

На правах рукописи

АЛЬОТАИБИ СУЛТАН МАДЖЕД

**ПРЕОБРАЗОВАНИЕ СЕМАНТИЧЕСКОЙ ИНФОРМАЦИИ В
МАШИННОМ ПЕРЕВОДЕ (НА МАТЕРИАЛЕ РУССКО-АРАБСКОГО
МАШИННОГО ПЕРЕВОДА ПОЛИТИЧЕСКИХ ТЕКСТОВ)**

Специальность 10.02.19. – теория языка

АВТОРЕФЕРАТ

Диссертации на соискание ученой степени

Кандидата филологических наук

Москва 2011

Работа выполнена на кафедре общего сравнительно-исторического языкознания филологического факультета ФГОУ ВПО «Московский государственный университет имени М. В. Ломоносова».

Научный руководитель:

доктор филологических наук, профессор **Марчук Юрий Николаевич**

Официальные оппоненты:

Крюкова Ольга Павловна

доктор филологических наук, профессор Научно-исследовательский
технологический университет
«МИСиС» - НИТУ «МИСиС»

кандидат филологических наук

Пак Виктор Михайлович

Кафедра арабской филологии
института стран Азии и Африки
ФГОУ ВПО «Московский
государственный университет
имени М. В. Ломоносова»

Ведущая организация:

Учреждение Российской
академии наук «Институт
востоковедения РАН»

Защита состоится 29 июня 2011 г. в ____ часов ____ минут на заседании диссертационного совета Д 501.001.24 при ФГОУ ВПО «Московский государственный университет имени М. В. Ломоносова» по адресу: 119991, Москва, Ленинские горы, МГУ им. М. В. Ломоносова, 1-й корпус гуманитарных факультетов, филологический факультет.

С диссертацией можно ознакомиться в читальном зале 1-го корпуса гуманитарных факультетов ФГОУ ВПО «МГУ им. М. В. Ломоносова».

Автореферат разослан 26 мая 2011 г.

Ученый секретарь

Диссертационного совета

А. М. Белов

Общая характеристика работы

Реферируемое диссертационное исследование посвящено проблеме преобразования смыслового содержания в машинном переводе с русского языка на арабский. Важность данной проблемы обусловлена необходимостью решения как практических, так и теоретических проблем машинного перевода. Новейшие исследования в данной области все еще не в полной мере отражают все стороны данного вопроса. Таким образом, возникает острая необходимость в продолжении изучения вышеупомянутой темы.

Для преодоления основных трудностей проблемы машинного перевода должны быть решены задачи автоматизированного представления контекста, смыслового содержания переводимого текста, знаний о понятиях предметной области, к которой относится переводимый текст.

Перспективы развития машинного перевода связаны с дальнейшей разработкой и углублением теории и практики перевода, как машинного, так и «человеческого». Для развития теории важны результаты сопоставительного языкознания, общей теории перевода, теории закономерных соответствий, способов представления знаний, оптимизации и совершенствования лингвистических алгоритмов. Новые и более эффективные словари с необходимой словарной информацией, строгие теории терминологизации лексики, теория и практика работы с подязыками помогут повысить качество перевода лексических единиц. Формальные грамматики, ориентированные на перевод, дадут возможность оптимизировать алгоритмы нахождения переводных соответствий в данной коммуникативной ситуации, которая может быть описана в рамках соответствующих прикладных теорий представления знаний. Наконец, новые возможности программирования и вычислительной техники также будут вносить свой вклад в совершенствование и дальнейшее развитие теории и практики машинного перевода.

Об **актуальности** избранной темы свидетельствуют разные факторы. Во-первых, объем переводов в мире неуклонно возрастает по мере того, как новые языки вступают в сферу мировой коммуникации. Перевод, являясь древнейшей областью применения интеллекта, культурного общения, фундаментом образования человеческого общества планеты Земля, в настоящее время, как и в течение всей истории, совершенствуется, благодаря новым информационным технологиям. Если на заре человеческого общества такими новыми технологиями были письмо и письменность, то в настоящее время перевод осуществляется с помощью новейших средств информатики, компьютерных технологий и современных способов коммуникации. Во-вторых, перевод органически включает большое количество проблем, которые отражаются в растущем наборе теорий и моделей перевода. Главная задача перевода – передавать в язык перевода из языка оригинала смыслы и содержание последнего. Однако что такое смысл и содержание? Каждая теория и модель перевода по-своему отвечает на этот вопрос. От принципиальной невозможности перевода (В. Гумбольд) до абсолютной адекватности пословного (буквального) перевода (Ю.В. Рождественский) функционируют самые различные теории и модели перевода, как для естественных, так и для искусственных языков. В-третьих, арабская и российская библиотеки, к сожалению, не насчитывают научных трудов, рассматривающих проблемы машинного перевода политических текстов с русского языка на арабский. Исследования в области машинного перевода с русского языка на арабский крайне редки, и даже можно сказать, что они практически не существуют. Такое неудовлетворительное положение

потребовало удвоения взятых усилий по анализу проблем систем машинного перевода в данной языковой паре. Данная работа представляет собой первую попытку и стартовый шаг в изучении проблем преобразования семантической информации в машинном переводе политических текстов (газетный интервью) с русского языка на арабский. Кроме того, на нынешнем рынке количества бесплатных современных онлайн-систем машинного перевода с арабского языка на русский и наоборот очень мало. Их даже можно считать по пальцам: google translator, windows live translator (bing) и ImTranslator. Данное скудное положение современных систем машинного перевода в русско-арабской языковой паре на нынешнем рынке требует новых идей и современных исследований. Основные положения данного исследования могут быть полезными для продвижения и усовершенствования существующих систем машинного перевода.

Объектом настоящего исследования являются известные теории и модели перевода, рассматриваемые с точки зрения их возможностей передачи смысловой (содержательной) семантической информации. Модели перевода как процесса требуют использования самых различных сведений как лингвистического, так и экстралингвистического порядка. Практически целью каждой модели является передача семантического содержания, точнее того, что под этим понимается в каждой модели. Объектом исследования являются также современные системы машинного перевода, результаты их работы, качество и перспективы совершенствования. На материале русско-арабского машинного перевода политических текстов, имеющих характер интервью, изучаются положительные и отрицательные стороны действующих моделей машинного перевода.

В качестве **предмета** исследования выступают образцы машинного перевода политического текста с русского языка на арабский (газетный интервью) и те ошибки, а также и правильные решения, которые можно видеть на материале таких переводов. Формальный и алгоритмический подход, реализуемый системами машинного перевода, иногда даже выигрывают по сравнению с переводом «человеческим». Здесь следует упомянуть о том, что английский термин human translation в противопоставление к переводу машинному не имеет точного соответствия и поэтому приходится в данной работе пользоваться термином «человеческий» перевод.

Цель работы: сравнить возможности существующих теорий и моделей перевода с точки зрения их допустимой формализации и эффективности в передаче информации. При этом в исследовании проводится разница между понятиями «теория перевода» и «модель перевода». Теория есть некоторое абстрактное, теоретическое, почти философское построение, объясняющее такое сложное лингвистическое явление как перевод. Модель – это реальное воплощение теории в виде некоторого набора алгоритмов преобразования текстов и словарей сопоставляемых языков.

Задачами исследования являются:

- а) провести детальный анализ полученных образцов машинного перевода русского политического текста (газетный интервью) на арабский язык, проанализировать как положительные результаты, так и ошибочные переводы;
- б) уточнить причины появления ошибок;
- в) наметить пути положительного решения затронутой проблемы;
- г) сформулировать требования к лингвистическому содержанию соответствующей модели перевода.

Метод лингвистического анализа: сопоставительный анализ переводимых текстов и результатов перевода.

Материал исследования: тексты электронных газет, переведенные профессиональными переводчиками с русского языка на арабский из известного арабскоязычного сайта Russian Today. «Человеческие» переводы сравниваются с машинными переводами, которые сделанные с помощью системы Windows Live Translator. В основном анализируются ошибки в переводе на семантическом уровне и сравниваются с правильными переводами, сделанными человеком-переводчиком. Результаты сравнения по возможности систематизируются и обобщаются. В необходимых случаях используются данные известных словарей русского и арабского языков.

Научная новизна исследования заключается в том, что:

- 1) Впервые проанализированы результаты машинного перевода с русского языка на арабский;
- 2) Предпринята попытка сравнить возможности современных теорий перевода обеспечить передачу смысла при помощи алгоритмов, базирующихся на формальных признаках текста;
- 3) Подчеркнуто различие в понятиях «теория перевода» и «модель перевода»;
- 4) Намечены пути использования контекста для совершенствования результатов работы системы машинного перевода;

Теоретическая значимость работы может быть усмотрена в том, что в широком масштабе сопоставляются теории и модели перевода как общелингвистической задачи с возможностями и современным состоянием систем машинного перевода, базирующимися на чисто формальной основе анализа и синтеза текстов.

Практическая значимость исследования – возможные рекомендации для использования систем машинного перевода в русско-арабской языковой паре для актуального перевода современных текстов политического характера (интервью), в ситуациях, когда «человеческий» перевод требует много времени, а ситуация вызывает необходимость быстрого перевода.

