Нижегородский государственный

лингвистический университет

им. Н.А. Добролюбова

**дипломная работа**

**на тему**

**Электронные словари**

**и их применимость**

**для традиционного**

**машинного перевода**

*Выполнил:*

студент гр. 503А Тараскин А.А.

*Научный руководитель:*

к.ф.н. доц. Яржемский А.Н.

**Нижний Новгород, 2003 г.**

**Содержание:**

**Введение** 3

**Глава I. Место лексикографии среди   
лингвистических дисциплин.** 5

* 1. Лексикография как одна из наук о языке. 6
  2. Типы словарей. 8
  3. Составление словарей. 22

**Глава II. Электронные словари.** 24

1. Электронные словари "МультиЛекс". 24
2. Электронные словари "Lingvo". 25
3. Преимущества электронных словарей. 26
4. Новые мехи требуют нового вина. 29
5. Читатели и писатели. 30
6. Проблема актуальности. 31
7. Соответствие уровню достижений   
   лингвистической науки. 31

**Глава III. Машинный перевод.** 33

1. Немного истории. 33
2. Классификация систем машинного перевода   
   по Лари Чайлду. 34
   * 1. *Полностью автоматизированный МП.* 34
     2. *Автоматизированный МП   
        при участии человека.* 35
     3. *Перевод, осуществляемый человеком   
        с помощью компьютера.* 35
3. Системы машинного перевода. 37
   * 1. *Система МП Retrans Vista.* 40
     2. *Система МП PROMT XT.* 43

**Глава IV. Практическая часть.** 49

**Заключение** 54

**Библиография** 56

**Введение.**

В настоящее время компьютеры занимают все более значительное место не только среди программистов и инженеров, но и в среде самых разнообразных пользователей, включая лингвистов, переводчиков и специалистов, нуждающихся в оперативном переводе иноязычной информации. В этой связи компьютерные словари являются очень удобным подручным средством в целях экономии времени и оптимизации процесса понимания иноязычной информации. Кроме того, сейчас имеются программы-переводчики, которые могут производить более или менее адекватный перевод иноязычных текстов и могут являться подспорьем в работе специалистов различных профилей. Данная работа посвящается изучению указанных проблем, а также анализу некоторых лингвистических программных продуктов, направленных на автоматизацию процесса перевода.

Настоящую тему исследования можно считать вполне современной, поскольку история развития и внедрения в повседневную жизнь персональных компьютеров (а тем более таких, каким было бы 'под силу' осуществлять более или менее современные программы машинного перевода) насчитывает едва ли более пятнадцати лет. Особую актуальность эта тема приобретает, если учесть тот факт, что именно в настоящее время Россия все более интегрируется в международное сообщество и что, наряду с экономическими и политическими барьерами, препятствуют этому во многом барьеры языковые. Вместе с тем, профессиональных переводчиков, способных и желающих осуществлять подобный процесс коммуникации сообществ во всех сферах науки и культуры, не так много, следствием чего является тот факт, что услуги их недешевы. Поэтому именно сейчас особенно актуален поиск путей максимально автоматизировать процесс перевода, осуществляемого человеком, чтобы, с одной стороны, максимально облегчить нелегкий труд человека-переводчика, а с другой - сделать этот труд максимально эффективным. Осуществить подобное можно, лишь максимально интегрировав усилия специалистов в областях кибернетики, программирования, психологии, а главное - лингвистики.

Цель данной работы - определение того, насколько можно использовать современные программные продукты для осуществления перевода, а также выявление наиболее перспективных, на наш взгляд, направлений исследований в области его автоматизации.

В соответствии с поставленной целью, **задачами** исследования являются:

* уяснение местоположения компьютерной лексикографии среди лингвистических наук;
* уяснение различий между электронными и традиционными бумажными словарями;
* определение максимально эффективного способа организации электронных словарей;
* уяснение различий между разнообразными системами МП и классификации последних;
* анализ работы систем машинного перевода на примере перевода, произведенного программой PROMT XT, и сравнение образцов перевода с переводом, сделанным человеком. Анализ причин несоответствия.

