based translation - PBT);
- по синтаксису (Syntax-based translation - SBT);
- по иерархическим фразам (Hierarchical phrase-based translation - HPBT).

Статистические системы отличаются быстротой настройки и легкостью добавления новых направлений перевода. Среди недостатков наиболее значительными являются наличие многочисленных грамматических ошибок и нестабильность перевода.

Гибридные системы сочетают в себе подходы, описанные ранее, что позволяет объединить все их преимущества. Статистический машинный перевод стремится использовать лингвистические данные, а системы с «классическим» подходом, основанном на правилах, применяют статистические методы. Добавление некоторых «сквозных» правил, то есть создание гибридных систем, несколько улучшает качество переводов, особенно при недостаточном объеме входных данных, используемых при построении индексных файлов хранения лингвистической информации машинного переводчика, базирующегося на N-граммах.

Гибридная технология перевода предполагает использование статистических методов для построения словарных баз автоматическим путем на основе параллельных корпусов, формирования нескольких возможных переводов, как на лексическом уровне, так и на уровне синтаксической структуры предложения выходного языка, применения постредактирования в автоматическом режиме и выбор лучшего (наиболее вероятного) перевода из возможных на основе языковой модели, построенной по определенному корпусу выходного языка.

В качестве примера можно привести архитектуру гибридной технологии «SMT и RBMT». RBMT-система дополнена двумя компонентами: модулем статистического постредактирования и модулем языковых моделей. Статистическое постредактирование позволяет сгладить RB-перевод, приближая его к естественному языку и при этом сохраняя четкую структуру синтезируемого текста.

Преимуществами гибридного перевода является быстрая автоматическая настройка на основе Translation Memories заказчика, терминологическая точность перевода, а также единство стиля и получение дополнительных полезных данных - двуязычного терминологического словаря.

Единственным минусом гибридных систем компьютерного перевода, с технологической точки зрения, является высокая сложность разработки и

настройки такой системы для новой пары языков, так как объема доступных параллельных текстов не хватает для того, чтобы обеспечить адекватный и эквивалентный вариант перевода. Поэтому был разработан компромиссный вариант, где для перевода использовался бы язык посредник – английский, количество параллельных текстов, с участием которого в значительной степени больше, чем параллельных текстов, например, для пары монгольский-испанский. Так, система сначала переводила текст с языка оригинала на английский, а затем с английского языка на язык перевода, таким образом, например, долгое время работала система Google Translate. Но в ноябре 2016 года корпорация Google представила свою обновленную систему компьютерного перевода – нейронный машинный перевод Google (Google's Neural Machine Translation System). С этого момента технология машинного перевода изменилась полностью. Благодаря развитию искусственного интеллекта и его глубокого обучения (deep learning) были пересмотрены сами подходы к процессу компьютерного перевода.

Основа, которая дала возможность развития систем компьютерного перевода – это технология машинного обучения. Машинное обучение – это способность компьютера учиться на опыте, т. е. модифицировать свою работу на основе недавно полученной информации.

Непосредственный переход от технологий гибридного машинного перевода к технологиям, основанным на глубоком машинном обучении, произошел в 2014 году, когда команда Kyung Hyun Cho осуществила прорыв, показав возможность перевода с одного языка на другой, без участия языка посредника, как это было в статистическом машинном переводе. Модель, предложенная в этой работе и реализованная в нейронном машинном переводе Google, использует систему кодировок, то есть преобразование информации в некую последовательность цифр. Система выполняет сохранение самого смысла, семантики, предложения, фразы, слова – N-граммы в его числовой эквивалент, а затем подбирает ему эквивалент на языке перевода, основываясь на частотности, на собственных ошибках,

сделанных ранее, на нормах и правилах языка перевода, усвоенных ей в ходе обучения на основе корпуса параллельных текстов.

Такой вид машинного перевода имеет ряд плюсов, таких как: ограниченность такого подхода к переводу лишь количеством анализируемых данных и вычислительной мощностью, ему предоставленной и независимость системы от правил языка, в том смысле, что системе не нужны исходные данные лингвистов о грамматических правилах используемых языков, она сама получает их, исходя из проанализированных корпусов параллельных текстов.

В настоящее время, только некоторые системы компьютерного перевода перешли на технологию нейросетей. Первой такой переход осуществила система Google, в конце 2016 года. Система Яндекс с 2017 года также использует нейросеть, но в отличие от первопроходца, система компьютерного перевода Яндекс производит сразу два варианта перевода – статистический и с помощью нейросети, а затем механизм Catboost, являющийся инновационным подходом к машинному обучению, выбирает наиболее подходящий вариант.

Лари Чайлдс в рамках Международной Конференции по Техническим Коммуникациям 1990 года предложил актуальную до сих пор классификацию систем машинного перевода:

– **FAMT (Fully-Automated Machine Translation)** - инструменты полностью автоматизированного машинного перевода.

– **HAMT (Human-Assisted Machine Translation)** - приложения для автоматизированного машинного перевода текстов, выполняемого при участии человек. Весьма наглядно такой тип систем машинного перевода иллюстрируется системами переводческой памяти (Translation Memory, TM).

– **MAHT (Machine-Assisted Human Translation)** - вспомогательные средства для выполнения перевода человеком с использованием компьютера. К категории MAHT-приложений сегодня относится абсолютное большинство

«электронных переводчиков» и компьютерных словарей - как программных, так и онлайн-овых.

При этом возможны следующие формы взаимодействия компьютера и человека:

– с постредктированием: исходный текст перерабатывается машиной, а человек-редактор исправляет результат;

– с предредктированием: человек приспособливает текст к обработке машиной (устраняет возможные неоднозначные прочтения, упрощает и размечает текст), после чего начинается программная обработка;

– с интерредктированием: человек вмешивается в работу системы перевода, разрешая трудные случаи;

– смешанные системы (одновременно с пред- и постредктированием).

Большое значение в работе систем машинного перевода играет лингвистическая база данных, которая включает в себя накопленные лингвистические данные, объективированные текстами, картотеками, словарями, грамматиками и другими лингвистическими источниками.

Типичный состав лингвистической базы данных можно ограничить следующими компонентами [2] :

– лексико-грамматический классификатор свойств исходного языка и переводного языка (система морфологического кодирования). При анализе исходного текста каждое слово в нем должно получить соответствующие морфологические характеристики: признак части речи, род, падеж, склонение, число и др. Система кодирования должна быть единой для конкретной системы машинного перевода;

– базовый двуязычный морфологический словарь. В этом словаре устанавливается пословное соответствие каждой словоформы исходного языка словоформам языка перевода;

– словарь сокращений и аббревиатур. Словарь используется на этапе разбиения исходного текста на слова и предложения. Сокращения и аббревиатуры должны быть по возможности расшифрованы, так как они

могут являться членами предложения, следовательно, их необходимо учитывать при синтаксическом и семантическом анализе;

- словарь идиом. Данный словарь применяется до синтаксического анализа, поскольку очень часто идиома является одним членом предложения и рассматривается как единое целое; при переводе идиома на исходном языке может соответствовать одному слову на переводном языке;

- терминологические словари по предметным областям.

Дополнительные словари подключаются при необходимости перевода специализированных текстов.

- синтаксический словарь. В данном словаре должна содержаться информация о синтаксической сочетаемости членов предложения как в языке оригинала, так и в переводном языке, а также синтаксические соответствия, необходимые при переводе.

- семантический словарь (тезаурус, онтология). Данный компонент содержит информацию о семантической сочетаемости лексем, о лексикосемантических полях, применяется на этапе построения семантического графа предложения.

- корпус параллельных текстов. Корпус содержит тексты на языке оригинала и их переводы на другой язык. При нахождении предложения или его фрагмента в корпусе параллельных текстов в текст перевода вставляется его соответствие на переводном языке. На использовании корпуса текстов построена технология памяти переводов.

Лингвистический процессор предназначен для полного лингвистического анализа текста на исходном языке, а также синтеза текста на языке перевода. Лингвистический процессор включает следующие компоненты [4]:

- программа разбиения текста на предложения и слова;
- программа распознавания устойчивых словосочетаний. Идиомы должны анализироваться и переводиться как неделимое целое;
- программа расшифровки сокращений и аббревиатур;

- программа морфологического аннотирования исходного текста;
- программа синтаксического анализа и построения дерева зависимостей;
- программа семантического анализа и построения семантического графа каждого предложения исходного текста;
- программа выбора переводного соответствия из двуязычного словаря или корпуса параллельных текстов;
- программа семантического синтеза текста на переводном языке;
- программа построения синтаксической структуры предложения и определения порядка слов в синтезируемом предложении;
- программа морфологического синтеза словоформ в переведенном тексте [2] .

Перевод, осуществляемый человеком с использованием компьютера. При этом подходе человек-переводчик ставится в центр процесса перевода, в то время как программа компьютера расценивается в качестве инструмента, делающего процесс перевода более эффективным, а перевод - точным. Это обычные электронные словари, которые обеспечивают перевод требуемого слова, возлагая на человека ответственность за выбор нужного варианта и смысл переведенного текста. Такие словари значительно облегчают процесс перевода, но требуют от пользователя определенного знания языка и затрат времени на его осуществление. И все же сам процесс перевода значительно ускоряется и облегчается.

Вопросы:

1. В каких областях и сферах используется машинный перевод?
2. Каковы преимущества машинного перевода?
3. Какие виды перевода различают на современном этапе?
4. В чем сущность автоматизированного перевода?
5. Какие инструменты включены в системы автоматизированного перевода?

6. Какова роль человека-переводчика при использовании автоматизированного перевода?
7. Каковы формы взаимодействия «человек – машина» предполагает автоматизированный перевод?
8. Какие сегодня программные продукты пользуются наибольшей популярностью?
9. В чем специфика автоматического перевода?
10. Перечислите виды машинного перевода и раскройте сущность их работы.
11. В чем основные достоинства и недостатки систем автоматического перевода?
12. Какую классификацию систем машинного перевода предложил Лари Чайлдс? Дайте характеристику систем в рамках данной классификации.
13. Какое значение в работе систем машинного перевода играют лингвистические базы данных?
14. Для чего предназначен лингвистический процессор?

Практическая работа № 1

1. Прочитайте статью Бенжамина Фистера «Информационные технологии в помощь переводчику» [19].

Auteur (об авторе):

Après des études d'histoire aux États Unis, Benjamin Phister s'est converti dans l'informatique et a passé 16 ans dans cette industrie en France et aux USA. Il travaille depuis 1995 comme traducteur et rédacteur technique spécialisé dans l'informatique et les finances. Ancien membre du Comité directeur de la SFT, il est actuellement Président de SFT Services, la filiale focalisée sur les formations et les autres activités lucratives du syndicat.

Дайте эквиваленты на русском языке следующим терминам:

Traduction Assistée par Ordinateur (TAO) –

Mémoire de traduction –

Alignement –

Gestion de la terminologie –

Corpora –

Traduction automatique –

2. Каковы, по мнению автора, новые тенденции в использовании информационных технологий в деятельности переводчика?

L'informatique, une aide pour le traducteur

Le traducteur est un être curieux. Il passe des heures interminables devant un écran d'ordinateur, et a tendance à s'en plaindre. Mais bien souvent il connaît mal les possibilités de celui-ci, et passe à côté d'outils et de techniques qui lui permettraient... de passer moins d'heures devant l'écran.

Le travail quotidien du traducteur a été révolutionné par l'arrivée de l'informatique. Il n'y a pas si longtemps, le traducteur tapait son texte à la machine à écrire (s'il n'écrivait pas directement à la main, ou dictait). Après l'avoir révisé, il fallait le saisir à nouveau. Changer deux mots dans une phrase pouvait nécessiter de retaper la page entière... Aujourd'hui il est difficile d'imaginer travailler dans ces conditions, sans accès à un outil de traitement de texte.

Depuis l'arrivée des ordinateurs individuels dans les années 1980, le traducteur dispose de sa propre machine pour saisir sa traduction. Il peut chercher les termes dans les dictionnaires électroniques, ou dans les traductions précédentes stockées sur son disque dur. Il peut utiliser des outils développés spécifiquement pour lui faciliter la tâche.

Une autre technologie a bouleversé encore d'avantage nos méthodes de travail : l'Internet.

Aujourd'hui nous recevons et envoyons nos textes via courriel. Nous consultons des ressources documentaires en ligne, dans le monde entier, comprenant des dizaines voire des centaines d'articles sur des domaines

spécialisés. Nous pouvons faire de la recherche terminologique sur des millions de documents, et tester la validité de nos choix en cherchant le nombre d'occurrences du terme choisi, et en vérifiant les sources. Nous pouvons collaborer sur des projets avec des collègues à l'autre extrémité du pays, ou sur un autre continent – partageant textes, glossaires, et mémoires. Il serait bien difficile de travailler efficacement sans l'Internet.

L'éventail d'outils spécifiques pour traducteurs est bien trop important pour tout passer en revue dans cet article. Voici une sélection des outils les plus importants:

– Traduction Assistée par Ordinateur (TAO).

La Traduction Assistée par Ordinateur a beaucoup évolué depuis ses origines, et aujourd'hui représente un environnement offrant un ensemble d'outils conçus pour optimiser le travail du traducteur. Les composants principaux de cet environnement sont présentés ci-dessous. La plupart sont disponibles dans des suites de programmes tels que SDL Trados Studio 2009, Wordfast, ou Déjà Vu, pour n'en citer que trois (il y en a bien d'autres).