Положения, выносимые на защиту:

- 1) Современные теории перевода всесторонне описывают процесс и результат перевода, в том числе и особенно теория уровней эквивалентности, который больше подходит для перевода с русского на арабский;
- 2) Между теорией перевода и моделью перевода есть различие, которое заключается в том, что теория дает теоретическое основание для моделирования перевода, а модель есть конкретное воплощение положений теории, которое может быть положено, в частности, в основу алгоритмов преобразования исходного текста в текст перевода, и в организацию соответствующих словарей;
- 3) Современные системы машинного перевода дают положительный результат в том случае, если их структура и словари отвечают требованиям лингвистических характеристик переводимых текстов;
- 4) В рамках машинного перевода существуют эффективные средства добиться качественного перевода и адекватной передачи семантической информации,

такие, например, как контекстологические словари для определенных типов текстов.

Апробация работы: Теоретические положения и практические результаты данной работы обсуждались на заседании кафедры общего и сравнительно-исторического языкознания филологического факультета МГУ им. Ломоносова (2011). По материалам исследования были сделаны сообщения на научной конференции «Ломоносовское чтение» в Московском Государственном Университете им. Ломоносова, Интернет-конференции «Русский язык, культура, литература» (2010), конференции ««РКИ в современной образовательной и геополитической парадигме» (2010), VII международной конференции синхронное и диахронное в сравнительно-историческом языкознании (2011).

Структура исследования:

Диссертация состоит из введения, трех глав, заключения и списка литературы, включающего 127 наименований. Библиография включает список источников, использованных при работе над настоящим диссертационным исследованием, а также материалов из информационной сети Интернет для выполнения практической части исследования.

Основное содержание работы

Во *Введении* обосновывается актуальность исследования, его теоретическая и практическая значимость, аргументируется выбор предмета исследования, определяются цель, задачи работы, методы лингвистического анализа.

В *первой главе «Основные понятия языкового перевода»* рассматриваются виды, жанры перевода, излагается исторический очерк о возникновении и развитии переводоведения. Также рассматриваются основные теории перевода, касающиеся описания последовательности действий переводчика в процессе перевода, и разработки теоретических принципов, на основе которых переводчик мог бы принимать оптимальные решения для достижения поставленных целей.

Существуют две основные классификации видов перевода: первая классификация основывается на характере переводимых текстов (жанрово-стилистические особенности оригинала), а вторая - на характере речевых действий переводчика в процессе перевода (психолингвистические особенности речевых действий в письменной и устной форме). Жанрово-стилистическая классификация переводов в зависимости от жанрово-стилистических особенностей оригинала выделяет два функциональных вида перевода: художественный перевод и специальный перевод (Комиссаров В.Н., 1990: 95). Считая, что общественно-политический вид в отличие от других видов характеризуется пропагандистской или агитационной установкой, В.В. Алимов приводит более обширную классификацию, состоящую из трех видов перевода, позволяющую определить объект каждого вида перевода: художественный, общественно-политический, и специальный (Алимов В. В., 2006:26).

Художественным переводом называется перевод художественной литературы. Его объектом являются художественные произведения. Основная задача любого художественного произведения заключается в достижении образно-эмоционального и эстетического воздействия на читателя. В целях достижения определенного эстетического воздействия на переводящий язык (ПЯ) используется огромное количество разнообразных языковых средств, от эпитета (красочное определение) до ритмико-синтаксического построения фразы (Алимов В. В., 2006: 26). Такая

эстетическая направленность отличает художественный перевод от остальных видов перевода. Следует отметить, что помимо трех выделяемых Алимовым подвидов в художественном переводе, В. Н. Комиссаров выделяет еще перевод пьес, перевод сатирических произведений, перевод текстов песен (Комиссаров В.Н., 1990: 97).



Общественно-политическим переводом называется перевод текстов общественно-политического и публицистического характера с пропагандистской или агитационной установкой. Именно эти тексты являются объектом общественно-политического перевода. Общественно-политический перевод характеризуется яркой эмоциональной окраской с большой насыщенностью различной терминологии.

Специальный перевод обслуживает различные предметные отрасли знаний, имеющие специфическую терминологическую номенклатуру (Нелюбин Л. Л., Хухуни Г.Т., 2008: 7). Следовательно, объектом специального перевода являются материалы, которые относятся к различным сферам человеческого знания и практики науки и техники. Эти материалы характеризуются предельно точным выражением мысли, следовательно, широким использованием терминологии (Алимов В. В., 2006: 28).

Под «*формами перевода*» понимается способ, при котором осуществляется перевод. Исходя из этой точки зрения, выделяются **письменный** и **устный** переводы, а также их разновидности. Письменный перевод включает в себя следующие разновидности: письменно-письменный, зрительно-письменный, письменный перевод на слух (или устно-письменный). В устном переводе обычно выделяются следующие разновидности: *устный перевод на слух*, *зрительно-устный перевод* или *перевод с листа*, т. е. устный перевод зрительно воспринятого исходного письменного текста. В устном переводе на слух выделяются *последовательный* и *синхронный переводы*. Ввиду того, что переводчик испытывает при синхронном переводе нервное и физическое напряжения, этот вид перевода считается самым трудным, требующим особенных навыков и умений.

На основе соотношения содержания и формы ПЯ с содержанием и формы ИЯ выделяются следующие **типы** перевода: *вольный (свободный), дословный, буквальный, пословный и адекватный (эквивалентный)*. *Вольным* называют перевод, при котором основную информацию оригинала передается на другом языке с возможными отклонениями – добавлениями, пропусками и т. п. При *дословном переводе* воспроизводятся семантико-структурные характеристики языка оригинала на языке перевода. Данный тип перевода очень заметен в переводах сакральной литературы. В *буквальном переводе* наблюдается стремление к сохранению формальных и семантических компонентов оригинала при передаче его на другой язык. Другими словами, он предлагается рассматривать текст не как последовательность отдельных слов, а как последовательность предложений. *Пословный* перевод представляет собой полексемную передачу смысла и содержания оригинала, при которой учитываются синтаксические и стилистические соотношения между ИЯ и ПЯ. «*Адекватным переводом* называется перевод, который обеспечивает прагматические задачи переводческого акта на максимально возможном для достижения этой цели уровне эквивалентности, не допуская нарушения норм или узуса ПЯ, соблюдая жанрово-стилистические требования к текстам данного типа и соответствуя общественно-признанной конвенциональной норме перевода» (Комиссаров В.Н., 1990: 233).

Что касается истории перевода как профессии, то она восходит, по словам Л. Л. Нелюбина, к шумерской цивилизации конца четвертого тысячелетия до н.э. и к ранним этапам существования египетской цивилизации, а именно к эпохе Древнего Царства и периода XXVIII в. до н.э. (Нелюбин Л. Л., 2009: 30). Найдены древние рукописи показывают, что переводчики в Древнем Египте представляли особые профессиональные группы и имели собственную иерархию. В конце третьего тысячелетия до н.э. в Аккаде отмечено существование специальных образовательных учреждениях (э-дубах, «домах табличек»), выпускники которых помимо того, что упражнялись в знании шумерского языка, еще переводили письменно и устно с аккадского на шумерский и обратно. История перевода показывает, что переводческое дело проходит через всю историю человечества, через цивилизации Востока и Запада: переводческие памятники исследователи находят и у финикийцев, и у карфагенян, и в иранской цивилизации у древних персов, и в древней Индии, и в Китае, и в Японии. Греко-римская цивилизация — это эпоха в истории перевода, когда, во-первых, в полной мере осознается просветительский потенциал перевода, а во-вторых, впервые формулируются некоторые основополагающие его принципы. Фактически в 240 г. до н.э. началась эпоха активной переводческой деятельности в Древнем Риме, связанной с усвоением богатейшего наследия древнегреческой литературы. Что же касается принципов перевода, сформулированных греко-римской цивилизацией (пусть еще в самых общих чертах), то некоторые из них не утратили своей актуальности до сих пор. Так, именно со времен Античности утвердилось противопоставление двух типов перевода — буквального и вольного. Первым устным переводчиком, упоминаемым в истории Рима, был римский сенатор Гай Ацилий. В 155г. до н.э. он выступил в Сенате устным переводчиком при приеме греческого посольства (Нелюбин Л. Л., 2009: 34). Греко-римская античность также оставила нам сведение о переводческой деятельности в области литературы, ибо литературоведческий перевод достиг в Риме особого расцвета. Здесь следует упомянуть такие имена Луция Ливия Андронника (275-200 гг. до н.э.), как родоначальника перевода письменных