**Глава I. МЕСТО ЛЕКСИКОГРАФИИ СРЕДИ   
ЛИНГВИСТИЧЕСКИХ ДИСЦИПЛИН**

Слово «лексикография» греческого происхождения, lexikos — относящийся к слову, словарный и grapho — пишу. Следовательно, лексикография означает: «пишу сло­ва» или «пишу словари». В современном значении лексико­графия — это теория и практика составления словарей, главным образом языковых, лингвистических, в отличие от неязыковых, энциклопедических.

Лексикография как научный термин появилась в широ­ком обиходе сравнительно недавно. Например, в энциклопедическом словаре Брокгауза и Ефрона (1896 г.) нет статьи на слово «лексикография», однако есть статья на слово «лексикология». Справедливости ради следует отме­тить, что в статье «словарь» этого же справочника имеется слово «лексикография», где оно является синонимом сло­восочетания «словарная техника».

В энциклопедическом словаре братьев А. и И. Гра­нат (1916 г.) уже есть статья на слово «лексикография», которая определяется как «научные способы обработки словесного материала языка для составления лексикона». Отметим в этом определении акцент на «научные способы обработки».

В первом издании «Большой советской энциклопедии» (1938 г.) в статье на слово «лексикография» дано: «Лексикография *(греч.),* работа по составлению словарей». И только во втором (1953 г.) и третьем (1973 г.) изданиях этот термин толкуется достаточно современ­но: «Лексикография — раздел языкознания, занимающийся практикой и теорией составления словарей». (БСЭ, 3-е изд. 1973, т. 14).

С другой стороны, статьи на слово «лексикография» нет ни в энциклопедии «Британика», ни в энциклопедии «Аме­рикана», хотя в обеих есть статьи на слово «словарь». Отсутствие термина «лексикография» в таких солидных совре­менных справочниках, как британская и американская энциклопедии, отнюдь не случайно. Это объясняется, во-первых, молодостью лексикографии как науки и, во-вто­рых, тем, что даже среди самих языковедов до сих пор еще идет спор, является ли лексикография наукой, точнее, частью науки о языке, или же она просто техника со­ставления словарей, в лучшем случае, искусство их со­ставления.

Видный испанский лексикограф X. Касарес в своей широко известной книге «Введение в современную лексико­графию» (которая переведена на русский язык) утверждает, что лексикография — это техника и искусство составления словарей. (Вспомним определение «лексикографии» в пер­вом издании БСЭ: «работа по составлению словарей».)

Известный английский лексикограф, автор знаменитых справочников по английскому и американскому сленгу Эрик Патридж, посвятивший всю свою жизнь составлению сло­варей, не случайно назвал свою последнюю книгу, плод многолетних разысканий в области словарного дела, так: «Благородное искусство лексикографии как объект занятий и опыт страстного его приверженца».

Даже Ф. Гоув, главный редактор третьего издания сло­вари Уэбстера (1461 г.), крупнейшего лексикографического предприятия нашего времени, пишет в программной статье «Успехи лингвистики и лексикография»: «Лексикография еще не наука. По-видимому, они никогда не будет наукой. Однако это сложное, топкое и порой всепоглощающее искусство, требующее субъективного анализа, произвольных решений и интуитивных доказательств".

**Лексикография как одна из наук о языке**

Существует, однако, другая точка зрения на лексикогра­фию. Ее сторонники считают, что лексикография — это не просто техника, не просто практическая деятельность по составлению словарей и даже не искусство, а само­стоятельная научная дисциплина, имеющая свой предмет изучения (словари различных типов), свои научные и ме­тодологические принципы, свою собственную теоретическую проблематику, свое место в ряду других наук о языке.

Впервые эту точку зрения на лексикографию со всей определенностью высказал известный советский языковед академик Л. В. Щерба. В предисловии к русско-фран­цузскому словарю (1936 г.) он писал: «Я считаю крайне неправильным то пренебрежительное отношение наших квалифицированных лингвистов к словарной работе, благо­даря которому почти никто из них никогда ею не зани­мался (в старые времена это за гроши делали случайные любители, не имевшие решительно никакой специальной подготовки) и благодаря которому она получила такое нелепое название «составление» словарей. И действи­тельно, наши лингвисты, а тем более наши «составители» словарей просмотрели, что работа эта должна иметь науч­ный характер и никак не состоять в механическом со­поставлении каких-то готовых элементов».