– Mémoire de traduction.

On peut considérer une mémoire de traduction comme une base de données avec un ensemble de phrases traduites. Quand il faut traduire une nouvelle phrase, l'outil recherche dans la base une phrase similaire ou identique, et affiche la traduction précédente, mettant en évidence le(s) mot(s) qui ont changé et remplaçant les chiffres modifiés.

Pour traduire un document, l'outil divise le texte en « segments ». En général un segment est égal à une phrase mais on peut demander une segmentation en fonction de la marque de fin de paragraphe et, dans ce cas, chaque segment contient toutes les phrases d'un paragraphe.

La fonction de recherche contextuelle est très pratique pour trouver des phrases, termes ou même mots traduits auparavant. Cette recherche est utile même pour le rédactionnel, quand le traducteur ne veut pas exploiter les fuzzymatches

mais veut simplement retrouver les termes et tournures de phrases dans la mémoire.

Les systèmes de TAO permettent de gérer des formats de fichier contraignants. L'outil s'occupe des « balises » indiquant le formatage (gras, italique, souligné, changement de polices ou de taille de caractères...) ; le traducteur doit simplement indiquer à quel endroit dans la phrase il faut ajouter la balise.

– Alignement.

Pour certains projets, le traducteur dispose de textes de référence en langues source et cible (comme par exemple un rapport annuel de l'année précédente, ou un manuel utilisateur à mettre à jour).

S'il veut simplement rechercher des phrases ou des termes dans le texte de référence, le traducteur peut utiliser des outils pour générer un « bitexte » ; il s'agit d'un document avec 2 colonnes, l'une pour le texte source et l'autre pour le texte cible. Chaque phrase du texte source est affichée avec sa traduction dans la colonne en face.

Par contre, si le traducteur veut inclure les textes de référence directement dans sa mémoire de traduction, il doit « aligner » les textes pour indiquer que telle phrase dans le texte source correspond à telle phrase dans le texte cible. La plupart des systèmes de TAO offrent des outils pour faire un alignement manuel. Mais ce travail d'alignement est fastidieux et chronophage. Certains outils ont été développés pour automatiser ce processus, comme AlignFactory Light de Logiterm. Ils proposent des filtres dotés de multiples critères pour contrôler la qualité de l'alignement (p. ex. : rejeter un segment si un des côtés est vide, ou si l'écart de longueur est excessif, etc.).

– Gestion de la terminologie.

Pratiquement tous les outils de TAO (Transit, Trados, Déjà Vu, Wordfast, MemoQ...) offrent des moyens de gestion de la terminologie. On peut aussi utiliser des outils indépendants tels que Lingo de Lexicool, ou Anylexic de AIT. Avec ces outils le traducteur peut importer des glossaires – fournis par le client, développés

par le traducteur, ou trouvés sur le Web – en divers formats (Word, Excel, HTML...). Il peut ajouter de nouveaux termes, et (avec les systèmes plus avancés) inclure diverses informations sur chaque terme : le nom du projet, du client, et du domaine ; le genre ; une définition ; le sigle ou acronyme associé ; une image...

Quand le traducteur commence à traduire un nouveau segment, l'outil cherche les termes à traduire dans la base terminologique ; quand il en trouve, il indique ces termes dans la fenêtre avec le texte source à traduire. Le traducteur peut insérer la traduction du terme par un simple clic ou un raccourci clavier.

On peut créer de multiples bases terminologiques par domaine et/ou par client. Cette technique permet d'appliquer la terminologie propre à chaque client.

– Contrôle qualité.

Beaucoup de systèmes de TAO (comme Trados Studio 2009 et Transit) ainsi que des outils indépendants (comme QA Distiller et ErrorSpy) offrent aussi des fonctions de contrôle qualité.

En plus de la vérification orthographique, ces programmes peuvent vérifier que les chiffres sont identiques entre le texte source et cible et qu'ils sont dans le bon format. Ils peuvent contrôler et corriger automatiquement la ponctuation en fin de phrase, les espaces en trop, la casse, les balises, et la terminologie utilisée. Tous ces contrôles automatiques permettent d'améliorer la qualité de la traduction tout en faisant gagner du temps au stade de la relecture.

– Recherche dans des corpora.

Depuis quelques années, le traducteur peut utiliser l'Internet pour accéder à des corpora. Il s'agit de grands volumes de textes déjà traduits et alignés. Le traducteur définit la langue source et la langue cible, puis saisit un terme ou quelques mots à traduire. Le système (tel que Linguee ou l'outil du Taus Data Association) affiche une table avec les phrases source et cible, en surlignant le terme recherché. Les nouvelles versions des outils de TAO (notamment MemoQ) permettent de lancer des recherches dans les corpora ou les fichiers de référence, directement depuis l'éditeur.

– Traduction automatique.

Ces systèmes font traduire le texte entièrement par la machine, sans intervention humaine. Certaines solutions (comme Systran) sont sur le marché depuis des décennies ; d'autres (comme Google Translate) sont plus récentes. La traduction machine est souvent méprisée par le traducteur – nous avons tous vu des exemples peu flatteurs des traductions automatiques. Elle n'est pas appropriée pour tout type de texte, notamment pour le rédactionnel mais ces systèmes s'améliorent petit à petit. Pour certains types de textes pas trop compliqués, ils peuvent rivaliser avec le traducteur humain.

Enfin, on constate l'émergence d'une autre évolution à fort potentiel, rendue possible par la combinaison de différentes technologies. Certaines organisations comme Google ou l'association TAUS ont compilé d'énormes mémoires de traduction de bonne qualité, basées sur des traductions humaines. Ces organisations peuvent aussi trouver d'excellentes bases de données terminologiques. Si elles intègrent une grande mémoire, une base terminologique, et une traduction machine, ces organisations peuvent obtenir – pour certains types de texte – des traductions de qualité entièrement automatisées, pour un prix imbattable (voire gratuitement). En effet, Google génère des bénéfices via la publicité sur les sites Internet. Il a donc tout intérêt à augmenter le nombre de sites localisés en les traduisant automatiquement.

Tendances nouvelles. TAO : environnement de travail intégré

Les premiers outils de TAO étaient étroitement liés à MS Word (en fait, ils étaient souvent écrits comme une série de macro-commandes sous Word). Cette conception a permis au traducteur de travailler dans l'environnement Word qu'il connaît bien. Mais il y a un inconvénient important : chaque fois que Microsoft offre une nouvelle version de Word, l'outil TAO n'est plus compatible. Il fallait laisser au développeur le temps d'adapter toutes les macros à la nouvelle version. Autre désavantage : l'éditeur était obligé de développer des outils nouveaux pour traduire des fichiers dans d'autres formats tels que PowerPoint, Excel, ou HTML.

Depuis quelques années, quasiment tous les éditeurs ont changé de modèle. Maintenant leurs outils utilisent un seul environnement de travail, utilisable pour

tout type de fichier (Word, PPT, HTML...) et intégrant des dictionnaires et autres outils pour gérer la terminologie. Le traducteur « importe » le(s) fichier(s) à traduire dans cet éditeur, traduit le texte, et « exporte » le texte vers le format cible voulu par le client. Ceci permet d'oublier les questions de format et de mise en page pendant la phase de traduction : le traducteur peut se concentrer sur le texte à traduire. Ce modèle ne plaît pas à tous les traducteurs mais il va falloir s'y habituer...

Concordance au niveau du sous-segment.

Lorsque l'on a un segment à traduire, les outils de TAO peuvent rechercher dans la mémoire des phrases similaires. La plupart des outils font cette recherche au niveau du segment entier, utilisant des formules mathématiques pour comparer les mots entre le segment source et le segment cible. Maintenant les outils comme Similis, Déjà Vu, et Transit NXT permettent aussi de trouver juste quelques mots dans un segment déjà traduit. Cette technique permet de beaucoup mieux exploiter le contenu de la mémoire.

Traduction automatique.

Les récentes versions des outils de TAO (Trados Studio 2009 et Transit NXT, par exemple) permettent au traducteur de demander la traduction automatique de la phrase courante, voire du document entier. Cette intégration de la traduction machine offre un outil de plus à la disposition du traducteur ; à lui de décider s'il veut s'en servir. Il est à parier que ces fonctions seront utilisées de plus en plus, avec les avancées de la traduction machine.

Ces avancées ont donné naissance à un nouveau métier, la post-édition¹, où le traducteur révise une traduction faite entièrement par la machine. Cette approche est utilisée depuis longtemps dans certains secteurs et elle va certainement se développer avec le temps.

Cloud computing.

«L'informatique dématérialisée » est un terme couvrant un ensemble de technologies permettant d'utiliser des ressources matérielles et logicielles situées à distance. On externalise la gestion et l'administration des programmes et des

ordinateurs : l'utilisateur se sert essentiellement de son navigateur Internet pour accéder à ces ressources lointaines. Ceci peut réduire les coûts et le temps nécessaires pour installer, mettre à jour, et maintenir son système informatique.

La boucle est bouclée.

Avec l'arrivée du cloud computing, l'informatique referme une grande boucle. Dans les années 1950 et 1960, les ordinateurs centraux (mainframes) étaient des monstres occupant des pièces entières et nécessitant une armée de techniciens pour les alimenter, les dresser, les nourrir, et les manipuler. Très chers à fabriquer et à entretenir, ces mastodontes étaient réservés aux grandes entreprises et instituts de recherche. Les ordinateurs centraux puissants exécutaient les programmes et les utilisateurs tapaient leurs commandes à distance via des terminaux simples, sans intelligence.

Dans les années 1970, l'arrivée des micro-processeurs – des calculateurs entiers sur un seul circuit – a permis la création des premiers micro-ordinateurs.

Le concept de « l'ordinateur individuel » était tout nouveau : quelques individus en effet ont rêvé qu'un jour chacun pourrait avoir un ordinateur sur son propre bureau. Je ne sais pas s'ils avaient imaginé que quelques décennies plus tard on en aurait aussi dans nos voitures, nos machines à laver, et jusque dans notre poche.

La percée des micro-ordinateurs dans les années 1980 a donc permis à chacun d'avoir un ordinateur sur son bureau et il est rapidement devenu évident qu'il fallait partager des données entre utilisateurs. On a alors mis en place les réseaux locaux, en tirant un câble entre les machines pour qu'elles puissent communiquer entre elles. Puis on a vu qu'il fallait des machines puissantes dédiées au stockage de données et à l'exécution de certains programmes tels que les bases de données : l'architecture « client-serveur » est née.

Malheureusement le concept même de l'ordinateur individuel pose des difficultés aux utilisateurs. Avec la disparition de l'armée des techniciens qui s'occupait des calculateurs centraux, chaque utilisateur est devenu responsable de l'installation de tous les logiciels sur son propre ordinateur ; il doit les configurer,

les faire cohabiter et évoluer, et résoudre les problèmes techniques inévitables – d'où une perte de temps considérable.

L'Internet offre la possibilité de se connecter sur des serveurs distants et d'y stocker ses données. Avec l'informatique dématérialisée, les applications tournent aussi sur ces serveurs à distance : nous y accédons via un simple navigateur qui affiche des informations calculées et stockées sur les serveurs. Donc plus de maintenance en local, de reconfiguration, etc. Il suffit que l'éditeur du logiciel mette à jour le logiciel sur le serveur central et tous les utilisateurs en bénéficient immédiatement.

Ce concept de terminaux non intelligents reliés à un serveur central puissant ressemble étrangement à ce que l'on a connu il y a 50 ans.

2 Онлайн-словари и переводчики

Онлайн-словарями называют программы, чьей основной функцией является поиск значения или перевода отдельных слов. Большинство современных онлайн-словарей выдает несколько значений слов или вариантов перевода, с так называемыми словарными пометами. Онлайн-переводчики тоже могут переводить отдельные слова, но их основная задача – перевод текстов.

Онлайн-словари бывают разных видов:

– Двухязычные словари (билингвальные) – это словари с иностранным словом слева и переводом на русский справа или наоборот. Иностранные слова часто снабжены озвучкой. Переводов обычно несколько, они снабжены комментариями, примерами. Все онлайн-словари позволяют свободно переключаться между направлениями. Многие онлайн-словари имеют большую базу из нескольких языков, поэтому правильнее их было бы назвать многоязычными.

– Толковые словари – это словари, в которых иностранному слову дается объяснение на иностранном же языке. Помимо значений слова, обычно приводятся примеры из речи, озвучка.

Приведем краткий обзор онлайн-словарей.

«ABBYY LINGVO» (www.lingvo.ru)

Системные требования ABBYY Lingvo (многоязычная версия):

Операционная система: Microsoft Windows Vista/Microsoft Windows Server 2003/Microsoft Windows XP, 1 ГГц процессор, 512 МБ ОЗУ, от 350 до 1200 МБ на жестком диске (в зависимости от количества установленных словарей), звуковая плата, наушники или колонки.

Lingvo — универсальный словарь. Он подходит для бизнеса, работы и учебы, изучения языка, для путешествий, профессионального перевода. Для такого полнофункционального пакета системные требования невелики.