памятников и значинателя римской литературы, Квинта Энния (239-169 гг. до н.э.) – драматурга и переводчика, владевшего тремя языками: родным окским, греческим и латинским; Публиция Теренция Аффа (190-159 гг. до н.э.) – драматурга и переводчика и др. Первым, кто затронул вопросы теоретических проблем перевод, был Марк Тулий Цицерон (106-43 гг. до н.э.) – яркая фигура в политической и литературной жизни Рима, создатель норм латинского языка. Цицерон всегда выступал против дословного перевода. По его мнению, дословный прием отражает языковую бедность и беспомощность переводчика. В своих теоретических основах художественного перевода он выдвинул положение о том, что перевод должен быть рассчитан на реципиента (получателя информации), сообразен с законами языка оригинала, должен отвечать эквивалентности перевода, соблюдать использование реалий языка оригинала, отвечать стилю и языку подлинника, отражать его смысл. Что касается формы перевода, то в Древнем Мире существовали практически все известные нам ныне формы перевода: устная и письменная. В 1130 г. в Толедо по инициативе Великого канцлера Кастилии Раймундо была организована школа переводчиков с арабского языка. Наряду с литературными произведениями переводчики Толедской школы переводили труды по философии, астрономии, медицине. Благодаря трудам толедских переводчиков европейцам стали доступны достижения арабской науки и культуры: труды по математике, астрономии, физике, алхимии, медицине. На Руси центром переводческой деятельности был Новгород. Первая полная Библия на Руси появилась в XV в. При архиепископе Геннадии в 1499 г. был завершен перевод новой Библии на церковнославянский язык. Новой ее можно назвать условно, ибо она включала значительную часть, хотя и отредактированных, но старых переводов. Первый перевод Библии на русский язык с греческого был осуществлен лишь в середине XIX века. В 1976 году отмечалось 100-летие издания Библии на русском языке. Вплоть до середины прошлого века перевод считался литературоведческой наукой. На протяжении эпохи Средневековья, всего периода Нового времени возникали различные суждения, взгляды, положения о том, каким должен быть переведенный художественный текст, что является основным и важным в переводе, какие типы перевода должны быть основополагающими. С первых десятилетий XX века наблюдается возрастание интереса к переводу и теоретическому осмыслению проблем перевода. Начинают возникать школы и теории лингвистического перевода. Краеугольный камень прагматики перевода и теории перевода, как области лингвистических знаний, заложила вышедшая осенью 1953 года книга А.В. Федорова «Введение в теорию перевода», в которой сформулировались решительные мысли, касающиеся важности языковедческого подхода к проблеме. Основные задачи теории перевода А.В. Федоров видит в двух планах. С одной стороны следует анализировать проблемы эквивалентности при передаче различных элементов отдельных уровней структуры языка (лексики, фразеологии, грамматики), что достигается при учете широкого контекста, а с другой стороны, перевод связан с жанровой природой и спецификой переводимого материала. После выхода в свет работы А.В. Федоров разработкой лингвистических основ перевода занимались Я.И. Рецкер, Л.С. Бархударов, Р.К. Миньяр-Белоручев, А.Д. Швейцер, В.Н. Коммисаров, В.Г. Гак, А.Ф. Ширяев и др. в результате предпринятых теоретиками перевода изысканий были созданы и сформулированы некоторые теории и модели перевода. Под моделированием перевода понимается условное описание процесса перевода, т. е.

процедур порождения текста перевода переводчиком. Или как определяет В.Н. Комиссаров «условное изображение процедуры осуществления процесса перевода» (Комиссаров В.Н., 1999:158). Реальный процесс перевода недоступен для непосредственного исследования. Он осуществляется в сознании переводчика в момент преобразования содержания, выраженного в одной языковой форме, в то же содержание, материализованное в другой языковой форме. Поэтому, изучение процесса перевода производится косвенным путем при помощи разработки различных теоретических моделей (теорий), с большей или меньшей приближенностью описывающих процесс перевода в целом или какую-либо его сторону. Задачи модели перевода заключаются в систематизации и описании последовательности действий, которые осуществляет переводчик в процессе перевода с одного естественного языка на другой, и в разработке теоретических принципов, на основе которых переводчик мог бы принимать оптимальные решения для достижения поставленных целей. Однако, большинство моделей перевода ограничено по своей объяснительной силе, и на их основе не гарантируется реальный перевод любого текста с необходимой степенью эквивалентности: «Модель перевода носит условный характер, поскольку она необязательно отражает реальные действия переводчика в процессе создания текста перевода. Большинство таких моделей имеет ограниченную объяснительную силу и не претендует на то, что на их основе может быть реально осуществлен перевод любого текста с необходимой степенью эквивалентности» (Комиссаров В.Н., 1990: 158-159).

Наиболее распространенными в настоящее время гипотетическими моделями процесса перевода являются: *ситуативная, семантическая, трансформационная, семантико-семиотическая, закономерных соответствий, коммуникативно-функциональная, информативная, теория уровней эквивалентности и др.* Каждая модель отражает какие-то определенные аспекты перевода, один аспект или несколько. Диссертационное исследование показало, что среди всех рассмотренных в данной работе теорий больше подходит для перевода с русского языка на арабский теория уровней эквивалентности, ибо она сохраняет цель коммуникации на всех уровнях и учитывает как лингвистические, так и экстралингвистические факторы. Суть этой теории заключается в потенциальном ограничении возможности полного сохранения содержания оригинала в переводе из-за различий в системах ИЯ и ПЯ и особенностях создания текстов на каждом из этих языков. Поэтому переводческая эквивалентность может основываться на сохранении (и соответственно утрате) разных элементов смысла, содержащихся в оригинале. В передаваемом от Источника к Рецептору тексте В. Н. Комиссаров выделяет 5 основных уровней (типов) эквивалентных отношений между текстами оригинала и перевода для обеспечения переводческой эквивалентности: 1) *уровень цели коммуникации*; 2) *уровень описания ситуации*; 3) *уровень сообщения*; 4) *уровень высказывания*; 5) *уровень языковых знаков (слово)*. На любом уровне эквивалентности перевод может обеспечивать межъязыковую коммуникацию, т. е. сохраняются цель коммуникации и способ ее достижения через идентификацию определенной ситуации.

Первый тип (эквивалентность на уровне цели коммуникации):

Постыдился бы – عيبٌ عليك

Цель коммуникации здесь заключается в выражении негативной эмоции говорящего по отношению к собеседнику. В переводе употреблена одна из типичных для таких

ситуаций фраз, служащих для выражения возмущения в арабском языке, хотя составляющие ее языковые средства не соответствуют единицам оригинала.

Второй тип (эквивалентность на уровне описания ситуации):

Он снял трубку – اجاب على الهاتف

Этот тип характеризуется несколько большей общностью содержания разноязычных текстов (Нелюбин Л. Л., Хухуни Г. Т., 2008: 339). Общая часть содержания обоих текстов не только передает одинаковую цель коммуникации, но и отражает одну и ту же внеязыковую ситуацию, хотя использованные в обоих текстах языковые единицы несопоставимы.

Третий тип (эквивалентность на уровне способа описания ситуации):

Осторожно, стекло – قابل للكسر

При таком переводе кроме сохранения цели коммуникации и общности ситуации, сохраняются еще общие понятия, с помощью которых эта ситуация описана в исходном тексте, хотя ни синтаксическая структура, ни использованные в переводе слова не воспроизводят синтаксической структуры и значений слов оригинала.

Четвертый тип (эквивалентность на уровне структурной организации высказывания):

Я сообщил ему о том, какие чувства я испытываю перед ней

اخبرته عن مشاعري تجاهها

При таком переводе добавляется к описанным выше трем компонентам содержания еще инвариантность синтаксических структур оригинала и перевода, т.е. значительная часть значений синтаксических структур оригинала сохраняется в переводе. Иначе говоря: здесь уже сохраняются значения, заключенные в связях между языковыми единицами, однако эквивалентности слов нет.

Пятый тип (эквивалентности на уровне языковых знаков):

Письмо написано на арабском языке – الرسالة مكتوبة باللغة العربية

В этом типе эквивалентности сохраняются все содержательных компонентов для обеспечения полной переводческой эквивалентности: цель коммуникации, описание одинаковой ситуации, смысле сообщения, значение синтаксических структур и значение слов, т.е. достигается максимальная степень близости содержания оригинала и перевода.

Таким образом, В. Н. Комиссаров разработал универсальную типологию семантических состояний, в которых может пребывать исходный текст относительно переводного. Данная модель подсказывает переводчику, в каких пределах и какую часть содержания исходного текста он волен или не волен жертвовать. После анализа данной теории становится ясным, что ни при каких условиях нельзя жертвовать целью коммуникации. Однако это не значит, что эквивалентность на уровне цели коммуникации лучше, чем эквивалентность на уровне, например, языковых знаков. Выбор подходящего типа эквивалентности для перевода условен самой ситуацией и характером содержания оригинала. В целом эта теория позволяет переводчику определить, какой степени близости к оригиналу он может достичь в каждом конкретном случае.

Что касается понятий «теория перевода» и «модель перевода», то они рассматриваются в настоящей работе как разные термины. Теория перевода служит основанием для модели перевода, которая более конкретна и содержит не только описание подхода к переводу, но и определенные характеристики алгоритма перевода.

Модель можно уподоблять чертежу, по которому строится какой-либо проект (на практической почве), а теория представляет собой основные теоретические мысли, с которых исходит данный проект (на теоретической почве). Модель не может возникать без определенной теории, в то время как теория может и не приводить к созданию модели.