Развивая выдвинутые в 1936 г. положения, Л.В. Щерба публикует в 1940 г. статью (ставшую потом широко известной и за рубежом), в которой на большом фактичес­ком материале начинает разрабатывать основные теорети­ческие вопросы лексикографии. Л.В. Щерба мыслил написать несколько статей (этюдов, как он их называл) по общей теории лексикографии, в которых намеревался обсу­дить такие важные проблемы, как основные типы словарей, природа слова, значение и употребление слова, построение словарной статьи в связи с семантическим, грамматическим и стилистическим анализом слова и др. Однако преждевременная смерть помешала осуществлению этого замысла. Л.В. Щербой был написан лишь первый этюд «Основные типы словарей», который начинался следующим хорошо известным сейчас положением: «Одним из первых вопросов лексикографии является, конечно, вопрос о различных ти­пах словарей. В основе его лежит ряд теоретическихпротивоположений, которые и необходимо вскрыть». С тех пор тезис о том, что лексикография — это не только практика составления словарей, но и теоретическая научная дисциплина, прочно вошел в ряд отправных положений советской лексикографической школы.

Здесь, однако, можно задать вопрос: такое ли большое точение имеет антиномия «наука или искусство» для лексикографии? Ведь ясно, что в том и другом случае составлением словарей следует заниматься, ибо они нужны; нужно много хороших и разных словарей. На это со всей определенностью следует ответить, что данный вопрос имеет принципиальное значение, и вот почему.

Что такое наука вообще? Каковы ее важнейшие черты? Основными и важнейшими чертами науки, любой научной дисциплины являются следующие: наличие системы знаний и необходимость их объективного исследования. Эти две существенные черты взаимосвязаны, тесно переплетены, ибо только тогда можно строить адекватную действительности систему знаний, когда эта действительность будет объек­тивно изучена. В применении к лексикографии это выглядит следующим образом.

Если принять тезис о том, что лексикография есть ис­кусство, то тогда открываются двери субъективному пониманию задач и пред­мета лексикографии, приемов и методов ее исследования, субъективному решению ее проблем. Вряд ли такой подход будет плодотворен и уж никак не научно объективен. Если же принять тезис о том, что лексикография есть просто техника составления словарей, некий вид сугубо практической деятельности, то тогда придется передать другим наукам (лексикологии, семантике, стилистике, эти­мологии и т. п.) решение всех теоретических вопросов, и лексикография должна будет использовать лишь готовые решения этих наук. Вряд ли это будет плодотворно, ибо другие науки о языке недостаточно хорошо знакомы с по­ложением дел в лексикографии. Они, следовательно, будут решать вопросы лексикографии со своих позиций, со своих точек зрения, а потому ущербно для лексикографии. Та­ким образом, тезис о том, что лексикография есть наука, является единственно правильным и наиболее плодотвор­ным. Из этого следует, что лексикография как наука имеет свой собственный предмет исследования, свои специальные методы исследования, свою структуру, свое место в ряду других лингвистических дисциплин.

Как и всякая наука, лексикография имеет две стороны: научно-теоретическую и практически-прикладную. Первая (теоретическая лексикография) ставит общие теоретиче­ские проблемы и работает над их решением. Вторая (прак­тическая лексикография) занимается непосредственно со­ставлением словарей различных типов на базе теоретичес­ких решений основных проблем. Конечно, деление лексикографии на две части в высшей степени условно. Эти две стороны лексикографии идут всегда вместе, они взаимосвязаны: лексикограф-теоретик не может заниматься голым теоретизированием без работы над конкретным материалом, без участия в какой-то практической лекси­кографической работе; и, наоборот, никакой лексикограф-практик не может окунуться только в свою чисто эмпи­рическую работу, не зная новейшей проблематики лекси­кографии как науки. Тем не менее, принципиальное раз­граничение двух сторон лексикографии чрезвычайно важно.

Из изложенного выше можно заключить, что термин «лексикография» имеет в настоящее время три значения: 1) наука, точнее, особая область языкознания, изучающая принципы составления словарей разных типов; 2) сама практика словарного дела, т.е. составление словарей; 3) совокупность словарей данного языка.