Разработка Lingvo началась в 1989 году по инициативе Давида Яна и Александра Москалёва. Сейчас это популярнейший электронный словарь. Переводчик Lingvo распространяется в нескольких изданиях, в том числе и в jewel-версии, с существенными ограничениями. Для перевода можно выбрать один из 11 языков: английский, немецкий, французский, испанский, итальянский, португальский, китайский, турецкий, украинский, латинский или русский.

Пакет включает в себя более 150 словарных баз и 8,7 млн. статей. Эти словари разрабатываются на профессиональном ресурсе LingvoDA.ru независимыми лексикографами. Поэтому разница между Lingvo и обычными программами-словарями — огромна. Каждое слово Lingvo переводит предельно подробно, перевод проиллюстрирован примерами из книг, газет и др. От версии к версии в программе появляются новые словари (в том числе и орфографические, толковые), базы озвученных слов.

Лингво – «говорящий» словарь, т.е. многие слова можно прослушать. Для каждого слова в Лингво есть не только перевод с полноценной словарной статьей, но и примеры из литературы, примеры и перевод словосочетаний. Это очень удобно, ведь по-настоящему смысловое богатство слова раскрывается только в контексте.

С помощью данного словаря можно самостоятельно изучать языки, расширить словарный запас с помощью дополнения к словарю «Lingvo Tutor». Tutor представляет собой разговорник на 6 языках. Наиболее употребляемые разговорные фразы озвучены профессиональными переводчиками, что немаловажно для понимания и правильного произношения заученных слов.

Мультитран (<https://www.multitran.ru>).

Разработчик: Андрей Поминов

Для загрузки доступны несколько дистрибутивов «Мультитрана», в зависимости от количества терминов. Есть версия на 2000 терминов (размером в 2 Мб), а есть и на 400000 (31 Мбайт), и даже на миллион слов (65

Мбайт). Прежде всего, программа предназначена для профессиональных переводчиков. Она с легкостью переводит профессиональные термины и жаргонизмы. Словарные базы постоянно расширяются — практически ежедневно. Совершенно свободно вы можете обновлять словари программы посредством онлайн-обновления (в ручном или автоматическом режимах).

Мультитран не содержит озвучки, но он популярен среди переводчиков, так как у него более простой интерфейс, в окне показаны варианты перевода из всех словарей сразу. Зачастую слова, особенно технические термины, имеют очень различные значения в разных сферах деятельности, поэтому полезно иметь перед глазами сразу много толкований: строительное, экономическое, космическое, морское и другие.

Le petit Robert / Малый Робер (<https://www.lerobert.com>).

Одноязычный словарь для иностранцев, изучающих французский язык. К каждому слову приводятся примеры употребления, грамматические пометы, произношение. Словарь содержит около 60 000 слов.

Larousse (<https://www.larousse.fr>).

Одноязычный словарь для иностранцев, изучающих французский язык. Содержит 135 000 определений и 6 000 статей, позволяющих избежать трудности в словоупотреблении, 92000 синонимов и 29000антонимов, 34000 выражений, 15000 омонимов. Помимо этого на сайте представлены ссылки на энциклопедические статьи по различной тематике и форумы, где можно обсудить интересующие вопросы на французском языке. Словарь также включает в себя полные таблицы спряжения глаголов.

Le dictionnaire.com (<http://www.le-dictionnaire.com>).

Одноязычный словарь, включающий 95000 французских слов с определениями, а также их синонимы, спряжение.

Dictionnaire.net (<http://www.dictionnaire.net>).

Одноязычный словарь с возможностью перехода на ссылки словарей различной тематике.

Онлайн-переводчики

Программа: «Google Translate»

Работа системы машинного перевода «Google Translate» осуществляется по технологии, основанной на статистическом вычислении вероятности совпадений. Этой системой используется множество баз параллельных текстов, в которых попарно хранятся словосочетания и их переводы. В процессе перевода осуществляется статистический анализ: система подбирает эквивалент для перевода, основываясь на частоте употреблений, и в итоге подставляет вариант, имеющий наиболее высокий процент совпадений. Следует отметить, что «Google Translate» изначально разрабатывался для перевода с английского языка и на него, и до сих пор английский является языком-посредником при работе с другими парами языков. Иначе говоря, перевод осуществляется не напрямую: сначала происходит трансфер текста с языка-оригинала на английский язык, а уже потом – на необходимый язык перевода, что во многом влияет на качество перевода.

Любопытная функция «Google Translate» – озвучка текста. Если онлайн-словари позволяют прослушать звучание одного слова, то здесь можно озвучить целый текст. Другая функция – перевод с помощью фото. Вы наводите камеру на иностранный текст, предварительно выбрав язык, и прямо на экране он тут же превращается в русский.

Компания Google не перестает создавать интересные проекты. Из недавно появившихся стоит отметить Google Translation Center (www.google.com/accounts/ServiceLogin?service=gtrans). Сервис предлагает следующий способ перевода текстов. По специальной форме вы заполняете заявку, загружаете документ для перевода и запрашиваете перевод. Ждете, пока не найдется доброволец, который сможет перевести ваш текст. Сервис предоставляет средства для общения между заказчиком и переводчиком, а также предоставляет средства для редактирования документа. На сервисе действует Google Translation Center рейтинговая система, которая

контролирует качество перевода. Сервис бесплатен в использовании, однако он закрыт для свободного доступа, поскольку проходит тестирование.

Программа: «Яндекс. Переводчик»

«Яндекс. Переводчик» ничем особенным от «Google Translate» не отличается. В нем тоже доступен перевод с\на десятки языков, для некоторых языков есть озвучка. Главное отличие в том, что у «Яндекс. Переводчика» два режима: перевод текстов и перевод веб-страниц. Перевод текстов работает, как в «Google Translate» с той лишь разницей, что если переводить одно слово, то появится не перевод, а короткая словарная статья, как в Lingvo, но с минимумом информации. А вот в режиме перевода веб-страниц нужно вводить не текст, а ссылку на страницу – в окне появляется ее переведенная версия. Можно настроить просмотр в двух окнах, получаются своего рода параллельные тексты.

Программа «PROMT» (www.promt.ru)

Системные требования PROMT Professional 10.0:

Операционная система Windows 2000 SP 3 (или выше), Windows XP Service Pack 2 (или выше); Windows Vista; компонент .NET Framework 3.0 (устанавливается вместе с продуктом), процессор Pentium 500 МГц или выше; 64 Мбайт оперативной памяти.

«PROMT» работает по принципу «перевода по правилам». Технология этого перевода состоит в применении алгоритмов, в соответствии с которыми программа анализирует текст и на основе проведенного анализа синтезирует вариант перевода.

Компания PROMT предлагает на выбор четыре редакции словаря:

- Expert для профессионалов в области перевода;
- Professional для корпоративных заказчиков;
- Standard для малого офиса;
- PROMT 4U для домашнего пользователя.

Полная версия переводчика требует достаточной ресурсоемкости, хотя официальные требования весьма занижены.

Программа интегрируется с офисными приложениями (Microsoft Office Word, PowerPoint 2000–2007), Microsoft Outlook 2000–2007, ICQ, Adobe Acrobat, OpenOffice Writer и Mozilla Firefox. В контекстном и главном меню приложения появляется меню с командами для быстрого перевода текста, документа. В других случаях, для перевода нужно обращаться к окну «Лингвистический редактор».

Качество перевода зависит от выбранного тематического словаря. На сайте www.promt.ru за отдельную плату можно приобрести дополнительные словари и коллекции словарей самых различных направлений (Авиационный, Автомобильный, Банк, Биология и т. д.). Кроме того, в версии Professional возможно подключение онлайн-словарей из базы Multitran.

Практическая работа № 1

1. Зайдите на сайт www.rvb.ru/soft/catalogue/index.html. В разделе 7 — Словари и тезаурусы — выберите «Словарь сокращений русского языка». Протестируйте предлагаемый онлайн-словарь, введя любое сокращение русского языка. Представьте результаты работы в таблице.

Введенное сокращение	Расшифровка сокращения

2. Протестируйте работу разных систем машинного перевода, размещенных в интернете, от компании Promt (<https://translate.ru>) и от компании Google (<https://translate.google.ru>).

Выполните автоматический перевод следующего теста.

Текст.

Cette machine à espresso automatique au design sobre et épuré est doté de caractéristiques exclusives pour des moments placés sous le signe de la sérénité et

de la performance. Grâce à son broyeur à grains vous apprécierez toute la saveur du café fraîchement moulu.

La durée de la garantie dépend de la législation de votre pays avec un maximum de 3000 cycles par an. Cet appareil est destiné à une utilisation domestique uniquement et ne convient pas à une utilisation commerciale ou professionnelle. Toute utilisation de cette machine dans un contexte autre que domestique n'est pas couvert par la garantie du fabricant.

La garantie ne couvre pas les dommages et détériorations provenant d'une mauvaise utilisation, de réparations effectuées par des personnes non habilitées ou encore au non respect du mode d'emploi. La garantie ne sera pas assurée si toutes opérations ou instructions d'entretien ne sont pas respectées ; si des produits de nettoyage ou de détartrage non conformes aux spécifications mentionnées dans cette notice sont utilisés. La garantie ne s'applique pas dans le cas où la cartouche filtrante Claris Aqua Filter System n'est pas utilisée selon les instructions ROWENTA. L'usure anormale des pièces (meules de broyage, valves, joints) est exempte de la garantie ainsi que les dommages causés par des objets extérieurs dans le broyeur à café (par exemple : bois, cailloux, plastiques, pièces...).

3. Охарактеризуйте протестированные онлайн-переводчики по следующим критериям.

Критерий	Перевод 1	Перевод 2
Затраты времени на выполнение перевода		
Необходимость специальной подготовки пользователя(компьютерные, языковые и т.п.)		
Качество перевода (целостность текста, наличие ошибок, стилистическая однородность)		
Необходимость постредактирования		

3. Отредактируйте варианты перевода. Проанализируйте объем своей работы и заполните таблицу.

Тип редактирования	Частота	
	Перевод 1	Перевод 2
Лексические замены		
Удаление вариантов перевода		
Лексические замены переводов словосочетаний		
Исправление неверного согласования		
Исправление неверного управления		
Вставка дополнительных слов		
Вставка дополнительных словосочетаний		
Удаление лишних слов		
Изменение структуры предложения		

4. Прокомментируйте результаты. Какой вид редактирования самый сложный и почему?

5. Сравните результаты переводов различной тематики. Прокомментируйте, какие недостатки содержат результаты перевода на основании ваших комментариев, занесенных в таблицу.

Тема	Комментарии
Экономика	
Культура	
Здоровье	
Компьютерные технологии	

– Экономика

A l'origine, la convertibilité de la monnaie signifie, la conversion de la monnaie de papier en or. A sa création, la banque d'Amsterdam inventa le procédé suivant ; elle délivrait des billets dont elle garantissait le remboursement en pièces métalliques. Les économistes distinguent généralement la convertibilité interne de la convertibilité externe.

La convertibilité interne. Elle signifie que les agents économiques résidents peuvent échanger les billets en leur possession en quantité équivalente or (la monnaie est totalement gagée sur l'or).

La convertibilité externe. C'est-à-dire que seuls les non résidents sont autorisés à demander à la banque centrale la conversion de la monnaie papier en or.

[\(\[http://pf-mh.uvt.rnu.tn/312/1/L%E2%80%99ECONOMIE_MONETAIRE_PAS_A_PAS.pdf\]\(http://pf-mh.uvt.rnu.tn/312/1/L%E2%80%99ECONOMIE_MONETAIRE_PAS_A_PAS.pdf\)\)](http://pf-mh.uvt.rnu.tn/312/1/L%E2%80%99ECONOMIE_MONETAIRE_PAS_A_PAS.pdf)

– Культура

Depuis une dizaine d'années, une attention nouvelle est prêtée, dans les pays industrialisés comme dans les pays en développement, aux problèmes de la préservation et du développement des cultures et des valeurs qu'elles expriment.

La question des «valeurs» et des «valeurs culturelles» est à la fois très vaste et très complexe. On désigne par valeurs culturelles «les relations symboliques qui assurent la cohésion d'une société donnée ou d'un groupe, maintiennent et renforcent le sentiment d'appartenance de ses membres, perpétuent la richesse de son patrimoine social-spirituel, assurant à sa vie la plénitude et donnent sens aux existences individuelles». Selon M. Makagiansar, sous-directeur general de l'Unesco pour la culture, les valeurs, et par extension les valeurs cultureiles, sont «l'ensemble des signes et des symboles par lesquels s'exprime un système commun d'orientations et de comportements».

[\(<http://unesdoc.unesco.org/images/0005/000546/054681fo.pdf>\)](http://unesdoc.unesco.org/images/0005/000546/054681fo.pdf)

– Здоровье

La maladie d'Alzheimer est une maladie neurodégénérative qui provoque une destruction des cellules nerveuses et donc une diminution continue des facultés

cognitives et de la mémoire. Les symptômes apparaissent en règle générale après 65 ans. Avec 65 % des cas de démence, elle représente la forme de démence la plus fréquente chez les personnes âgées. D'après des études, 1 homme sur 8 et 1 femme sur 4 sont touchés par cette maladie.