Во второй главе **«Модели машинного перевода»** излагается краткая история машинного перевода, рассматриваются общие характеристики и основные положения теорий «смысл↔текст» и «текст-текст», приводится анализ теории машинного перевода и модели перевода на основе переводных соответствий, и также рассматриваются современные системы машинного перевода и даются результаты их работы.

История машинного перевода насчитывает немногим более 50 лет. За это время сменилось несколько поколений систем машинного перевода - от первых программ, использовавших ограниченные ресурсы универсальных компьютеров первого поколения до современных коммерческих продуктов, использующих мощные ресурсы серверов и персональных компьютеров, включая ПК, в которых можно размещать карманные словари, а также компьютерные сети.

Системы первого поколения

Фактически история машинного перевода начинается с "Джорджтаунского эксперимента". В январе 1954 г. в Нью-Йорке состоялась первая публичная демонстрация системы машинного перевода с русского языка на английский, разработанной компанией IBM совместно с Джорджтаунским университетом (Hutchins W. J., 1995: 434). Для испытания был специально подготовлен словарь из 250 русских слов, записанных латинскими буквами. Слова были подобраны таким образом, чтобы каждое из них имело одно или, в крайнем случае, два английских значения. В отдельных случаях давались только корни, а некоторые слова были представлены полностью даже с наличными окончаниями (глаголы) или во множественном числе. Для перевода были выбраны простейшие русские фразы, состоящие из слов, которые входили в словарь, и разработаны шесть синтаксических правил, обеспечивающих с помощью кодовых чисел правильный перевод (Семенов А. Л., 2008: 6). Следующие десять лет разработка систем машинного перевода осуществлялась в США университетом Brigham Young University в Прово, штат Юта (ранние коммерческие системы WEIDNER и ALPS) и финансировалась Мормонской церковью, заинтересованной в переводе Библии; в Европе – группами GENA (Гренобль) и SUSY (Саарбрюкен). Особого внимания заслуживают труды русских лингвистов, таких как И. А. Мельчук и Ю. Д. Аперсян, результатом которых стал лингвистический процессор ЭТАП. В 1960 г. в состав Научно-исследовательского института математики и механики в Ленинграде вошла экспериментальная лаборатория машинного перевода, преобразованная затем в лабораторию математической лингвистики Ленинградского государственного университета. Системы «первого поколения» использовали пословный метод перевода без учета встроенного лингвистического компонента и не обладали средствами решения проблем многозначности, следовательно, допускали много ошибок.

Системы второго поколения

В 1980-х годах наблюдалось бурное продвижение в области вычислительной лингвистике, что предоставило исследователям систем машинного перевода

возможность перейти на более сложный уровень разработки. К переводу стали принимать «косвенный» подход, который был основан на определенных лингвистических правилах. Этот подход позволял анализировать текст на исходном языке и абстрактно репрезентировать его с использованием программ, которые устанавливали соответствие между словам и структурами предложений с попыткой решить проблему многозначности. До конца 1980-х гг. разработка систем машинного перевода опиралась на два научных подхода – прямой и непрямой. Непрямой подход представляли две системы: «интерлингва» и «трансфер». Хорошо известны системы машинного перевода прямого типа для универсальных вычислительных машин – Systran, Logos и Atlas. К универсальным вычислительным машинам относят большие, дорогие и наиболее сложные компьютеры, которые обрабатывают огромные данные, например, статистические. Не менее известны системы трансферного типа – Ariane фирмы GETA и EUTROTRA. Однако обе системы, в том числе и мультязычная система машинного перевода фирмы CICC, не давали того, что от них ожидалось. Большинство исследований этого периода обращали внимание на трансферный подход к переводу по переводным моделям или на интерлингвальный подход, т.е. через язык-посредник. Уровень потребности профессиональных переводчиков снижался. Поэтому некоторые коммерческие компании, как ALPNET, Inc, воспользовались возможностью предоставить на рынке ряд переводческих инструментов, одним из которых была первая версия системы класса «память переводов».

Современный период

Начало 1990-х гг. ознаменовалось огромным и весьма значительным событием: разработкой фирмой IBM системы машинного перевода под названием Candide, использующей «статистические методы». В то же время в Японии были ведены эксперименты с методикой, основанной на корпусах примеров переводов. Позже этот метод стал известен как «перевод по переводным примерам». Однако, ни статический метод, ни перевод по переводным моделям не опирались на синтаксические и семантические правила, полагаясь вместо них на выборку подобных эквивалентов из огромного электронного корпуса готовых текстов. А это значит, что они отличались от более ранних (до 1990 г.) методик. Идея статистического метода была взята из теории коммуникации, предложенной почти 60 лет назад в меморандуме У. Уивера. В отличие от других подходов, основанных на переводных моделях, новый метод анализа строится на сопоставлении текстовых пар из оригинала и переводящего языка, полагая, что на уровне предложения эти тексты структурно однотипны. Статистические подсчеты выполняются на сопоставлениях билингв для установления вероятностей различных переводных эквивалентов, которые могут быть извлечены из сопоставительных билингв по принципу совпадения (последовательностей слов, словосочетаний или предложений на исходном и переводящем языке). Несмотря на возникновение этих новых методов, старые методы продолжали свое существование. Тем не менее, произошел сдвиг в исследованиях машинного перевода от теоретических аспектов к прикладным. Эта тенденция продолжается до сих пор вместе с интеграцией систем машинного перевода с другими переводческими инструментами, особенно в области локализации программного обеспечения. Одновременно происходит расширение возможностей от автономных режимов к сетевым и от офлайновых к онлайн.

Что касается теории «смысл↔текст» (ТСТ), то она представляет собой многоуровневую модель непосредственного перехода от текста к его глубинной смысловой записи на некотором языке смысла, универсальном для всего человеческого общения, и затем обратный переход – от языка смысла к любому естественному языку на «поверхностном» уровне. Данная теория была впервые разработана И. А. Мельчуком в середине 1960-х гг. в Москве при активном участии ряда других российских лингвистов – прежде всего А. К. Жолковского, а также Ю. Д. Апресяна. ТСТ, по мнению её создателей, является универсальной, т.е. она может быть применима к любому языку. Однако она до сих не в состоянии реализовать систему семантических атомов (сем) (как это когда-то с неосторожностью прозвучало как обязательство), через которые можно представить любые смыслы. Это пока не получилось ни в этой, ни в какой-либо иной теории. Форма порождения смыслов в человеческом мозге недоступна для непосредственного наблюдения, поэтому трудно представить надежную блок-схему для установления верного порядка формулирования смыслов с самого начального этапа до последнего. А. И. Новиков пишет: «Одной из основных трудностей изучения такого явления, как смысл, является его непосредственная ненаблюдаемость. Косвенным проявлением смысла, как известно, могут служить разного рода вторичные тексты: пересказ своими словами исходного текста, аннотация, реферат, конспект, наконец, представление содержания текста в виде набора ключевых слов, основных тезисов, планов и др.» (Новиков А. И., 2002: 157). Десятилетия работ по машинному переводу показали, что трудности перехода на язык смысла от нормального естественного языка чрезвычайно велики. Прежде всего, само понятие «смысл» так и не было точно определено (Галактионов В. А., Мусатов А. М. и др., 2007: 20-21). Однако, следует упомянуть, что в своей работе «методы моделирования перевода» Ю. Н. Марчука различает смысл текста и смысл слова. Смысл текста представляет собой ролевую (смысловую) структуру, которую актуализируют поименованные в тексте объекты (предметы действительности, на которые указывают слова), а смысл слова – это указание на роль, которую играет поименованный этим словом объект, если слово есть имя объекта (Мурчук Ю. Н., 1985: 113). В. А. Звегинцев еще в 1976 году высказал достаточно убедительную мысль, что предложение, взятое вне дискурса (контекста), имеет не смысл, а «псевдосмысл». А разные перифразы одного и того же высказывания имеют и разные псевдосмыслы (Звегинцев В. А., 1976: 306). Перифразировки типа «*Охотник ударом ноги убил волка*», «*Ударом ноги охотника волк был убит*», «*Волк был убит охотником посредством удара ногой*» и пр. нельзя считать имеющими один и тот же смысл. Смысл определяется в конкретном дискурсе, и перестановка слов, замена словоформ, замена конструкций меняют и смысл, определяемый в зависимости от контекста, хотя, как полагает Ю.Н. Марчук, с некоторой абстрактной точки зрения можно утверждать, что некоторый «глубинный» смысл всех высказываний одинаков. Дело в том и заключается, что «глубинность» и «одинаковость смысла» является некоторой обманной величиной – смысл не существует вне контекста (Марчук Ю. Н., 2007: 128).