Являясь частью науки о языке, лексикография тесным образом связана с такими лингвистическими дисциплинами, как лексикология, семантика, стилистика, этимология, фонология и т. п. С этими дисциплинами лексикография имеет общую проблематику. Порой она использует резуль­таты их исследований, а часто и опережает их в решении некоторых проблем.

Таким образом, постепенно, шаг за шагом, лексикография оформляется и самостоятельную лингвистическую дисциплину, становясь равноправной среди других языковедческих наук.

**Типы словарей.**

Существующие типы словарей разнообразны. Это разнообразие объясняется, прежде все­го, сложностью и многоаспектностью самого объекта лексикографического описания, т. е. языка. Кроме того, многочисленные потребности общества в получении самой различной информации о языке также усложняют и рас­ширяют репертуар словарей. Практически нет никакой возможности дать в одном словаре всю в той или иной мере исчерпывающую информацию о языке, которая удовлетворила бы в равной степени все общество в целом и отдельные его слои и частности. Именно поэтому в лю­бой национальной лексикографии мы находим десятки, ато и сотни словарей самых разных типов.

Деление словарей на типы происходит, как говорят классификаторы, по различным основаниям: в зависимости от цели словаря, его объема, порядка расположения в нем слов, объекта описания и т. д. Многие из указанных момен­тов налагаются друг на друга, объединяясь в словаре одного и того же типа, другие стоят особняком, служа основанием для словарей совершенно отличного типа. Существуют переводные, толковые, диалектные и региональные словари, словари сленга, исторические, неологизмов, этимологические, крылатых выражений и много других. Необходимо от­метить, что в науке о языке пока еще нет сколько-нибудь общепринятой типологии словарей, хотя попытки создать таковую предпринимались многими лингвистами, в частно­сти Л. В. Щербой, П. Н. Денисовым, Б. Кемадой, Я. Малкилом, Л. Згустой и др.

Прежде всего, приходится различать словари ***лингвистические*** и ***нелингвистические***. Первые собирают и описывают под тем или иным углом зрения лексические единицы языка (слова и фразеологизмы). Особый подтип лингвистических словарей составляют так называемые ***идеографические*** словари, идущие от понятия (идеи) к выражению этого понятия в слове или словосочета­нии. В нелингвистических словарях лексические единицы (в частно­сти — термины, однословные и составные, и собственные имена) служат лишь отправной точкой для сообщения тех или иных сведений о предметах и явлениях внеязыковой действительности. Встречаются и промежуточные разновидности словарей. Кроме того, всякий словарь может быть отнесен либо к «***общим***», либо к «***специальным***».

Примерами общих лингвистических словарей могут служить обыч­ные ***толковые*** и ***переводные*** словари, охватывающие с той или иной степенью полноты всю лексику, бытующую в общенародном употребле­нии. Специальный лингвистический словарь разрабатывает какую-то одну область лексики, иногда достаточно широкую (например, ***фразео­логический*** словарь, словарь иностранных ***слов***), иногда же довольно узкую (например, словарь ***личных имен***, даваемых новорожденным). ***Общий нелингвистический словарь*** — это общая ***энцикло­педия*** (например, БСЭ—Большая Советская Энциклопедия). Спе­циальный нелингвистический словарь — это ***специальная (отрасле­вая) энциклопедия*** (медицинская, юридическая и т. д.) или же краткий словарь той или иной (обычно — более узкой) области знания, или биографический словарь деятелей той или иной отрасли (писателей, художников и т.д.), или же той или иной страны (словарь-справоч­ник типа «Who is who»).