(<https://comparateur-assurance-dependance.fr/la-sante-des-seniors>)

– Компьютерные технологии

La technologie a toujours été un moteur extraordinaire dans l'informatique. Au fil des années, les machines deviennent plus petites et moins chères, plus puissantes et moins gourmandes en ressources. Les logiciels progressent en même temps, profitant des nouvelles possibilités et d'autres performances accrues du matériel. Les éditeurs inventent de nouveaux outils, puis les améliorent sans cesse en ajoutant de nouvelles fonctions et possibilités. La concurrence entre fabricants d'une part, et entre éditeurs d'autre part, fait que l'informatique évolue sans cesse, à une cadence infernale.

Le travail quotidien du traducteur a été révolutionné par l'arrivée de l'informatique. Il n'y a pas si longtemps, le traducteur tapait son texte à la machine à écrire (s'il n'écrivait pas directement à la main, ou dictait). Après l'avoir révisé, il fallait le saisir à nouveau. Changer deux mots dans une phrase pouvait nécessiter de retaper la page entière... Aujourd'hui il est difficile d'imaginer travailler dans ces conditions, sans accès à un outil de traitement de texte.

(<https://journals.openedition.org/traduire/251>)

3 CAT-программы

CAT-программы – это программное обеспечение, используемое человеком-переводчиком в процессе перевода для повышения производительности труда. Данное программное обеспечение использует программы Translation memory (TM). Translation memory (переводческая память, накопители переводов) – программы, позволяющие «не переводить одно и то же два раза». Это базы данных, которые содержат ранее переведенные единицы текста. Если в новом тексте обнаруживается единица, которая уже есть в базе, система автоматически добавляет ее в перевод. Такие программы значительно экономят время переводчика, особенно если он работает с однотипными текстами.

TM-программы можно различать по нескольким критериям:

- По технической реализации можно выделить локальные TM-программы и доступные в режиме онлайн.

При использовании онлайн-версий работа происходит в окне интернет-браузера. Локальные программы могут представлять собой отдельные приложения или встраиваемые приложения.

Встраиваемые приложения устанавливаются в текстовый редактор (например, Microsoft Word). Перевод происходит непосредственно в текстовом редакторе. Преимущество такого подхода в том, что пользователь работает с привычным интерфейсом редактора, ему доступны расширенные возможности редактирования текста. Недостаток заключается в ограниченности поддержки форматов файлов - обычно доступны лишь форматы текстового редактора. При работе с отдельными приложениями весь процесс перевода происходит в среде приложения. При этом могут быть доступны дополнительные форматы. Недостатком может быть необходимость привыкать к интерфейсу программы.

- По дополнительным функциональным возможностям.

Часто в ТМ-программы встраиваются программы по управлению терминологией, что позволяет предоставлять общий доступ к терминологическим базам данных, хранить и обрабатывать терминологию. Это позволяет обеспечить единство терминологии в переводе, обеспечить соблюдение языковой политики в рамках отдельного предприятия или предметной области.

ТМ-программы предоставляют возможность построить конкорданс слова - найти список всех употреблений слова в контексте, одновременно предъявляя способы перевода лексической единицы в каждом случае. Данная функция представляет собой доступ к многоязычному корпусу текстов, и поиск лексических единиц осуществляется через интерфейсы, подобные тем, которые применяются при работе с корпусами текстов.

– По общим критериям: бесплатные и условно-платные программы, поддержка различных операционных систем.

Например, STAR Transit – это условно-бесплатная программа и поддерживает только работу в ОС Windows, OmegaT - бесплатная программа с открытым исходным кодом, поддерживает работу в ОС Windows, Linux, Mac OS).

Использование ТМ-программ позволяет значительно увеличить производительность труда переводчика, упрощается работа по соблюдению единообразия терминологии в тексте. Память переводов эффективна при переводе текстов со значительным количеством повторов.

Процесс перевода с использованием памяти переводов включает несколько этапов:

– **Подготовка.**

На стадии подготовки следует решить вопрос о создании новой памяти переводов или осуществить выбор той памяти переводов, которую необходимо использовать для перевода данного текста, произвести необходимые настройки (например, произвести фокусировку на способах

перевода термина, соответствующих ранее созданным проектам для того же заказчика) и корректировки выбранной памяти переводов.

– Собственно перевод.

Предварительный перевод (pre-translation) представляет собой автоматическую замену найденных в базе данных точных соответствий сегментам исходного текста на их эквиваленты в языке перевода. При этом нечеткие совпадения обычно не включаются, так как затраты на пост-редактирование при таком подходе могут оказаться выше выгоды от использования базы примеров перевода. Сегменты текста, которые не были найдены в базе, переводятся вручную или - при наличии - с помощью системы машинного перевода.

Для перевода в узких предметных областях важным компонентом перевода является использование терминологических баз. Даже если при посегментном сравнении исходного текста и базы данных не найдено никаких соответствий, автоматическая замена терминов обеспечивает соблюдение языковой политики предприятия, единство терминологии и предоставляет релевантные в данной сфере эквиваленты терминов человеку-переводчику.

В связи с этим, ключевым аспектом применения ТМ-программ является создание глоссария, ориентированного на данного заказчика или предметную область, который может содержать в своей основе имеющиеся глоссарии или создаваться вновь. При этом создание базы терминов вручную - весьма трудоемкая задача. Поэтому при создании глоссариев прибегают к инструментарию для автоматического или полуавтоматического извлечения терминов из корпусов текстов, основанному на идентификации лексических единиц с применением лингвистического, статистического или - чаще - гибридного подходов. В коммерческих ТМ-программах, как правило, имеются встроенные системы извлечения терминологии и составления глоссария терминов (SDL MultiTerm Extract, SDL MultiTerm, PROMT TerM, Mono-Cone Pro, Simple Concordance Program и др.).

«Сырой» текст, включающий автоматически замененные термины и сегменты текста на их переводные аналоги, на конечной стадии подвергается постредактированию.

– Проверка качества.

Проверка качества включает формальную проверку выполненного перевода на полноту, грамматическую правильность, корректный перевод релевантной терминологии, которая может осуществляться самим переводчиком, (возможно) заказчиком, а также на промежуточном звене, обеспечивающем связь между переводчиком и заказчиком. При этом большое значение имеет процесс фиксации выполненных правок в памяти переводов, иначе ошибки могут повторяться при последующих переводах.

Таким образом, при работе с ТМ-программой выделяются содержательный (включающий использования ряда функций по автоматизации переводческой деятельности), формальный и экономический аспекты перевода. Обозначенные аспекты являются важной составляющей использования памяти перевода в профессиональной деятельности переводчиков.

Основные виды программ:

Trados (<http://www.translationzone.com/trados.html>) – одна из самых популярных программ Translation memory. Позволяет работать с документами MS Word, презентациями PowerPoint, HTML-документами и файлами других форматов. В Trados есть модуль для ведения глоссариев. Сайт: Déjà Vu. Также один из лидеров по популярности. Позволяет работать с документами практически всех популярных форматов. Есть отдельные версии программы для переводчиков-фрилансеров и для бюро переводов.

OmegaT (<http://www.omegat.org>) поддерживает большое количество популярных форматов, но документы в MS Word, Excel, PowerPoint требуется конвертировать в другие форматы. Программа является бесплатной.

Wordfast (<http://www.wordfast.com/>) – программа ориентирована, прежде всего, на переводчиков-фрилансеров. Предлагается несколько версий программы, в том числе – бесплатная, которая доступна онлайн (Wordfast Anywhere).

MetaTaxis (<http://www.metataxis.com/>) позволяет работать с документами основных популярных форматов. Предлагается два варианта программы – модуль для MS Word и серверная программа.

MemoQ (<http://kilgray.com/>) – программа, функционал которой схож с Tradosом и Déjà Vu, стоимость программы ниже, чем у более популярных систем.

Star Transit (<http://www.star-group.net/DEU/group-transit-nxt/transit.html>) предназначена для перевода и локализации. На данный момент совместима только с ОС Windows.

WordFisher (<http://www.wordfisher.com/>) – бесплатная система Translation Memory, созданная и поддерживаемая профессиональным переводчиком.

Across (<http://www.across.net/us/translation-memory.aspx>.) предлагается четыре различных версии программы, отличающихся по объему функционала.

Практическая работа

1. Зайдите на сайт программы «Trados Studio» <https://www.tradostudio.com/>. Ознакомьтесь с общей информацией и ответьте на следующие вопросы.

1. На чем основаны Решения SDL Trados?
2. Что позволяет выполнить программа «Trados Studio»?
3. Какие услуги предоставляет создание баз в формате Translation Memory?

2. Соотнесите этапы перевода текста в программе «Trados Studio» и действия на необходимые действия на данных этапах.

1. Открытие файла
2. Подключение базы переводов
3. Настройка
4. Проверка совпадений
5. Подстановка распознаваемых элементов
6. Перенос тегов
7. Перевод и подтверждение
8. Предварительный просмотр
9. Проверка перевода
10. Сохранение перевода

- a) перейти в меню «Файл» выбрать «Перевести документ и выбрать файл для перевода»;
- b) в диалоговом окне «База переводов и настройки документа» подключить готовую или создать новую базу переводов (ТМ);
- c) для изменения стандартных настроек во вкладке «Главная» выбрать «Настройки проекта», где можно подключить дополнительные ресурсы – терминологические базы, словари или службы машинного перевода;
- d) программа осуществляет автоматически поиск совпадений в базе переводов и терминологической базе;
- e) распознаваемые элементы – это непереводимые элементы (теги, числа, даты, единицы измерения);
- f) в тегах содержится информация о структуре или форматировании исходного текст; необходимо перенести теги из исходного текста в перевод, чтобы конечный документ выглядел как и оригинал;
- g) для подтверждения перевода необходимо нажать сочетание клавиш Ctrl+Enter;
- h) необходимо нажать кнопку «Предварительный просмотр» для предварительного просмотра переведенного документа в программе;

- i) после завершения перевода необходимо запустить «Проверку»; перейти на вкладку «Редактирование» и выбрать «Проверить орфографию», а затем «Проверить». Обнаруженные ошибки отображаются в окне «Сообщения»;
- j) для сохранения двуязычного файла необходимо щелкнуть «Сохранить»; для получения конечного переведенного текста щелкните «Файл – Сохранить как».

Комментарии:

1. Полезные сочетания клавиш:

Ctrl + , – отображение списка Быстрая подстановка (отображение чисел, аббревиатур, форматирования текста и тегов для выделенного сегмента);

Ctrl + Enter – подтверждение сегмента (подтверждение перевода сегмента и сохранение единицы перевода в памяти);

Ctrl + Shift + F2 – пополнение терминологической базы (для пополнения терминологической базы выделите исходное слово или словосочетание и его перевод, а затем нажмите кнопку «Быстрое добавление термина»);

Ctrl + T – подстановка совпадения (подстановка совпадения из базы переводов. Используйте сочетания клавиш Ctrl + 1 , Ctrl + 2 , и т.д., чтобы подставить нужное совпадение);

F3 – запуск поиска Конкорданс (выделите слово в исходном или в переведенном сегменте, чтобы выполнить по нему поиск Конкорданс, результаты поиска отображаются в окне Конкорданс, искомые термины выделяются желтым цветом);

Ctrl + Ins – копирование исходного текста (копирование исходного сегмента в целевой. Это особенно удобно, когда в исходнике много тегов);

Shift + F3 – смена регистра (смена верхнего и нижнего регистра в выбранных словах);

F7 – проверка орфографии (запуск проверки орфографии. Файл > Параметры > Редактор > Орфография изменение инструмента проверки орфографии по умолчанию);

Shift + F12 – вызов диалогового окна Сохранить как (отображение окна Сохранить перевод как для сохранения перевода в исходном формате);

2. Используйте автоподстановку слов. Автоподстановка позволяет ускорить перевод, предлагая подсказки при наборе текста. Источниками подсказок являются база переводов, терминологические базы, машинный перевод и словари AutoSuggest.

– для создания словаря AtoSuggest перейдите в режим Базы переводов и нажмите кнопку > Создать словарь AutoSuggest;

– чтобы подключить словарь AutoSuggest к проекту, щелкните Настройки проекта > Языковые пары > [выберите нужное направление];

– для настройки автоподстановки перейдите в Файл > Параметры > Автоматические подсказки.

3. Пользуйтесь поиском Конкорданс, чтобы найти в базе переводов определенное слово, словосочетание или фразу. Поиск Конкорданс сопоставляет часть выделенного текста или сегмента даже если порядок слов и контекст в единице перевода отличается. Нажмите клавишу F3, чтобы запустить поиск Конкорданс.

4. Используйте функцию Быстрая подстановка для переноса форматирования из исходного текста в перевод. Выделите текст перевода, затем нажмите сочетание клавиш Ctrl + , , чтобы выбрать и применить необходимое форматирование из списка.

5. Для пополнения терминологической базы выделите исходное слово или словосочетание и его перевод, а затем нажмите кнопку Быстрое добавление термина.

6. Чтобы изменить параметры перевода только для текущего проекта, меняйте настройки в режиме Проекты на вкладке > Главная с помощью кнопки > Настройки проекта. Чтобы изменить стандартные параметры Studio и далее применять их для всех новых проектов, меняйте настройки через Файл > Параметры.

7. Открыть файл для перевода можно перетащив его на панель навигации Редактор или в окно Редактор.

8. Для настройка рабочей среды Studio вынесите на Ленту самые необходимые кнопки и команды, для этого перейдите на вкладку Вид в группу > Пользовательский интерфейс и выберите > Настройка ленты. В режимах Проекты, Файлы, Редактор и Базы переводов можно изменить размер и расположение окон.