Второе направление в конструировании систем машинного перевода Ю. Н. Марчук называет подходом «текст-текст». Теоретические основы этого подхода представляются в моделировании переводческой деятельности человека-переводчика, которую можно явно разбить на два этапа. Первый этап – рутинный, т.е. его суть заключается в поиске по словарю незнакомых слов, уточнении значений,

морфологическом анализе, использовании устойчивых словосочетаний, фразеологизмов, оборотов и пр. этот этап простой. Другой этап сложный и творческий. Он состоит из поиска переводных соответствий в трудных случаях, разрешения неоднозначностей смыслового и синтаксического характера, поиска переводных эквивалентов для несовпадающих реалий. Неоднократно переводчик попадет в неловкую ситуацию, когда просто невозможно добиться адекватности перевода, полного соответствия перевода оригиналу по всем параметрам, поэтому теория человеческого перевода признает неизбежной потерей той или иной информации при всяком переводе. Реализация такой же человеческой творческой особенности в системах машинного перевода требует конструирования чрезвычайно мощного автомата, а это сложная задача и вряд ли достижимая на нынешнем уровне развития. В принципе теория «смысл↔текст» именно на это и направлена. Путь «текст-текст» вместо использования универсальных моделей опирается на рабочие, прикладные, воспроизводящие инженерно-лингвистические модели, более точно описывающие сущность конкретной задачи – найти переводное соответствие на языковом уровне, что представляет возможность избежать некоторых весьма серьезных трудностей первого подхода (Марчук Ю. Н., 1983: 10). Переводческие трудности, для решения которых требуется творчество, оставляются на разрешение человеку-редактору. Поэтому понятие «полностью автоматизированный высококачественный машинный перевод» как ближайшая задача, в настоящее время не ставится разработчиками промышленных систем.

Вообще говоря, между этими двумя подходами противоречий нет и один дополняет другой. Решение подобной проблемы Ю. Н. Марчук видит в создании некоторой универсальной достаточно общей модели, так или иначе описывающей языковую деятельность человека в целом или во всяком случае ту ее сторону, часть которой является перевода с одного языка на другой. Но универсальная модель, если речь идет о переводе, а не о нахождении смысла высказывания и пр. должна обязательно дополняться частными моделями, точно описывающими самую сущность перевода как языкового процесса. В системе АМΠΑР, НЕРПА и последующих использована рабочая прикладная модель перевода по переводным соответствиям, о которой речь пойдет в следующей части. В ней учтены особенности конкретных языковых пар, предусмотрены возможности пополнения словарей и схем алгоритма, перехода на новые тематики при условии создания новых тематических словарей. В нынешнем состоянии машинного перевода создание систем идет от прикладного конца, от конкретных пар языков, с использованием современных мощностей вычислительной техники. Теория «текст-текст» представляет собой дальнейшее развитие модели перевода по переводным соответствиям (МПС). Следует отметить, что современные трансферные системы машинного перевода основаны на концепции переводных соответствий, предложенной Марчуком Ю.Н. (Марчук Ю.Н., 1983: 56-125). Модель машинного перевода на основе переводных соответствий имеет два основных режима: генерацию и трансляцию. Этап генерации характеризуется преобладанием систематизации элементов описания: слов, единиц перевода, основных закономерностей преобразований. Данный этап предметно наполняет систему языковым материалом. Этап генерации имеет две основные задачи: исходное состояние системы – основной набор языковых данных, выявленных при нахождении переводных соответствий в исходном конкордансе и параллельных текстах, и более

точная настройка системы, ее словарей и алгоритмов анализа и синтеза на специфику данного подъязыка. Этап трансляции представляет собой непосредственное осуществление перевода на основе языкового материала. Кроме того, данный этап имеет также элементы пополнения словарей и алгоритмов системы, однако они вторичны по отношению к основной операции, т.е. осуществлению перевода.

В своей модели Марчук Ю.Н. различает два основных компонента: предметный и динамический. Предметный компонент отвечает на вопрос о том, **что** должно быть переведено, а динамический – **как** должен быть осуществлен перевод. В состав предметного компонента входят два основных элемента: предметный и переводной. Составляющие предметного элемента – лексика, грамматика, семантика. Каждая составляющая представляет собой множество элементов, связанных между собой. Лексика представляет список слов данного языка (подъязыка), грамматика – множество грамматических признаков данного языка, и наконец семантика – набор семантических признаков (или семантико-синтаксических, в той части синтаксиса, которая не охватывается грамматикой). Перечисленные составляющие во взаимодействии образуют текст. Переводной элемент предметного компонента образуется из типов переводных соответствий. В соответствии с принятой Марчуком Ю.Н. типологией переводных соответствий различаются три типа соответствий: *эквивалентные, вариантные и трансформационные*. Предметный элемент предметного компонента представляет исходную структуру данного компонента. На эту исходную структуру, как описывает Марчук Ю.Н., накладывается сеть, образованная составляющими переводного элемента. В целом предметный компонент отражает предмет и представляет статическое описание сущности исходных данных, над которыми производится операция перевода. Предметный компонент модели переводных соответствий есть специфически переводческое описание статики языковых данных, участвующих, необходимых и достаточных для перевода с данного естественного языка на другой, с учетом всех преобразований, которые могут для этого понадобиться. По своей концепции Марчук Ю.Н. представляет структуру предметного компонента следующим рисунком:

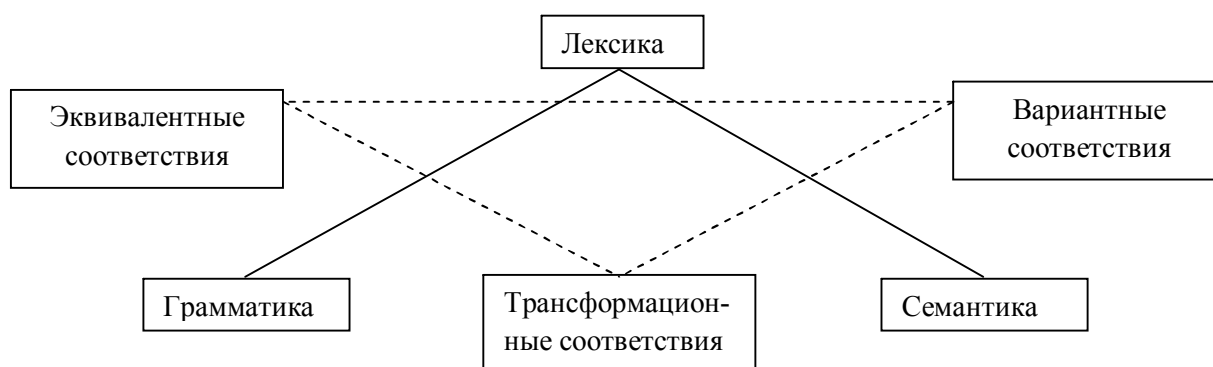


Рис. 1. Структура предметного компонента модели МПС.

Эквивалентные соответствия устанавливаются всегда простой операцией соотнесения. Суть данной операции заключается в поиске тождественных эквивалентов в словарях: так, в результате поиска входного слова в списке слов словаря может быть сразу получен постоянный переводной эквивалент этого слова (например, وقت «время»). Конечно, эквивалентность соответствий имеет место, как правило, только в рамках достаточно четко очерченных подъязыках, например, научно-технические термины. Вариантные соответствия суть такие, которые

устанавливаются с помощью некоторого формализованного выбора. В принципе любой из рассматриваемых в процессе обработки вариантов сможет передать основной смысл предложения. В зависимости от правильного выбора алгоритма можно говорить о большей или меньшей степени потери информации. Примером вариантных соответствий могут служить простые случаи контекстного разрешения многозначности отдельных слов. Так, существительное «свет» может переводиться на арабский язык либо как «ضوء-свет», либо «عالم-мир». Для решения подобных проблем необходимо создать автоматический словарь, устанавливающий семантические и контекстологические отношения между языковыми знаками и их референтами (об этом будем говорить подробно далее). Ясно также, что вариантные соответствия могут не охватить всего разнообразия переводных соответствий, кроме того, могут быть контекстные ограничения. Тем не менее, группа соответствий такого типа достаточно представительна, интуитивно допустима, безусловно существует в распоряжении переводчика и поэтому может быть выделена в составе модели (Марчук Ю.Н., 1976: 108). Трансформационные соответствия образуют самый сложный вид соответствий, для их нахождения требуются сложные алгоритмические операции анализа, синтеза и межъязыковых преобразований. В случае ошибки алгоритма следует уже говорить не о потере информации, а о невозможности правильной передачи смысла.

Задача динамического компонента заключается в установлении соответствий, динамике нахождения переводных соответствий в исходном тексте и конструировании некоторого выходного текста, находящегося со входным в отношении коммуникативной эквивалентности. Для выполнения такой задачи динамический компонент содержит два основных элемента: алгоритмического и программного. Алгоритмический элемент, в свою очередь, состоит из алгоритма, трансляционной грамматики и словаря. Алгоритм, который подразделяется на общий алгоритм, организующий систему в целом, и частные алгоритмы (этапов, частей этапов, отдельных решений анализа и синтеза), управляет использованием трансляционной грамматики и словаря на соответствующих стадиях нахождения соответствий и построения эквивалентного текста. Трансляционная грамматика представляет собой специфическую бинарную грамматику, которая оптимальным для перевода образом организует все необходимые для соответствий грамматические признаки. Сюда входят признаки всех языковых уровней: морфологического, синтаксического, семантического (или семантико-синтаксического) и лексического. Наконец словарь представляет собой удобно организованную систему собрания лексических единиц с необходимой информацией. Это собрание лексических единиц упорядочивается по трем основаниям: по признаку входного/выходного языка, однозначности/многозначности слов, фразеологических единств/отдельных лексем.