**Толковые словари.** Толковым словарем называют такой словарь, главной задачей которого является толкование значений слов (и фразеологизмов) ка­кого-либо языка средствами самого этого языка. Толкование дается с помощью логического определения концептуального значения (на­пример, *накалиться —* нагреться до очень высокой температуры; *рекордсмен —* спортсмен, установивший рекорд), посредством под­бора синонимов *(назойливый —* надоедливый, навязчивый) или в форме указания на грамматическое отношение к другому слову *(прикрывание—*действие по значению глаголов *прикрывать* и *прикрываться).* В некоторых толковых словарях значения слов раскрываются в нуж­ных случаях с помощью рисунков. Эмоциональные, экспрессивные и стилистические коннотации указываются посредством специальных помет («неодобр.», «презр.», «шутл.», «ирон.», «книжн.», «разг.» и т. п.). Отдельные значения по мере надобности и возможности (в за­висимости от объема словаря) иллюстрируются примерами — типич­ными сочетаниями, в которых участвует данное слово (например, *утюг накалился, атмосфера накалилась —* где глагол выступает уже в переносном значении: “стала напряженной”), или же (особенно в словарях большего объема) цитатами авторитетных писателей. Как правило, толковые словари дают также грамматическую характеристику слова, указывая с помощью специальных помет на часть речи, грам­матический род имени существительного, вид глагола и т. д. и приводя в нужных случаях кроме «представительной», или «словарной», и некоторые другие грамматические формы данного слова. В той или иной мере указывается и произношение слова (например, в русских толковых словарях — ударение), иногда сообщаются и разные другие, добавочные сведения.

Обычно толковые словари являются словарями современного лите­ратурного языка. Некоторые из них носят строго нормативный характер, т. е. отбирают только факты, полностью соответствующие литературной норме, рекомендуют эти факты как единственно «правильные» и отсекают все, что хотя бы немного отклоняется в сторону просторечья. Типичным примером может служить академический сло­варь французского языка (Dictionnaire de I'Academie Française). Для многих других толковых словарей характерно более широкое понимание литературного языка и, соответственно, включение в сло­варь разговорной и даже просторечной лексики (кроме лишь узко-областных, диалектных, узкопрофессиональных и сугубо арготических элементов). К этому типу относятся академические словари русского языка — 17-томный «Словарь современного русского литературного языка» Академии наук СССР (1950—1965) и 4-томный «Словарь русского языка» (1957—1961), а также однотомный «Словарь русского языка» С. И. Ожегова (9-е испр. и доп. издание под ред. Н. Ю. Шведовой 1972 г.), который очень полезен для практических целей, и более ранний «Толковый словарь русского языка» коллектива авторов под ред. Д. Н. Ушакова (4 тома, 1935—1940). Особое значение для русской лексикографии имеет, конечно, 17-томный академический «Словарь современного русского литературного языка». Он включает более 120 тыс. слов. В 1970 г. он отмечен Ленинской премией.

Иной характер носит знаменитый, не раз переиздававшийся «Тол­ковый словарь живого великорусского языка» В. И. Даля (4 тома, первое изд. 1863—1866 гг.), в изобилии включающий областную и диа­лектную лексику середины 19 века и в отношении полноты охвата этой лексики и обилия народных выражений до сих пор непревзойденный. Он включает около 200 тыс. слов литературного языка и диалектов. С 1965 г. начал выходить «Словарь русских народных говоров» под редакцией Ф. П. Филина, в котором представлены диалектная лексика и фразеология всех русских говоров 19-20 веков.

Главная задача толкового словаря – истолковать значение слов и их применение в речи, отграничить правильное от неправильного, показать связь слов со стилями языка, дать читателю сведения об особенностях падежных, родовых, залоговых, видовых и иных грамматических форм слова; попутно указывается, как слова пишутся и произносятся.

Толковые словари, как правило (но не всегда), оказываются и ***нормативными***, т.е. объясняющими слова в соответствии с требованиями литературно-языковых норм (норма применительно к языку – это выработанное при участии литературы и принятое обществом в качестве обязательного правило, регулирующее применение слова в речи, его написание, произношение и ударение). Так, все перечисленные толковые словари русского языка – нормативные, за исключением «Толкового словаря живого великорусского языка» В.И. Даля.