4 CAT-программа «Omega T»

CAT-программа «Omega T» (<http://omegat.org/ru>) это свободная система автоматизированного перевода, поддерживающая память переводов, написанная на языке программирования Java. Является бесплатной программой с открытым исходным кодом, которая работает на Windows, Mac и Linux-системах и постоянно совершенствуется сообществом.

Данная программа разделяет исходный текст на сегменты — строки, предложения, параграфы или абзацы. Человек переводит сегмент один за другим, а перевод сохраняется в специальную базу данных — память перевода (Translation memory, TM). Если переводчику встретится похожий сегмент, программа покажет подсказку или возможный перевод. А одинаковые сегменты программа может переводить сама. Когда перевод закончен, программа создаёт документ, идентичный оригиналу — сохраняя структуру и форматирование, но заменяя исходный текст переводом.

Профессиональные функции CAT-программы «Omega T»:

- подбор неточных совпадений;
- размножение совпадений;
- одновременная обработка проектов с большим числом файлов;
- одновременное использование нескольких памятей переводов;
- использование глоссариев с признанием флективных форм;

Эффективные инструменты программы:

- поддержка уникода (UTF-8): используется для алфавитов без латиницы;
- поддержка языков с письмом справа налево;
- встроенное средство проверки орфографии;
- совместимость с другими программами автоматизированного перевода.

Программа «OmegaT» работает не с отдельными файлами, а с «проектами». Проект — набор папок с определённой структурой. Чтобы перевести файл, нужно создать проект, а потом добавить туда файл.

Внутри папки проекта есть несколько под-директорий:

- dictionary — можно добавить словари в формате StarDict;
- glossary — база терминов по проекту;
- omegat — память перевода и резервные копии проекта;
- source — папка с исходными файлами;
- target — папка, в которой будут появляться переводы;
- tm — папка для дополнительных памятей перевода.
- omegat.project с конфигурацией текущего проекта.

Практическая работа

1. Установите на ваш персональный компьютер программу «Omega T», следуя следующему алгоритму:

– Скачайте дистрибутив с сайта <http://omegat.org/ru>. Будем использовать русскоязычную Latest Version - OmegaT 4.1.5 update 1 для Windows. Для запуска требуется Java. Если не уверены, есть ли она у вас, скачайте версию с пометкой JRE.

– Запускаем программу «OmegaT».

– Переходим в параметры > настройки > параметры орфографии. Ставим галочку *«автоматически проверять орфографию текста»*. Нажимаем установить. Выбираем язык (например, ru_RU для русского), нажимаем «Установить», затем «Заккрыть». В списке видим установленные языки. Выходим из настроек.

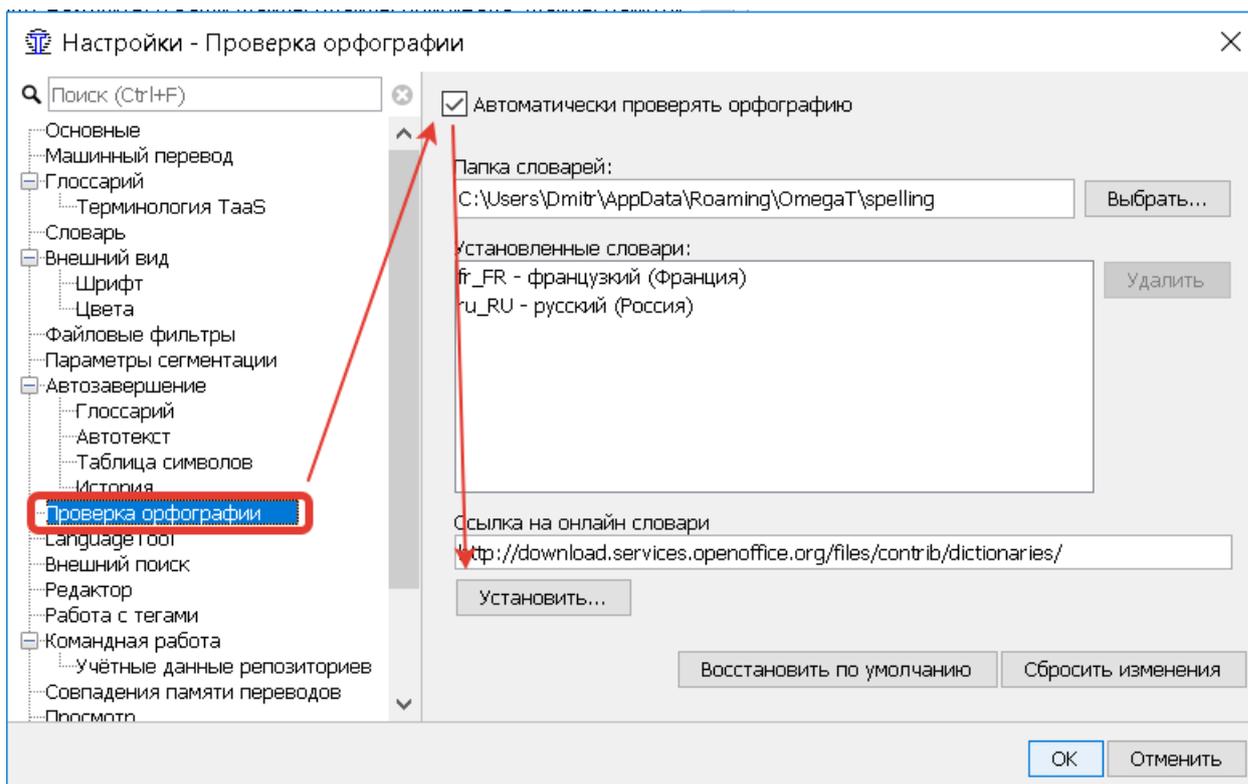


Рисунок 1 – Установка программы

2. Переведите следующий текст с помощью программы «OmegaT».

Текст. Utilisation

N'utilisez pas cette brosse pour nettoyer des personnes ou des animaux.

N'utilisez pas l'appareil à l'extérieur ou sur des surfaces humides. Ne nettoyez que des surfaces sèches.

Utilisez l'appareil conformément aux instructions du présent manuel avec les accessoires recommandés par Miele.

Cet appareil est destiné uniquement à un usage domestique.

Cet aspirateur n'est pas un jouet. Il est nécessaire d'être vigilant lorsqu'il est utilisé à proximité d'enfants.

N'utilisez jamais l'aspirateur si le sac à poussière ou les filtres ne sont pas en place.

Vous pourriez endommager l'appareil.

2.1 Для перевода текста в программе «OmegaT» создайте проект (набор папок с определённой структурой).

Для создания проекта следуйте следующему алгоритму:

– Запускаем программу «OmegaT».

- Проект > Создать. Выбираем место для сохранения и имя проекта.
- В появившемся окне укажите языковую пар. Обратите внимание, что язык файлов оригинала — это язык, с которого вы переводите, язык переведенных файлов — это язык, на который вы переводите. Указывать нужно в двух- или четырех-буквенном коде. Чтобы работала проверка правописания, код должен совпадать с кодом, указанным в настройках орфографии (если в орфографии установлен RU-RU, а в проекте будет RU, то проверка работать не будет).
- Ниже отметьте галочки «включить сегментацию по предложениям (делить сегменты по предложениям, а не по абзацам)» и «автоматическое копирование переводов (подставлять переводы автоматически)». Галочку «убрать теги» лучше снять.
- Нажимаем ОК.

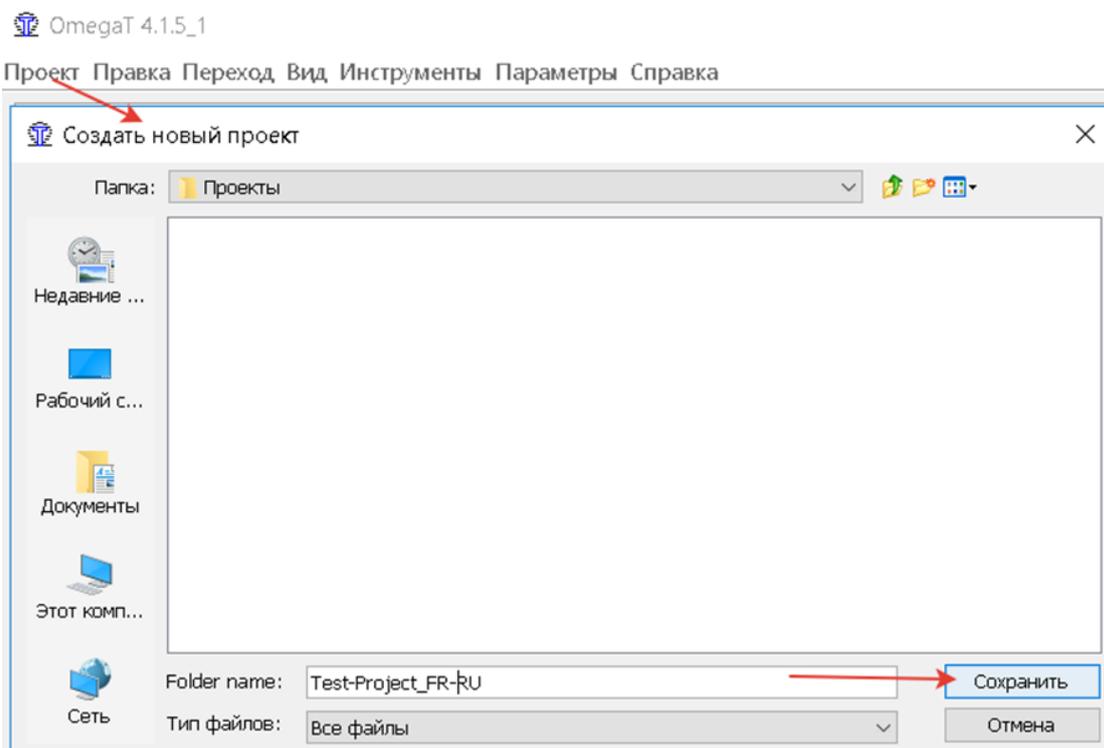


Рисунок 2 – Создание проекта (наименование)

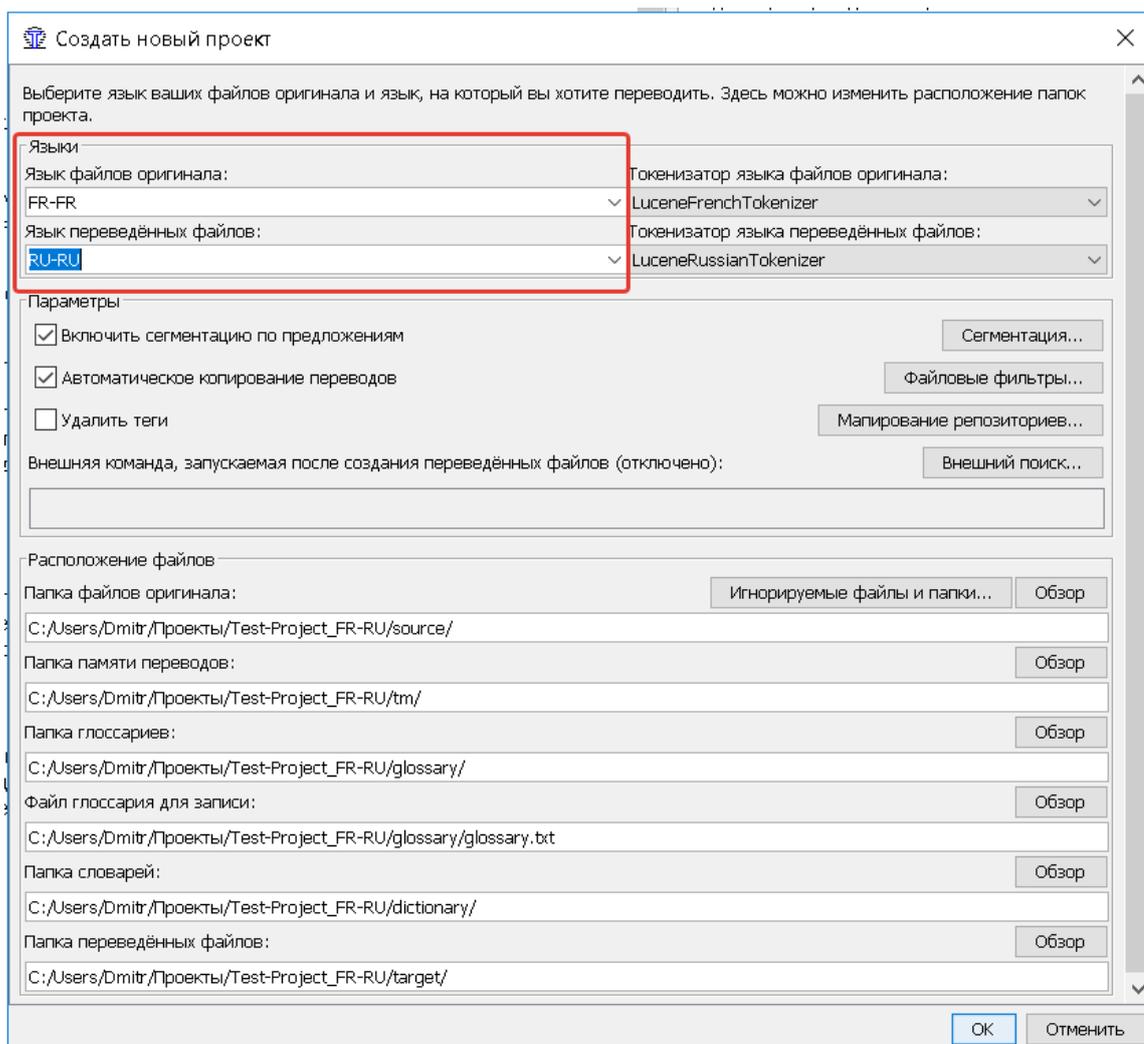


Рисунок 3 – Создание проекта (выбор языков)

– Для перевода текста скопируйте документ. Для этого нажмите «скопировать документы в папку файлов оригинала» и выберите файлы, которые вы хотите перевести. Файлы будут скопированы в папку «*source*», только что созданного проекта. Вы можете добавить туда файлы вручную, скопировав файлы в «*source*» через проводник.