Программный элемент представляет собой ряд программных решений для этапов, назначенных и описанных в алгоритме. Составляющие программного элемента: списки, схемы лексического анализа и перевода, схемы грамматического анализа. Каждая из этих перечисленных составляющих отличается уровнем сложности программно-лингвистического аппарата: так, списки (сюда включаются также списки и таблицы для синтеза) реализуют простейшее соответствие типа словарей перевода однозначных слов и т. п.; схемы лексического анализа – перевод сложных словосочетаний и многозначных слов – характеризуются элементарным исходным набором стандартных операторов и моделей перевода; схемы грамматического

анализа (куда включаются также схемы, организующие перестановки слов в предложении и пр.) используют уже максимально полный набор стандартных операторов. В анализирующей части составляющие программного элемента динамического компонента взаимодействуют с типом переводных соответствий: так, списки реализуют соответствия эквивалентного типа, схемы лексического анализа – варианты, схемы грамматического анализа – трансформационные соответствия. Оба компонента модели (предметный и динамический) по-разному взаимодействуют в разных режимах работы системы (генерации и трансляции). В обоих режимах используются оба компонента, однако порядок работы их меняется. По словам Марчука Ю.Н. в режиме генерации работа происходит согласно рис. 2, т. е. первым вступает в действие предметный компонент, который обеспечивает создание необходимых для модели составляющих и элементов.

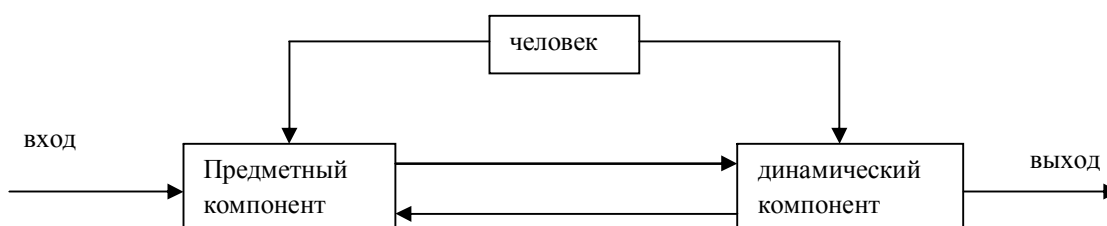


Рис. 2. Генерация системы МП по модели МПС.

Физически объем языковой информации необходимый для создания модели в целом, может быть представлен в виде конкорданса, составленного на материале достаточного количества исходных текстов. После того, как модель с ее двумя компонентами создана, работа системы уже осуществляется по схеме рис. 3.



Рис. 3. Работа модели в режиме трансляции.

Как показывает наше исследование, данная модель представляет весьма гибкую систему и подходит к языкам различного строя. При этом следует учесть, что реализация ее должна проходить по общим законам корпусной лингвистики. Корпусная лингвистика, базирующаяся на больших массивах языковых текстов, позволяет извлекать объективные лингвистические данные и делать содержательные выводы на их основе с высокой степенью достоверности, что гарантирует эффективность работы системы машинного перевода, построенной с учетом этих данных.

Что касается типов современных систем машинного перевода, то согласно классификации, предложенной А. Д. Бакуловым, Н. Н. Леонтьевой (Бакулов и др., 1990), Э. И. Королевым и другими исследователями, выделяются три типа: системы прямого типа; трансферные системы; и системы семантического типа.

Системы прямого типа базировались на принципе перекодирования текста на одном языке в текст на другом языке (прямой перевод). Слово в языке перевода

порождается от слова в языке оригинала. Грамматика в традиционном понимании в них отсутствовала полностью. Предложения на переводящем языке конструировались прямым замещением последовательности слов исходного языка последовательностью слов переводящего языка, следовательно, конечное множество лексических единиц из исходного текста согласовывалось с ограниченным контекстом из лексических единиц текста перевода (Whitelock P., Kilby K., 1995: 6). Системы прямого перевода не обладали средствами решения проблем многозначности, не справлялись с несвязанными языковыми парами, не проводили никакого лингвистического анализа перед генерированием перевода, повторяли синтаксические структуры языка оригинала, не учитывали минимальных потребностей синтаксического и семантического анализа, и не устанавливали различия частей речи, например, существительных и глаголов (Hutchins W.J., 1979: 29). Первые системы перевода, разработанные по принципу «слово за словом», посылающие семантические и синтаксические потребности на задний план, часто выдавали на выход очень низкое качество перевода, что акцентировало сложность языка и необходимость лучшего анализа и синтеза текстов. К системам прямого перевода относятся ранние версии таких программ, как SYSTRAN, LOGOS, SPANAM, TRANSOFT, АМПАР-АНРАП-СПРИНТ, СИЛОД, GERENG-GERRUS и др. (Канападзе, 1989, Королев, 1991, Кулагина, 1990, Хроменков, 2005, ИИ-90, 1990, Toma, 1977, Sinaiko et al., 1972, Vasconcellous, 1984).

Системы трансферного типа отличались от систем прямого тем, что при их проектировании использовалась концепция переводных соответствий, что означало возможность обновления грамматических правил и словарей, а также добавления новых языковых пар без какого-либо негативного воздействия на работоспособность всей системы. Существует два типа перевода по переводным моделям: через язык-посредник и трансфер. Перевод с языком-посредником проходит два этапа: анализ и синтез. На первом этапе текст оригинала трансформируется в представление на языке-посреднике с помощью словаря и грамматических правил исходного языка, а на этапе синтеза предложение на языке перевода порождается из представления на языке-посреднике уже с помощью словарей и грамматических правил на языке перевода (Lewis D., 1992: 78). Основной проблемой такого рода систем является установление достаточно универсального представления, чтобы распространяться на все языки. Но на практике весьма трудно построить представление на действительно естественном языке, которое отображало бы возможные аспекты синтаксиса и семантики всех языков.

Вторым типом перевода по переводным моделям является трансферный перевод, который включает в себя три этапа: анализ, трансфер, и синтез. На этапе анализа текст на исходном языке конвертируется в абстрактное представление также исходного языка. Затем происходит трансфер этого представления в его эквивалент в представлении языка перевода. На третьем этапе генерируется текст на языке перевода. На каждом этапе используются специфические словари и грамматическая информация. Следующая схема иллюстрирует этот процесс:



Трансфер подходит для построения многоязычной системы машинного перевода. В методике трансфера используются различные модели для всех языковых пар. Следует отметить, что на основе систем трансферного типа лежит модель переводных соответствий, описанная подробно в предыдущей части диссертации. Примерами систем с трансфером являются система TAUM (Traduction Automatique de l'Université de Montréal), предназначенная для перевода английских текстов на французский язык, разрабатываемая в Монреальском университете с 1965 г., системы семейства ЭТАП, созданные в ИНФОРМЭЛЕКТРО и Институте проблем передачи информации РАН (Баранов А.Н., 2003: 175-175).

Системы семантического типа появились в 80-ые гг. прошедшего века. Основным отличием данных систем от других является наличие компонента, включающего экстралингвистические знания (ИИ-90б 1990, Королев, 1991, Семенов А.Л., 2008, Хроменков П.Н., 2005, Goodman et al., 1991, Hutchins, 1986, Levin et al., 1992, Nirenburg et al., 1992, Onyshkevich et al., 1995, Rosetta, 1994, Viegas et al., 1999, Wilks, 1975). К этому классу относятся системы семантического типа, использующие модель «смысл↔текст». Разработчики данных систем имели целью установить универсальный смысловой язык-посредник, а когда стала ясна недостижимость этой цели (довольно быстро, впрочем), стремились создать отдельный семантический компонент систем, чтобы он контролировал и переводил семантику как высший уровень языковой структуры в переводе (Марчук Ю.Н., 2010: 119-120). Эти попытки также практически провалились, ни одной действующей системы с семантическим компонентом мы до сих пор не имеем нигде в мире. К системам семантического типа относится целое множество программ, не получивших широкого коммерческого распространения: ASCOF, ATLAS II, DLT, CONTRAST, NICATS/JE, KBMT-89, LUTE, PIVOT, PLAIN, ROSETTA, SEMSYN, TRANSLATOR (Боброва, 1990, ИИ-90, 1990, Капанадзе, 1989, Королев, 1991, Кулагина, 1990, Hutchins, 1986, Lawson, 1991, Levitt, 1993, Maxwell et al., 1988, Piggot, 1989, Tucker et al., 1984).

В третьей главе «Анализ русско-арабских переводов политических текстов (газетный интервью)» рассматриваются некоторые положения корпусной лингвистики, проводится анализ машинного перевода политических текстов (газетных) с русского языка на арабский, выявляются основные семантические трудности и переводческие решения, представляются лингвистические факторы, определяющие правильный перевод, в алгоритмическом описании создания словаря для системы машинного перевода и, наконец, излагаются общие характеристики и основные положения контекстологического словаря.

Проведенный анализ машинного перевода политических текстов (газетных) с русского языка на арабский показал, что система часто сталкивается со следующими ошибками:

1. Многоэквивалентность: когда, в зависимости от контекста, у однозначной исходной единицы имеются разные эквиваленты в ПЯ:

«Его особенность заключается в том, что безопасность обеспечивается через реализацию стратегических национальных приоритетов. В их числе определены оборона, государственная и общественная безопасность.»
"سمتها ضمان السلامة من خلال الاستراتيجية الوطنية ذات الأولوية. ومن بين هذه هي الدفاع والحكومة والسلامة العامة".