**Переводные словари**. Толковым словарям противостоят переводные, чаще всего двуязычные (скажем, русско-английский и англо-русский), а иногда многоязычные. В переводных словарях вместо толкования значений на том же языке даются переводы этих значений на другой язык, например, *накалиться* - become heated, *назойливый* – importunate, troublesome. В зависимости от того, предназначен ли словарь как пособие при чтении (слушании) текста на чужом языке, или как пособие при переводе с родного языка на чужой, его желательно строить по-разному. Так, русско-английский словарь для англичан может да­вать меньше сведений в «правой» (т. е. английской) части, чем их дает русско-английский словарь, предназначенный для русских. Например, переводя русское *обращение,* словарь для англичан может просто перечислить все возможные английские эквиваленты *(address, appeal; conversion; treatment, circulation* и т. д.), так как англичанину известны смысловые различия между этими английскими словами; в словаре же для русских придется указать, что *address* и *appeal* это ‘обращение к ...’, причем *appeal* это ‘обращение’ в смысле’призыв’; что *conversion* это ‘обращение в веру’ и т. п., что *treatment* это ‘обращение с...’, ‘обхождение с кем-либо’, a *circulation* ‘обращение товаров, денег и т. п.’; кроме того, придется указать, с какими предлогами употребляются эти ан­глийские существительные, даже указать место ударения, т. е. снабдить английские эквиваленты многими разъясне­ниями, которые помогут правильно употребить их, переводя текст со словом *обращение* с родного русского языка на иностранный англий­ский. Ясно, что в англо-русском словаре картина соответственно из­менится. В словаре, рассчитанном на русских, русская часть будет менее подробной, но в словаре, предназначенном для англичан, при­дется подробно указывать различия в значениях и в употреблении русских эквивалентов, снабжать их грамматическими пометами, указывать ударение и т. д. Хороший переводный словарь должен содержать также стилистические пометы и особо отмечать случаи, когда переводящий эквивалент является неточным в стилистическом отно­шении. Перевод слов всегда представляет большую трудность, т.к. объем значения слова в разных языках часто не совпадает, переносные значения в каждом языке развивается по-своему. Так, в русском *сон* означает и «сон» (состояние сна) и «сновидение», а в чешском первому соответствует *spanek*, а второму – *sen*, аналогично в английском различают *sleep* и *dream*, *slumber*; в немецком *Schlaf* и *Traum*. Наоборот, важное для русского языка различие глаголов *идти* и *ехать* не получит отражения в переводе на болгарский язык, где будет общий глагол *ида*, *идвам*, и французский, где *arriver* – и *идти*, и *ехать* и т.п.

Переводные словари могут быть ***двуязычными*** (русско-французский, англо-русский и т.п.) и ***многоязычными***. К последним относится составленный А. и В. Поповыми «Словарь на семи языках (французско-немецко-английско-итальянско-испанско-португальско-голладско-русский)», вышедший в свет в 1902 г. Теоретическое и практическое значение подобных словарей весьма невелико. Значительно важнее многоязычные специальные словари, дающие перевод какой-либо отраслевой терминологии на ряд языков, например, выпущенный в России в 1881 г. «Карманный русско-английско-французско-итальянско-датский и норвежско-латышский морской словарь». В последнее время довольно широкое распространение получили краткие многоязычные словари с подбором наиболее употребительных слов и выражений. Примером может быть «Славянский разговорник», выпущенный в Софии в 1961 г.. В нем приведены приветствия («Здравствуйте!» и т.п..), предостережения («Берегитесь!» и т.п.), слова для беседы на бытовые темы в гостях, в магазине, на почте и т.д. на русском, сербохорватском, болгарском, польском и чешском языках. Многоязычные словари могут иметь различную целевую установку. Так, в 18 и начале 19 века были распространены «каталоги языков», где к данному слову подбирались все известные переводы на любые языки; позднее этот тип стал более узким и практическим, объединяя переводы либо на группу родственных языков, либо на группу языков одной географической местности в помощь туризму и путешествиям.