Комментарии:

– При работе в программе «OmegaT» обратите внимание, что весь текст разделён на предложения, форматирования не видно, появились теги серого цвета, а заголовок дублируется.

– Текст разделился на сегменты. Каждое предложение выделилось в отдельный сегмент. Правила сегментации можно настроить самостоятельно при необходимости.

– Форматирование в программе не видно, его заменяют теги. Они представляют собой сокращения тегов из Word, которые иначе могли бы выглядеть как <t>. Чтобы сохранить оригинальное форматирование, нужно оставлять эти теги как есть, вписывая перевод между тегами по той же логике, что и в оригинале.

– Опция «*remove tags*» в настройках проекта убирает теги вместе с форматированием. Не рекомендуется использовать, если важно сохранить оригинальное форматирование.

– Заголовок дублируется. На самом деле, сверху (в зелёном цвете) всегда отображается текст на исходном языке, изменить его нельзя. Под ним находится текстовое поле, куда по умолчанию скопирован тот же самый текст. Его нужно удалить и вписать перевод.

– В правой части программы есть два сектора: «*нечеткие совпадения*» и «*гlossарий (словарь проекта)*».

Нечеткие совпадения — результаты поиска по базе данных проекта. Здесь отображаются подсказки по переводу, основанные на ваших предыдущих переводах.

Гlossарий (словарь проекта) — результат поиска по гlossарию, который вы составляете самостоятельно. В отличие от памяти перевода, это не готовый текст, а лишь подсказки по определённым терминам. Это мощный инструмент, который помогает сохранять единообразие в терминологии.

2.2. Для перевода текста выполните следующие действия:

– Кликните на сегмент для перевода. Под оригинальным текстом появится редактируемая текстовая строка, курсор будет в её начале, а в строке будет продублирован оригинальный текст.

– Впишите свой перевод.

– Нажмите «Enter». При нажатии перевод сохранится, а курсор перейдёт к следующему сегменту.

– Повторяйте, пока не закончите документ. В любой момент можно вернуться к предыдущему сегменту, дважды щёлкнув на него. В правом нижнем углу есть удобный индикатор прогресса. Кликните на него, чтобы переключить режим просмотра.

20% (осталось: 8) / 20% (осталось: 8), 10 71/71

Рисунок 4 – Режим просмотра

В этой строке указано, что в текущем файле переведено 20% уникальных сегментов, осталось перевести ещё 8, а их общее число в проекте — 10. Цифры 71/71 в конце не относятся к счётчику проекта. Это — индикатор длины сегмента, с которым вы работаете. Он говорит, что в оригинале было 71 символов, и в переводе их тоже 71. Эта функция полезна в тех случаях, когда нужно строго соблюдать длину строки, например, при переводе интерфейса программ.

– При переводе текста используйте инструмент «Нечёткие совпадения Fuzzy Matches», самый главный инструмент любого CAT-приложения, ради этого они и существуют.



Рисунок 5 – Инструмент «Нечёткие совпадения Fuzzy Matches»

В документе-образце *второе* предложение очень похоже на *одиннадцатое*. После перевода второго предложения программа сразу же показала нечёткое совпадение с текстом одиннадцатого предложения.

В верхней части отображается текст на исходном языке, который был сохранён в памяти перевода. Синим цветом выделены слова, которые присутствуют в памяти перевода, но отсутствуют в текущем предложении (с которым сравнивается совпадение), зелёным — слова, расположенные рядом с недостающими частями. Ниже будет перевод, сохранённый в памяти.

Если нажать Ctrl+R, то он скопируется в поле для перевода.

2.3. При переводе текста используйте «Глоссарий». Чтобы добавить слово в глоссарий, выделите его, щёлкните правой кнопкой и выберите «Добавить в глоссарий». Теперь, когда слово встретится в тексте, в окошке *Глоссарий* сразу же отобразится подсказка. Таким образом, когда в новом предложении появился термин, вы сразу будете знать, как именно следует его переводить.

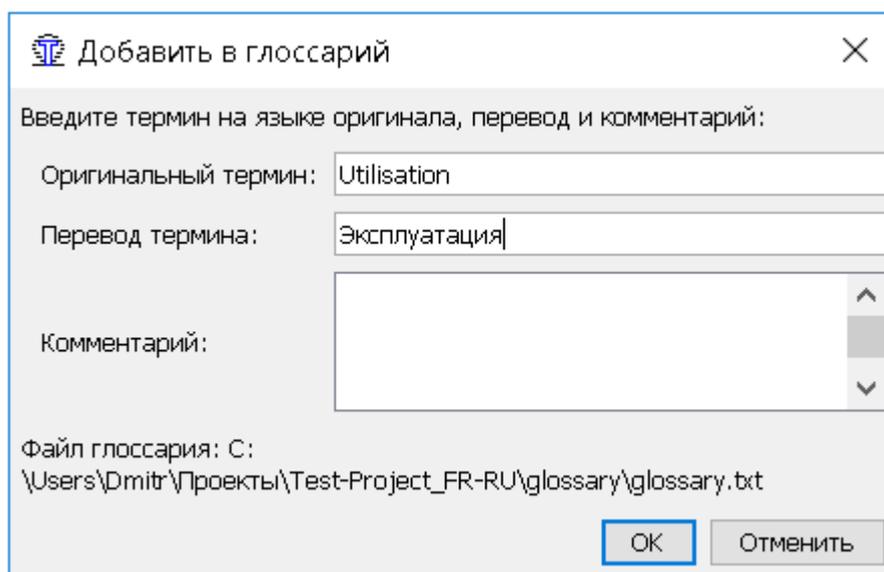


Рисунок 6 – Работа с глоссарием

2.4. По окончании перевода текста сохраните проект. Для этого выполните следующие действия:

– Пункт Проект > Сохранить означает «сохранение проекта», т.е. запись всех переводов в файл базы данных.

– Для того, чтобы получить готовый файл, нужно выбрать Проект > Создать переведенные документы.

По этой команде OmegaT создаст новый файл в папке \target\ с тем же именем, что и оригинал, а весь текст поменяет на перевод. Если какие-то сегменты вы не перевели, то в файле на их месте будет оригинальный текст.

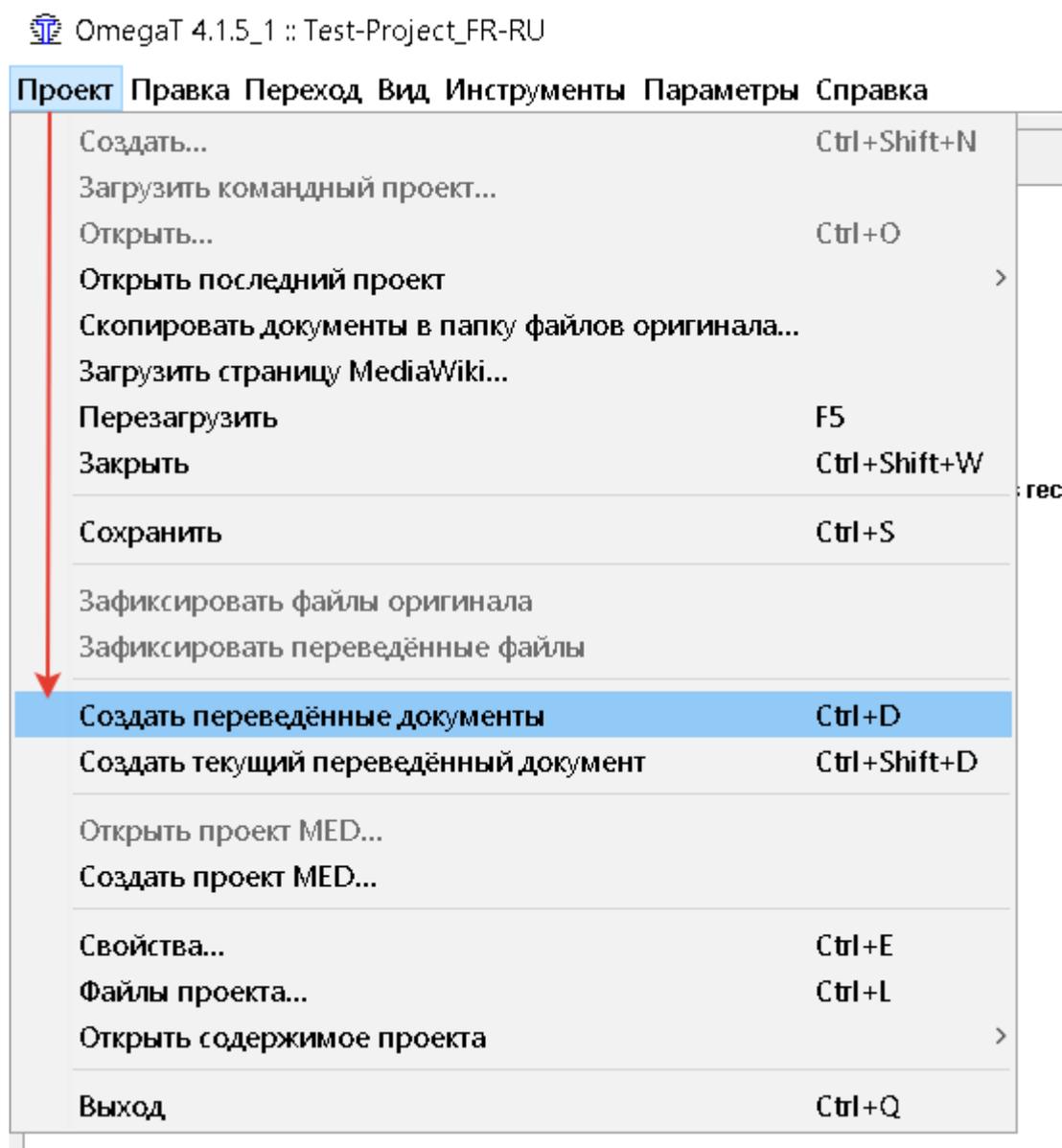


Рисунок 7 – Сохранение проекта

2.5. Изучите дополнительные материалы о работе программы по ссылке <http://omegat.org/ru/documentation#manual>

5 Онлайн-программа «Wordfast Anywhere»

САТ-программа «Wordfast Anywhere» (<https://www.freetm.com/>) - это свободная система автоматизированного перевода, поддерживающая память переводов, написанная на языке программирования Java. Данная программа является бесплатным онлайн-инструментом, который работает на Windows, Mac и Linux-системах.

Данная программа разделяет исходный текст на сегменты — строки, предложения, параграфы или абзацы. Человек переводит сегмент один за другим, а перевод сохраняется в специальную базу данных — память перевода (Translation memory, TM). Если переводчику встретится похожий сегмент, программа покажет подсказку или возможный перевод. А одинаковые сегменты программа может переводить сама. Когда перевод закончен, программа создаёт документ, идентичный оригиналу — сохраняя структуру и форматирование, но заменяя исходный текст переводом.

Профессиональные функции САТ-программы «Wordfast Anywhere»:

- подбор неточных совпадений;
- размножение совпадений;
- одновременная обработка проектов с большим числом файлов;
- одновременное использование нескольких памятей переводов;
- использование глоссариев с признанием флективных форм;
- Эффективные инструменты программы:
 - поддержка уникада (UTF-8): используется для алфавитов без латиницы;
 - поддержка языков с письмом справа налево;
 - встроенное средство проверки орфографии;
 - совместимость с другими программами автоматизированного перевода;
 - интерфейс для Google Translate.

Практическая работа

1. Установите на ваш персональный компьютер программу «Wordfast Anywhere», следуя следующему алгоритму:

– Зайдите на официальный сайт Wordfast Anywhere (<https://www.freetm.com/>) и создайте учетную запись. Сразу после создания учетной записи программа предложит создать проект.