С точки зрения семантики, система в принципе исправилась с передачей общего смысла контекста, однако, в данном контексте выбранное машинной арабское слово «سلامة» не является подходящим эквивалентом для русского слова «безопасность». Русскому слову «безопасность» соответствуют два арабских слова «سلامة» и «أمن», выбор которых зависит от контекста. Слово «سلامة», по данным толкового словаря арабского языка «المحيط», обозначает отсутствие боли, вреда, поврежденья, недостатка, порока, например, سلامة الناس (букв. перевод: Безопасность людей), а слово «أمن» - мир, спокойствие, отсутствие чувства страха (антоним слова «страх»), например, «يسود الأمن» (букв. перевод: В стране царит безопасность). Исходя из этого, нам представляется, что для данного контекста подходит слово «أمن», а не «سلامة».

2. Выбор значения, не существующего в рамках исходной единицы, например, интерпретация рус. слова «замысел» как «конструкция»:

«По замыслу, структуре и содержанию Стратегия взаимоувязана с Концепцией долгосрочного социально-экономического развития России до 2020 года».

"حسب التصميم، هيكل ومحتوى الاستراتيجية قترتبط بمفهوم التنمية الاجتماعية الاقتصادية الطويلة الأجل لروسيا حتى عام 2020".

В данном переводе наблюдается неудачная передача значения, стоящего за словом «замысел». В толковом словаре русского языка слово «замысел» имеет два значения: 1) задуманный план действий, деятельности; 2) смысл, идея. Однако система bing далеко ушла от этих двух значений, которые в данном контексте очень подходят для перевода на арабский язык, и вместе них привела араб. слово «تصميم», соответствующее русскому слову «конструкция». В данном случае представляется правильным решением выбрать араб. слово «غاية», выражающее первое значение, как более подходящий эквивалент для рус. слова «замысел».

3. Игнорирование некоторых семантических признаков отдельных слов, например, глагол «взаимоувязать» в предыдущем переводе. Первая часть сложного слова (взаимо) добавляет дополнительное семантическое значение (взаимное действие с обеих сторон, проявляющееся по отношению друг к другу). Система опустила данный семантический признак и ограничилась араб. глаголом «ارتبط» (увязать). Для решения данной проблемы необходимо прибегать к словосочетанию «ارتبط ارتباطاً متبادلاً», буквальным переводом которого является «взаимно увязать».

4. Опускание некоторых исходных единиц. Таких ошибок много.

5. Игнорирование грамматических признаков слова при определении переводного соответствия. Например, при переводе языковой единицы «соответственно»:

«Военная доктрина должна быть непосредственной составной частью единого механизма обеспечения национальной безопасности, соответственно, изменения в системе касаются и ее составных частей».

"العقيدة العسكرية ينبغي أن تدمج مباشرة إليه واحدة لضمان الأمن الوطني، على التوالي، المتعلقة بالتغييرات في النظام والأجزاء المكونة لها".

Языковая единица «соответственно» выполняет разные морфологические функции. В предложении она иногда выступает наречием, предлогом или союзом. В зависимости от ее морфологической функции ее значение может меняться, следовательно, иметь разные эквиваленты в арабском языке. Выступая в данном контексте союзом, слово «соответственно» соответствуют араб. слову «لذا», а не «على التوالي» как предлагает система bing. Арабской языковой единице «على التوالي» соответствует русское слово «соответственно», выступающее в предложении наречием.

6. Многозначность: когда у одного исходного слова имеется несколько значений, следовательно, разные эквиваленты в ПЯ. Ввиду того, что система часто сбивается из контекста, она часто выбирает несоответствующее контексту значение, соответственно, она находит неправильный эквивалент в ПЯ.

7. Проблема антонимического перевода, например, «...угрозы для нашей страны не сняты». Суть данной проблемы заключается в трансформации отрицательной конструкции в утвердительную или наоборот, утвердительной в отрицательную, сопровождаемую заменой одного из слов переводимого предложения ИЯ на его антоним в ПЯ. Данное явление в переводческой литературе известно под названием «антонимический перевод» (Бархударов Л. С., 2010: 215). Так, русскую отрицательную конструкцию «...военные опасности и возможные военные угрозы для нашей страны не сняты» нужно было преобразовать в арабскую утвердительную, а глагол «снять» в данном контексте заменить его арабским антонимом «قائم». Система МП сохранила русскую отрицательную конструкцию при переводе и не заменила исходное слово «снять» на его антоним в арабском языке, что привело к невозможности воспринять высказывания.

8. Неполная расшифровка аббревиатуры, например, «МАГАТЭ»:

«Мы, конечно, считаем, что Иран должен сотрудничать с МАГАТЭ в большей степени».

"ونعتقد بالتأكد أن تتعاون إيران مع الوكالة الدولية".

В арабском языке аббревиатуры, указывающей на название какого-либо объекта, не существует. Система в нашем случае распознала аббревиатуру. Однако, она расшифровала только первую и вторую части аббревиатуры «МАГАТЭ» (международное агентство - الوكالة الدولية). Для решения подобных проблем необходимо снабжать систему всеми популярными аббревиатурами.

9. Омонимичность аббревиатуры: например «РФ» - Российская Федерация, и «РФ» - РФ-лифтинг. Словосочетание «Совет безопасности РФ» система переводит как «Совет безопасности РФ-лифтинг», а не «Совет безопасности Российской Федерации». РФ-лифтинг – это технический термин, который широко употребляется в науке о физике.

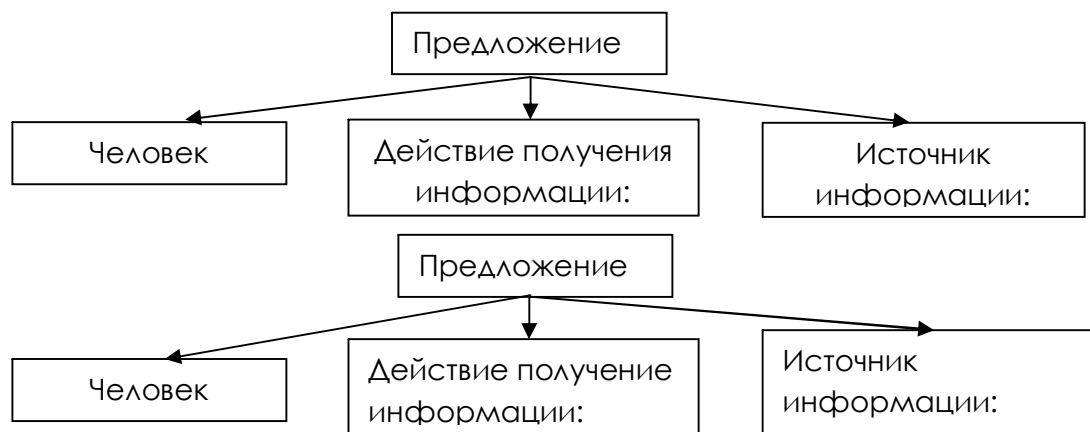
10. Перевод устойчивых словосочетаний, например, «оказывать влияние». Данное фразеологическое явление система упустила из виду, следовательно, опустила сказуемое в предложении, что привело к невозможности восприятия высказывания. Данное устойчивое глагольно-именное словосочетание можно было сохранить в переводе на арабский язык «أبدى تأثيراً» или, по крайней мере, ограничиться глаголом *влиять* «أثر».

Для решения подобных проблем в системах машинного перевода необходимо улучшить электронный словарь с учетом всех необходимых внутрилингвистических факторов, детерминант (окружающих слов, словоформы, грамматических признаков, синтаксического и/или семантического контекста, грамматических отношений между единицами и т.п.), и нужной информации для преодоления лингвистического барьера перед системой машинного перевода. Кроме семантического анализатора необходимо также улучшить морфологический и синтаксический анализаторы, так как во многих случаях морфологические и синтаксические признаки языковой единиц, как показывает анализ, эффективно влияют на правильный выбор значения многозначной языковой единицы. Эффективность систем машинного перевода во многом зависит от структурной организации и механизма работы словарей. В целях преодоления дефектов классических словарей при создании современных электронных словарей необходимо учесть смысл каждого слова и его свойств, решить все лингвистические вопросы (двусмысленность слов, семантическая неясность предложений и текстов, и др.), которые встречаются при автоматической обработке естественных языков. Другими словами, создатели словарей должны уделить особое внимание таким лингвистическим явлениям, как синонимы, антонимы и омонимы, ибо они имеют большое значение для устранения семантической неясности, и решения проблем машинного лингвистического анализа.