К общим словарям мы отнесем также словари, рассматривающие (в принципе) всю лексику, но под каким-либо специфическим углом зрения. Таковы, в частности, ***словообразовательные*** (де­ривационные) словари, указывающие членение слов на составляющие их элементы, т.е. дают справки о морфологическом составе слова. Примером может служить «Школьный словообразовательный словарь» З.А. Потихи (1964г.). Далее ***этимологические*** словари (одного языка или группы родственных языков), содержащие сведения о про­исхождении и первоначальной мотивировке слов. Краткие этимоло­гические словари обыкновенно ограничиваются приведением для каждого слова одной этимологии, представляющейся автору словаря наиболее вероятной. В более крупных и солидных словарях, как правило, приводятся соответствия в родственных языках и излага­ются «контроверзы», т. е. споры ученых, касающиеся этимологии тех или иных слов, даются краткие сводки предложенных гипотез и их критическая оценка. В этимологические словари принято включать и слова, этимология которых остается неясной (в этих случаях так и указывают: «неясно»). Производные и сложные слова, мотивировка которых очевидна, либо вовсе не включаются в этимологический сло­варь, либо основные из них перечисляются для иллюстрации слово­образовательной активности производящего слова, или в тех слу­чаях, когда в производных отражены связи с какими-то более старыми значениями, утраченными производящим словом. Примером этимологических словарей могут служить «Этимологический словарь русского языка» А. Преображенского, «Russisches etymologisches Wörterbuch» М. Фасмера, который с 1966 г. начал выходить в русском переводе. Для практических целей может быть полезен вышедший в 1961 г. «Краткий этимологический словарь русского языка» Н.М. Шанского, В.В. Иванова и Т.В. Шанской.

От этимологических словарей следует отличать ***историчес­кие*** словари, которые, в свою очередь, представлены двумя разновидностями. В некоторых из них ставится цель — проследить эволюцию каждого слова и его отдельных значений на протяжении письменно засвидетельствованной истории соответствующего языка, обычно вплоть до современности (или какого-то отрезка этой исто­рии, тоже вплоть до современности). Примерами словарей этого типа могут служить «большой Оксфордский словарь» английского языка**,** немецкие словари — начатый братьями Гримм и словарь Г. Пауля, большой словарь Шведской академии и некоторые другие. Ко второй разновидности исторических словарей следует отнести словари древ­них периодов истории соответствующего языка, например «Материалы для словаря древнерусского языка» (в трех томах) филолога и этнографа Изм. Ив. Срезневского, изданный в 1893-1903гг., а дополнения к нему в 1912 г., а также словари отдельных писателей прошлого (в том числе и недавнего прошлого) или даже отдельных памятников.

Предшественниками исторических словарей были азбуковники, лексиконы и так называемые притекстовые словари: они помещались прямо при текстах и в них объяснялись слова только конкретного данного текста. Сущность исторического словаря Л. В. Щерба в свое время охарактеризовал следующим образом: «Историческим в полном смысле этого термина был бы такой словарь, который давал бы историю всех слов на протяжении определенного времени, причем указывалось бы не только возникновение новых слов и новых значений, но и их отмирание, а также их видоизменение».

Знакомство с историческими (как и с этимологическими) словарями позволяет выяснить историю слов и выражений современного языка, заглянуть в их «биографию». Так, например, открыв словарь И. И. Срезневского, можно узнать, что такие однокорневые и близкие по значению современные слова, как *работник, рабочий, рабочая* (о лице), восходят к слову *раб*, претерпев длительную эволюцию в своих значениях. Теперь прежняя связь со словом *раб* этих и других однокорневых слов никем непосредственно не осознается, например: *работа – рабство, неволя..*. (т. 3, с. 2 указанного словаря); *работати, работаю – находиться в рабстве, в неволе*... (т. 3, с. 4); *работъникъ – раб, невольник*... (т. 3, с. 5); *работъница – служанка, раба...; работъный – относящийся к рабству...; рабъ – слуга, невольник*... (т. 3, с. 5) и т. д. Эти и другие слова с таким же корнем снабжены примерами из древних памятников письменности.

Другая разновидность исторического словаря – это ***словарь писателя***. Словарь писателя или отдельного памятника должен быть исчерпы­вающим, т. е. должен а) включать абсолютно все слова, употреблен­ные в сочинениях (также в сохранившихся письмах и т. д.) данного писателя и б) указывать все встретившиеся формы этих слов. Обычно такой словарь не только иллюстрирует цитатами из текста все выде­ленные значения и оттенки значений, но и дает «адреса» всех случаев употребления слова (например, том, страницу, строку для каждого случая употребления). Если подобным образом строится словарь не одного писателя, а целого периода в истории языка, такой словарь оказывается исчерпывающим для этого периода, или так называемым «тезаурусом». Хорошим примером словаря писателя может служить «Словарь языка Пушкина» (т.1-4, АН СССР, М, 1956-1961), за рубежом созданы словари Шекспира, Гете и других больших литераторов. Такие словари нужны науке для того, чтобы можно было полнее и правильнее понять, как развивается так называемый язык художественной литературы, т.е. тот стиль общего литературного языка, который обслуживает художественное творчество, словесное искусство. Прежде всего, составляются словари по произведениям наиболее крупных писателей и поэтов, имеющих национальное значение в развитии культуры.