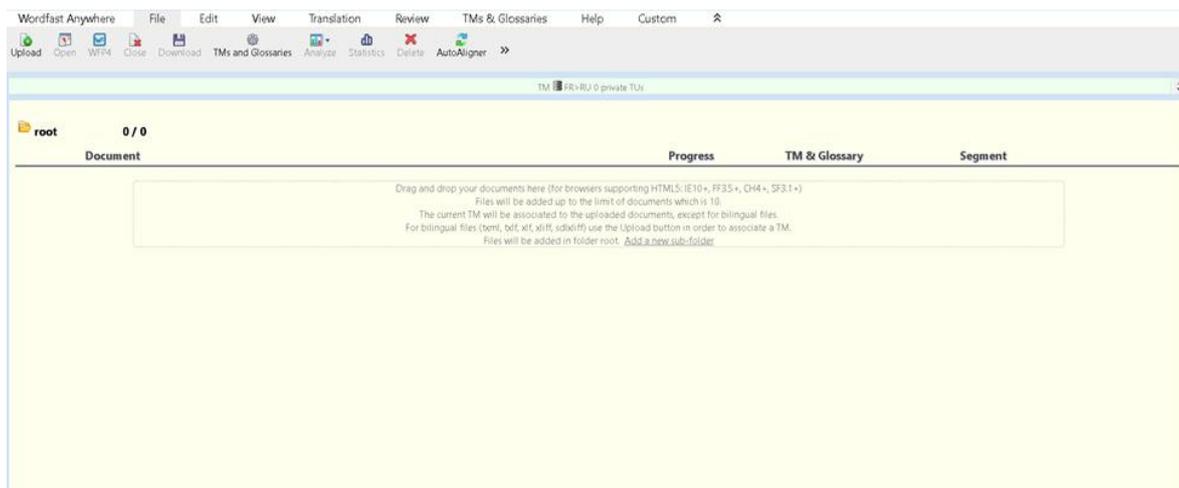


Рисунок 8 – Создание проекта

– Для создания проекта необходимо указать язык оригинала (Source language) и язык перевода (Target language). В появившемся окне укажите языковую пару. Указывать нужно в двух-буквенном коде. Нажимаем ОК.

Например, RU — русский язык.

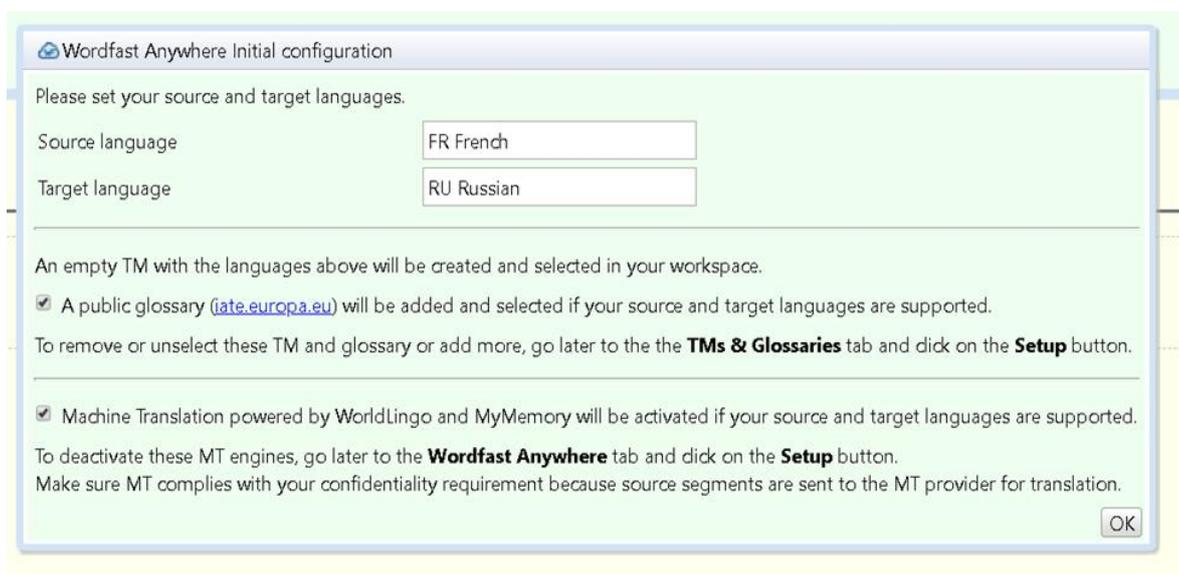


Рисунок 9 – Создание проекта (выбор языков)

2. Переведите следующий текст с помощью программы «Wordfast Anywhere».

Текст. Utilisation

N'utilisez pas cette brosse pour nettoyer des personnes ou des animaux.

N'utilisez pas l'appareil à l'extérieur ou sur des surfaces humides. Ne nettoyez que des surfaces sèches.

Utilisez l'appareil conformément aux instructions du présent manuel avec les accessoires recommandés par Miele.

Cet appareil est destiné uniquement à un usage domestique.

Cet aspirateur n'est pas un jouet. Il est nécessaire d'être vigilant lorsqu'il est utilisé à proximité d'enfants.

N'utilisez jamais l'aspirateur si le sac à poussière ou les filtres ne sont pas en place. Vous pourriez endommager l'appareil.

2.1. Для перевода текста в программе «Wordfast Anywhere» необходимо создать проект, а потом добавить файл. Для этого необходимо нажать File > Upload и выбрать файлы, которые необходимо перевести.

1	Utilisation	0
2	N'utilisez pas cette brosse pour nettoyer des personnes ou des animaux.	0
3	ⓧ N'utilisez pas l'appareil à l'extérieur ou sur des surfaces humides.	0
4	Ne nettoyez que des surfaces ⓧ sèches Ⓣ.	0
5	Utilisez l'appareil conformément aux instructions du présent manuel avec les accessoires recommandés par Miele.	0
6	Cet appareil est destiné uniquement à un usage domestique.	0
7	ⓧ Cet aspirateur n'est pas un jouet.	0
8	Il est nécessaire d'être vigilant lorsqu'il est utilisé ⓧ à proximité d'enfants Ⓣ.	0
9	Encore, n'utilisez pas cette brosse pour nettoyer des personnes ou des animaux.	0
10	ⓧ N'utilisez jamais l'aspirateur si le sac à poussière ou les filtres ne sont pas en place.	0
11	Vous pourriez endommager ⓧ l'appareil Ⓣ.	0

Рисунок 10 – Текст в программе «Wordfast Anywhere»

Комментарии:

– Весь текст разделён на предложения, форматирования не видно, появились теги красного цвета. Таким образом, текст разделился на сегменты, каждое предложение выделилось в отдельный сегмент, форматирование в wordfast не видно, его заменяют теги.

– Чтобы сохранить оригинальное форматирование, нужно оставлять эти теги как есть, вписывая перевод между тегами по той же логике, что и в оригинале.

2.2 Для перевода текста выполните следующие действия:

– Кликните на строчку справа от переводимого сегмента, если при создании проекта вы включили машинный перевод, то вы увидите вариант, предлагаемый машинным переводчиком, его можно удалить.

– Впишите свой перевод.

– Нажмите «Alt+down». При нажатии перевод сохранится, а курсор перейдёт к следующему сегменту.

– Повторяйте, пока не закончите документ. В любой момент можно вернуться к предыдущему сегменту, просто щёлкнув на него. В правом нижнем углу есть удобный индикатор прогресса.

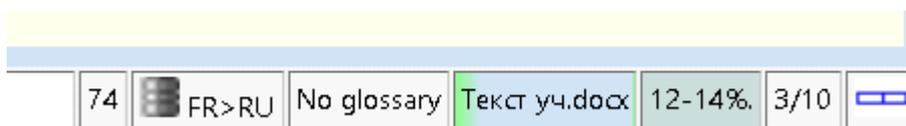


Рисунок – Индикатор прогресса

Здесь указано, что всего переведено 12-14% документа и 3 из 10 сегментов.

2.2. Для перевода текста используйте инструмент «Нечёткие совпадения».

	Использование	
1 Utilisation		100
2 N'utilisez pas cette brosse pour nettoyer des personnes ou des animaux.	Не используйте пылесос для чистки людей или животных.	100
3 N'utilisez pas l'appareil à l'extérieur ou sur des surfaces humides.		100
4 Ne nettoyez que des surfaces sèches.		0
5 Utilisez l'appareil conformément aux instructions du présent manuel avec les accessoires recommandés par Miele.		0
6 Cet appareil est destiné uniquement à un usage domestique.		0
7 Cet aspirateur n'est pas un jouet.		0
8 Il est nécessaire d'être vigilant lorsqu'il est utilisé à proximité d'enfants.		0
9 Encore, n'utilisez pas cette brosse pour nettoyer des personnes ou des animaux.	Не используйте пылесос для чистки людей или животных.	91
10 N'utilisez jamais l'aspirateur si le sac à poussière ou les filtres ne sont pas en place.		0
11 Vous pourriez endommager l'appareil.		0

Рисунок 12 – Использование инструмента «Нечёткие совпадения»

В документе-образце *второе* предложение очень похоже на *девятое*. Программа показывает нечеткое совпадение девятого предложения, если второе уже переведено.

2.3. Для перевода текста используйте «Глоссарий». Данная функция позволяет программе подсказывать как следует переводить те или иные термины. Для этого необходимо в ходе перевода добавлять слова (оригинал и перевод). Выполните следующие действия:

- чтобы создать глоссарий, нажмите TMs & Glossaries > Setup -> Создать;
- чтобы добавить слово в глоссарий, нажмите TMs & Glossaries -> add term, введите оригинал и перевод термина.

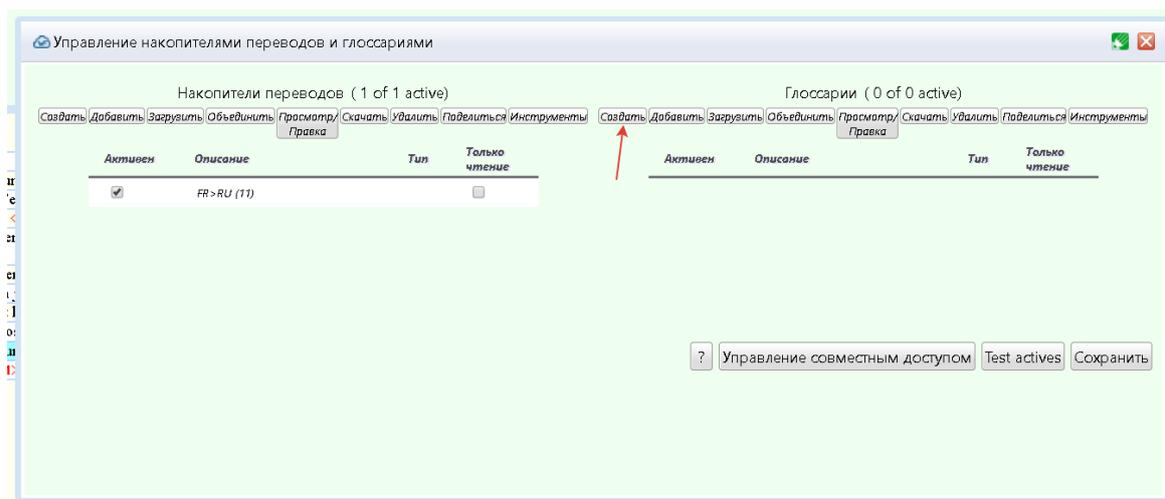


Рисунок 13 – Создание глоссария

2.4. По окончании перевода необходимо сохранить проект. Для этого нажмите File > Download. Проект сохраняется автоматически в виде двуязычного документа, либо документа только с переводом.

Список использованных источников

- 1 Андреев, А. Машинный перевод: правила против статистики / А. Андреев. – Электронные текстовые данные. – Режим доступа: <http://www.computerra.ru/cio/old/offline/2007/63/329838/>.
- 2 Баранов, А. Н. Введение в прикладную лингвистику / А. Н. Баранов // МГУ им. М. В. Ломоносова.- 3-е изд. – М.: ЛКИ, 2007. – 360 с.
- 3 Белоногов, Г.Г. Системы фразеологического машинного перевода политематических текстов [Электронный ресурс]. Режим доступа: <http://www.a-z.ru/person/belonogov/index.htm>
- 4 Зеевальд-Хег, Ю. Локализация программного обеспечения, информационно-технологические требования к переводчику в эпоху глобализации / Ю. Зеевальд-Хег // Вестник Перм. нац. исследоват. политехи, ун-та. Социально-экономические науки. - 2012. -№16.- С. 85-96.
- 5 Зубов, А.В. Основы искусственного интеллекта для лингвистов / А.В. Зубов, И.И. Зубова– М.: Унив. Книга: Логос, 2007. – 320 с.
- 6 Ильнер, А.О. Подготовка переводчиков в России и за рубежом: сравнительный анализ / А. О. Ильнер // Образование и наука. - 2010.- №8. - С. 65-70.
- 7 ИНТУИТ. Национальный открытый университет. Введение в машинное обучение. [Электронный ресурс]. – Режим доступа: <https://www.intuit.ru/studies/courses/10621/1105/lecture/17981>
- 8 Инфопедия: [сайт]: - Режим доступа: <https://infopedia.su/>
- 9 Каде, О Проблемы перевода в свете теории коммуникации / О. Каде // Вопросы теории перевода в зарубежной лингвистике. М., 1978.– С.75
- 10 Краткая история машинного перевода. Журнал «Русский репортер»: [сайт]. – Режим доступа: http://rusrep.ru/2010/24/istoriya_perevoda/
- 11 Кутузов, А.Б. Компьютерные технологии в формировании профессиональной компетенции переводчика / А.Б. Кутузов // Языки

профессиональной коммуникации: сборник статей Третьей международной научной конференции, т.2. – Челябинск, 2007. – 287 с.