Механизм построения автоматического словаря должен учесть следующие факты:

1. Набор словарных лексических единиц по контекстам: на основе собрания словарного материала лежит анализ текстов со всех сторон: тематической, стилистической и др. Данная процедура представляет возможность ограничить значения слова по контекстному употреблению и стилистической окраске. Больше употребляющееся значение в контекстах, зафиксировано как основное, а редко употребляемое - как второстепенное или исключительное (البعلبكي, 1987: 299-285).
2. Четко и иерархично разграничить и классифицировать семантические поля слов с главного общего «верхнего» класса до самого узкого подкласса.
3. Определить точку пересечения значений разных слов в пределах одного семантического поля и зафиксировать объем занимаемых им мест в данном поле.
4. Определить семантическое поле каждой группы синонимичных слов (например, группа глаголов *умереть*, *скончаться*, *задушиться*, *погибать* и т.д. можно объединить в одно семантическое поле «смерть»). Затем определить семантические признаки каждого слова данной группы, например, *убить* обозначает «смерть» в искусственном виде с возможностью употребления «орудия»; *умереть* обозначает «смерть» в нормальном виде (Сальва С. Х., Омар М., 2006: 11-34).
5. Создать специальную таблицу для исходных однозначных единиц, имеющих в ПЯ несколько эквивалентов, выбор которых зависит от контекста.
6. Снабжать систему определенным количеством контекстов для каждой многозначной языковой единицы с определением грамматических отношений между другими составляющими предложение. Данный набор контекстов служит эталоном для подбора более вероятного значения.

7. Снабжать систему популярными аббревиатурами.
8. За фоном каждого предложения стоит определенная семантическая микро-сеть, которая в свою очередь входит в состав семантической сети абзаца, затем целого текста. Данные семантические микро-сети необходимы для сравнения с другими семантическими микро-сетями, которые могут встретиться в новых контекстах. Так, следующих два русских предложения на определенном уровне отличаются между собой, соответственно, имеют разные арабские эквиваленты: 1) *Иван читает книгу*; 2) *Иван читает Пушкина*. Так, данные предложения, как нам представляется, имеют следующие семантические микро-сети:



Когда в тексте встретится одна из этих семантических микро-сетей система автоматически предлагает соответствующий арабский эквивалент, например, для первого предложения подбирается арабский эквивалент «يقرأ إيفان كتاباً» (букв. перевод: Читает Иван книгу), а для второго – «يقرأ إيفان لبوشكن» (букв. перевод: Читает Иван Пушкину). Современная система должна быть снабжена подобными микро-сетями для разрешения таких потенциальных проблем.

9. Создать в базе данных таблицу, специально предназначенную для организации микроконтекстуальных отношений некоторых многозначных или омонимичных слов с другими словами, и определить при какой синтаксической связи наблюдается данный тип отношений. Например, слово «коса» занимает место в разных семантических полях (орудие, прическа), и для выбора верного значения данного слова часто требуется знание о его синтаксической позиции в микроконтексте и употребляемом с ним глаголе. Синтаксическая позиция (подлежащее, сказуемое, дополнение, определение, обстоятельство) составляющих микроконтекст будем представлять кодами в двоичной системе счисления как показывает следующий рисунок:

подлежащее	сказуемое	дополнение	обстоятельство	определение
0	0	0	0	0

10000 = 1 = подлежащее.
 01000 = 2 = сказуемое.
 00100 = 4 = дополнение.
 00010 = 8 = обстоятельство.
 00001 = 16 = определение.

В качестве примера приведем следующие выражения и попробуем решить проблему нахождения правильного перевода для слов «коса»:

1. ...Но, вместо ответа, человек скрутил руки девушки и **привязал косой** к мачте, чтобы не бросилась в воду («Легенды Крыма»).
2. Он прожил не больше года, напакостил что-то в деревне, садовник хотел его **убить косой**, отец мой велел ему убираться (А.И. Герцен. Былое и думы).
3. Подскажите, пожалуйста, как **косить траву косой** на сено? (Пример автора).
4. **Косу заплетали** незамысловатее, на ленточках старались навязать побольше узлов, чтобы крёстная жениха не могла **расплести косу** самостоятельно, а просила бы девок ей помочь («Национальный корпус русского языка»).
5. Один дядя играл в гольф, как вдруг появилась тетка с косой и стала **косой** его **душить**. Охрана отогнала тетку (Интернет-сайт <http://www.liveinternet.ru/>).

Анализируя данные примеры, мы обнаружили, что слово *коса* в употреблении как дополнение [00100 (4)] с глаголами *привязать*, *душить*, и *[рас]плести/[за]плести* всегда представляется значением, сохраненным в семантическом поле «прическа», а с глаголами *косить/убить* — «орудие». Данные отношения мы можем представить следующими таблицами базы данных:

Таблица семантических полей

id сем поле	семантическое поле
0000000001	Прическа
0000000002	Орудие
0000000003	Инструмент
0000000004	Пространство
0000000005	Источник информации

Таблица словарных единиц

id слова	словарная единица
0000000001	Коса
0000000002	Плести
0000000003	Косить
0000000004	Душить
0000000005	Убить
0000000006	Привязать

Таблица значений

id значение	id слова	id сем поле	Значение
0000000001	0000000001	0000000001	ضفيرة
0000000002	0000000001	0000000002	منجل
-----	-----	-----	-----

Таблица контекстных отношений

id слова1	синтак позиция_1	id слова2	синтак позиция_2	id значение
0000000001	4	0000000002	2	0000000001
0000000001	4	0000000003	2	0000000002
0000000001	4	0000000004	2	0000000001
0000000001	4	0000000005	2	0000000002
0000000001	4	0000000006	2	0000000001

Таблица «контекстные отношения» играет главную роль в определении подходящего значения для определенного контекста. Так, первое поле «id_слова1» указывает на идентификационный номер определенного слова, которое может встретиться в каком-либо контексте. Например, идентификационный номер 0000000001 согласно таблице «словарные единицы» ссылается на слово «коса». Если данное слово занимает в определенном контексте синтаксическую позицию, обозначенную цифрой 4 (00100 - дополнение), и находится в сочетании со словом под идент. номером 0000000002 (*плести*), занимающим в свою очередь синтаксическую позицию под номером 2 (01000 - сказуемое), то арабским соответствием в данной ситуации является значение под идент. номером 0000000001, который согласно таблице «значения» ссылается на арабскую единицу «ظفيرة». Таблицу «контекстные отношения» можно снабдить другой необходимой информацией для разрешения подобных проблем.

Следует также отметить, что решение проблемы лексической многозначности известный ученый-лингвист Ю.Н. Марчук видит в точном учете специфических особенностей предметного поля и лингвистического состава конкретных подязыков (т.е. областей, заведомо существенно меньших, чем вся система естественного языка). Для таких языковых общностей как подязыки возможно определять значение отдельных языковых единиц таким образом, что их совокупность (линейная комбинация) не противоречит идее цельного текста как такового. На этой теоретической основе была впервые выдвинута Ю.Н. Марчуком в 1976 г. идея создания контекстологических словарей для определенных типов текстов в рамках определенных семантических полей и подязыков, которые сегодня помимо контекстов употребления включают и толкование. Контекстологические словари дают возможность создавать многоязычные терминологические базы данных для широких предметных областей. Подобные проблемы, связанные с контекстом и разрешением многозначности, также подробно рассматриваются в трудах профессора Марчука Ю.Н. (Контекстологический словарь для машинного перевода многозначных слов с английского языка на русский. М., 1976; проблемы машинного перевода. М., 1983; Методы моделирования перевода. - М, 1985; Компьютерная лингвистика. - М., 2007).

В **Заключении** подводятся итоги диссертационного исследования и формулируются основные выводы.

В **Библиографии** приводятся труды российских, арабских и других зарубежных ученых, словари и другие источники, используемые в исследовании.

Основное содержание диссертации отражено в следующих публикациях:

1. **Машинный перевод и семантическая обработка текстов // международный аспирантский вестник. Русский язык за рубежом. –М., 2011. -№1.-С. 9-12.**
2. **Основные трудности и переводческие решения в машинном переводе политического газетного текста с русского языка на арабский // Мир науки, культуры, образования. –Горно-Алтайск, 2011. –№2(27). –С. 41-46.**
3. Теория машинного перевода // Слово. Грамматика. Речь. –М. 2010. –№7. –С. 126-137.
4. Состояние современных систем машинного перевода с русского языка на арабский // Язык, сознание, коммуникация. –М. 2011. -№42. –С. 69-75.

5. Теории «смысл↔текст» и «текст-текст» // Актуальные проблемы в гуманитарных и естественных наук. –М. 2011. -№1(24). –С. 169-175.
6. Альотаиби С.М. Проблемы преобразования фоновых культурных знаний в системах машинного перевода. ? Сборник Материалов IV Международной научно-практической конференции «Русский язык как иностранный в современной образовательной и геополитической парадигме». – М.: Макс Пресс, 2010. С. 224-225.
7. Альотаиби С.М. Язык и культура в проблематике современного машинного перевода. ? Сборник материалов VII Международной научной конференции «Синхронное и диахронное в сравнительно-историческом языкознании». – М.: КДУ, 2011. С. 10-13.
8. Альотаиби С.М. Теория уровней эквивалентности. ? Материалы III Международной интернет – конференции «Русский язык, Литература, Культура: актуальные проблемы изучения и преподавания в России и за рубежом». – М., 2010. (электронный носитель - диск).