12 Молчанов, А., Статистические и гибридные методы перевода в технологиях компании prompt // [Электронный ресурс]. Режим доступа: http://controleng.ru/wpcontent/uploads/ce46_p68_statisticheskie_i_gibridnye_metody_perevoda_v_tekhnologiyakh_kompanii_prompt.pdf

13 Новожилова, А.А. Обучение студентов-переводчиков работе с электронными ресурсами как основа их будущей конкурентоспособности и успешности / А. А. Новожилова, Е. А. Шовгенина // Вестник Волгоградского государственного университета. Серия 6, Университетское образование. – 2013. – № 14. – С. 70–76.

14 Орел, М.А. Словарь переводчику – друг, товарищ и Брут / М.А. Орел // Перевод: информационные технологии. – М.: Всероссийский центр переводов науч.-техн. лит. и документации, 2009. – С. 79-106.

15 Переводоведческая лингводидактика: учебно-методическое пособие [Электронный ресурс]. Режим доступа: <https://librolife.ru/g4303518>

16 Усачева, А.Н. Перевод: от лингвистической теории к когнитивной модели / А.Н. Усачева // Вестник Волгоградского государственного университета. Серия 2, Языкознание. – 2011. – № 1 (13). – С. 131–137.

17 Хроленко, А.Т. Современные информационные технологии для гуманитариев: практическое руководство / А.Т. Хроленко, А.В. Денисов // М.: Флинта; Наука, 2007. – 128 с.

18 АBBYU LINGVO: [сайт]: - Режим доступа: <http://www.lingvo.ru/>

19 Benjamin Phister, «L'informatique, une aide pour le traducteur. [Электронный ресурс]. – Режим доступа: <http://journals.openedition.org/traduire/251>

20 Dictionnaire.net: [сайт]: - Режим доступа: <http://www.dictionnaire.net>

21 Larousse : [сайт]: - Режим доступа: <https://www.larousse.fr>

22 Le dictionnaire.com: [сайт]: - Режим доступа: <http://www.le-dictionnaire.com>

23 Le petit Robert / Малый Робер: [сайт]: - Режим доступа: <https://www.lerobert.com>

24 PROMT: [сайт]: - Режим доступа: <http://www.promt.ru/>

25 Wordfast Anywhere: [сайт]: - Режим доступа: <http://www.freetm.com/>

Приложение А

Vocabulaire (словарь компьютерных терминов)

enter, ввод (клавиша) – *entrée (f)* – клавиша на клавиатуре компьютера, чаще всего служит для выбора, подтверждения, а также для перехода на новую строку.

shift (клавиша) – *maj, majuscule (f)* – клавиша на клавиатуре компьютера, служит для изменения значения других клавиш, в частности, для набора заглавных букв.

амперсанд (знак &) – *et (m) commercial* – знак &, обозначающий and (союз «и» в английском, а теперь и не только в английском языке).

байт – *octet (m)* – единица информации, равен восьми битам, или двоичным знакам; можно также сказать, что представляет собой символ текстового файла (кроме языков со сложным письмом, напр., китайского, где символ занимает два байта в памяти компьютера).

буфер обмена – *presse-papier (m)* – участок памяти, в котором хранятся данные, которые необходимо перенести из файла в файл или из папки в папку (часть текста, изображения, целый файл или даже множество файлов и папок).

быстрый запуск (в Windows) – *lancement (m) rapide* – панель инструментов, позволяющая быстро обращаться к некоторым популярным приложениям.

вид (пункт меню) – *affichage (m)* – пункт меню, при активации которого пользователь может выбрать желаемое представление окна, документа.

видеокарта, видеоадаптер; видюха (*разг.*) – *carte (f) graphique* – устройство, транслирующее изображение на монитор.

вкладка (диалога) – *onglet (m)* – ссылка или закладка в окне диалога, являющаяся заголовком той или иной страницы этого диалога.

вставить (из буфера обмена) – *coller* – получить данные из буфера обмена и вставить их в текущей позиции.

выбирать, выделять (элемент списка) – *sélectionner* – выбирать нужный элемент (файл, папку, настройку).

вырезать (в буфер обмена) – *couper* – перемещать элемент из текущей позиции в буфер обмена, не оставляя его копии в текущей позиции.

гигабайт, Гб; гиг, гектар (*разг.*) – *gigaoctet (m)* [*jjigaoctet*], *Go*; *giga (m)* (*разг.*) – единица информации, равная 1 024 мегабайтам, или 1 048 576 килобайтам, или 1 073 741 824 байтам, обычно в гигабайтах измеряют объём жёстких дисков, интернет-трафика.

данные – *données (f, pl)* – информация вообще; какая-либо информация, которую нужно передать, разместить, сохранить, обработать.

двухъядерный – *à double cœur* – микропроцессор с двумя ядрами; таким образом, повышается быстродействие и многозадачность компьютера.

диалоговое окно, диалог – *boîte (f) de dialogue* – окно, предназначенное для действий пользователя (выбора из списков, отметки флажков и радиокнопок).

директория – *répertoire (m)* – группа файлов, имеющая название, может содержать любое количество файлов любого формата.

дисплей – *écran (m)* – экран (монитора, мобильного телефона, mp3-плеера).

драйвер; дрова (*pl, разг.*) – *pilote (m)* – программа, способствующая содействию того или иного устройства с компьютером.

жёсткий диск; винчестер, винт, хард (*разг.*) – *disque (m) dur* – магнитный носитель информации в виде одной или (чаще) нескольких металлических пластин в герметичном пластиковом корпусе;

жидкокристаллический – *à cristaux liquides* – основанный на жидких кристаллах.

зависать (о компьютере); виснуть, висеть (*разг.*) – planter – прекращать реагировать на действия пользователя вне зависимости от характера этих действий.

загружать(ся); грузить(ся) (*разг.*) – (se) démarrer – запускаться, приходиться в рабочее состояние (о компьютере); добавляться в память (о программе).

закладка (в тексте) – signet (*m*) – место в файле (в тексте, музыкальном произведении, фильме), к которому позднее можно обратиться.

запустить (приложение) – lancer – добавить приложение в оперативную память компьютера и начать его работу.

звуковая карта, звуковая плата; звуковуха (*разг.*) – carte (*f*) son – устройство, транслирующее звук в колонки или наушники, подсоединённые к компьютеру.

значок; иконка (*уст.*) – icône (*f*) – изображение, сопоставленное с определённым типом файлов, программой, диском.

клавиатура; клавиша (*разг.*) – clavier (*m*) – устройство ввода информации в компьютер.

клавиша – touche (*f*) – клавиша клавиатуры компьютера.

кнопка – bouton (*m*) – кнопка любого устройства, в том числе на корпусе компьютера, а также изображение в виде кнопки на экране.

компьютер; комп, машина (*разг.*) – ordinateur (*m*); ordi (*m*), machine (*f*) (*разг.*) исторически: вычислительная машина, выполняющая программы; в современном мире это понятие сузилось до персональной электронной вычислительной машины, предназначенной для выполнения различного рода программ.

компьютерная игра; игрушка, игруха (*разг.*) – jeu (*m*) vidéo – игра, предназначенная для установки на компьютер и/или игровую консоль.

копировать (в буфер обмена) – copier – перемещать элемент из текущей позиции в буфер обмена, оставив его копию в текущей позиции.

корзина (Windows) – corbeille (*f*) – место на диске, предназначенное для хранения случайно удалённых файлов.

корпус (комп.) – boîtier (*m*) – внешний кожух компьютера, содержащий все его основные компоненты.

курсор – curseur (*m*) – индикатор фокуса, то есть того конкретного места в окне, где находится в данный момент пользователь.

материнская плата, системная плата; мать, мама, мамка (*разг.*) – carte (*f*) mère – основная плата компьютера, на которой подключены все его модули и основные устройства, представляет собой металлическую пластину размером с одну из стенок корпуса.

мой компьютер (в Windows) – poste (*m*) de travail – элемент управления в Microsoft Windows, позволяющий осуществлять доступ к носителям информации данного компьютера.

мышь (комп.); мыша (*разг.*) – souris (*f*) – ручной манипулятор, с помощью которого выполняются наиболее частые действия.

обратный слеш, обратная косая черта (знак \) – barre (*f*) oblique renversée – знак \, часто употребляемый в записи пути к файлам.

окно – fenêtre (*f*) – область экрана, обладающая определёнными свойствами.

операционная система; операционка, ось (*разг.*) – système (*m*) d'exploitation; OS (*m*) (*разг.*) [oèsse] – среда, в которой пользователь взаимодействует с компьютером.

отмена (кнопка) – annuler – кнопка, позволяющая отменить изменения и выйти из диалогового окна.

панель инструментов – barre (*f*) d'outils – полоса кнопок, дающих быстрый доступ к часто употребляемым действиям (операции с файлом, правка и форматирование текста).

папка, каталог – dossier (*m*) – см. директория.

перезагружать(ся) – (se) redémarrer (*vt*) – выключать(ся) и снова включать(ся).

ПК – PC (*m*) [se prononce à la française] – персональный компьютер.

по умолчанию – par défaut – значение того или иного параметра, устанавливаемое разработчиками.

порт – port (*m*) – гнездо для подключения устройств к компьютеру.

применить (кнопка) – appliquer – кнопка, служащая для подтверждения изменения настроек.

пробел (клавиша) – barre (*f*) d'espacement – клавиша на клавиатуре компьютера, вводящая пробел между словами, имеет большие, чем другие клавиши, размеры, потому и названа на других языках «линией», «строкой» (*barre*).

пробел (символ) – espace (*m*) – пробел между словами, который с точки зрения компьютера также является символом со своим кодом.

программное обеспечение; софт (*разг.*) – logiciels (*m, pl*) – комплекс программ, предназначенных для одной конкретной цели.

прожечь, записать (на CD, DVD) – graver – записать информацию на лазерный диск (CD, DVD).

рабочий стол (Windows, etc.) – bureau (*m*) – пространство, занимающее основную часть экрана при запуске Windows.

раскладка (*ж*) клавиатуры – disposition (*f*) du clavier – определённый набор значений буквенно-цифровых клавиш, часто связанный с языком и страной использования.

свойства (файла, etc.) – propriétés (*f, pl*) – сведения об объекте: его размер, тип, размещение.

системный блок – unité (*f*) centrale – корпус со всеми внутренними структурами компьютера вместе.

скриншот – capture (*f*) d'écran – снимок экрана в конкретный момент.

слеш, косая черта (знак /) – barre (*f*) oblique – знак /, часто употребляемый в адресах Интернета.

строка (таблицы) – rang (*m*) – горизонтальная строка таблицы, часто ограниченная сверху и снизу чертой.

строка (текста) – *ligne (f)* – группа символов текста, умещающаяся на экране, или принудительно прерванная пользователем.

устанавливать отметку (на флажок, радиокнопку) – *cocher (vt)* – поставить отметку на элементе управления, таким образом выбрав возможность, предлагаемую им.

файл – *fichier (m)* – единица данных; единый комплекс информации: документ, фотография, песня, фильм.

формат – *format (m)* (в формате *txt* – *au format txt*) – тип файлов.

чип – *puce (f)* – микросхема, предназначенная для определённых целей (напр., для проверки отпечатков пальцев).

щёлкнуть (мышью); кликнуть (*разг.*) – *cliquer* (по – *sur qch.*) – нажать левую кнопку мыши, поставив указатель на объект.

щелчок (мышью); клик (*разг.*) – *clac (m)* – нажатие кнопки мыши на объекте; сопровождается характерным звуком, откуда и пошло название.

ярлык – *raccourci (m)* – ссылка на тот или иной объект, выносимая, напр., на рабочий стол для удобства пользователя.

Приложение Б

Интернет-ресурсы

- 1 Диалог: Международная русскоязычная конференция по компьютерной лингвистике. – Режим доступа <http://dialog-21.ru>
- 2 Корпусная лингвистика. Машинный перевод. Прикладная лингвистика // Фонд знаний «Ломоносов». – Режим доступа <http://www.lomonosov-mnd.ru/enc/ru/encyclopedia>
- 3 Корпусная лингвистика: тематический сайт СПбТУ и ИЛИ РАН. СПб., 2008. – Режим доступа <http://corpora.iling.spb.ru>
- 4 Лаборатория компьютерной лингвистики Института проблем передачи информации РАН. – Режим доступа <http://proling.iitp.ru/ru/node/1>
- 5 Международный каталог поисковых машин. – Режим доступа <http://www.searchenginecolossus>
- 6 Национальный корпус русского языка. – Режим доступа <http://www.ruscorpora.ru/index.html>
- 7 Переводческие глоссарии. – Режим доступа <https://www.proz.com/>
- 8 Портал переводчиков. – Режим доступа <http://translations.web-3.ru/>
- 9 Программы лингвистического анализа и обработки текста. – Режим доступа <http://asknet.ru/>
- 10 Речевые технологии. – Режим доступа <http://speech-soft.ru/index.php>
- 11 Русскоязычный каталог сайтов www.ru. – Режим доступа <http://www.rus-url.ru/>
- 12 Словарь английских аббревиатур. – Режим доступа <https://www.abbreviations.com/>
- 13 Справочник технического переводчика. – Режим доступа <http://intent.gigatran.com/>
- 14 Тезаурусы для описания отдельных предметных областей. – Режим доступа <https://www.thesaurus.com/>
- 15 Этимологический онлайн-словарь. – Режим доступа <https://www.etymonline.com/>