Особое место занимают ***диалектологические***, или диалектные сло­вари. Диалектный словарь может быть дифференциаль­ным, т. е. содержащим только диалектную лексику, отличающуюся от общенародной, либо полным, охватывающим в принципе всю лексику, бытующую в диалектной речи — как специфическую для данного диалекта, так и совпадающую с лексикой общенародного языка. Кроме того, он может быть либо словарем одного говора (даже говора одной деревни), либо словарем целой группы близких говоров, рассматриваемых как один диалект, либо, наконец, сравнительным словарем многих или даже всех территориальных диалектов какого-либо языка. К диалектологическим (в широком смысле) относятся словари слэнга и арго. Примерами словарей, включающих лексику одного диалекта, могут быть некоторые старые диалектные словари, такие, как «Материалы для объяснительного областного словаря вятского говора» Н. Васнецова (1908 г.), «Смоленский областной словарь» В. Добровольского (1914 г.), и новые: «Словарь современного русского народного говора» под ред. И.А. Оссовецкого, в котором дана лексическая система одного из говоров (дер. Деулино) Рязанской области, «Псковский областной словарь с историческими данными», который начал выходить с 1967 г.; «Словарь русских старожильческих говоров средней части бассейна р. Оби» и подобные. Словари, включающие разные диалекты языка, представлены «Опытом областного великорусского словаря» Академии наук (1852 г.), «Толковым словарем живого великорусского языка» В. Даля, «Словарем русских народных говоров» и подобными.

Интересный и сравнительно новый тип словарей — ***частотные*** словари. Их задача — показать сравнительную частоту употреб­ления слов языка в речи, что практически значит в некотором мас­сиве текстов. Примерами частотных словарей являются «The Russian Word Count» (Detroit, 1953) Йоссельсона, составленный на основе данных статистического анализа употребления около одного миллиона слов, и «Частотный словарь современного русского литературного языка», составленный Э. А. Штейнфельдт и изданный в Таллине в 1963 г. Словарь содержит 2500 наиболее употребительных слов, выбранных из современных текстов (художественной литературы для детей и взрослых, пьес, радиопередач, газет) общим объемом в 400 тысяч словоупотреблений. Важнейшими частями словаря являются: 1) об­щий список слов, расположенных в порядке убывания частоты, с ука­занием при каждом слове абсолютного количества встретившихся случаев его употребления; 2) список по частям речи с указанием (правда, не для всех частей речи) частоты отдельных грамматических форм (например, слово *год* встретилось 810 раз, в т. ч. 684 раза в ед. и 126 раз во мн. числе, 111 раз в именит., 244 раза в родит, п. и т. д.); 3) общий список слов по алфавиту с указанием частоты (для омони­мов — раздельно по частям речи; например, союз *а* встретился 3442 раза, частица *а —* 578 раз, междометие *а —* 54 раза). Кроме названных выше словарей существуют и другие. Например, "Подсчет русских слов" Х. Джоссельсона, «Частотный словарь немецкого языка» Ф. Кэдинга, «Сравнительный частотный список первой тысячи слов английского, французского, немецкого и испанского языков» Х. Итона. Частотные словари позволяют делать очень интересные выводы о функционировании в речи слов и грамматических категорий языка. Они имеют также большое практическое значение, в частности для рационального отбора лексики на разных этапах обучения данному языкутех, для кого он не является родным. Таким образом, развитие математических, прежде всего статистических методов изучения языка породило частотные словари, слова в которых получают числовой, статистический показатель, т.е. цифровые сведения о том, как часто применяется то или иное слово в языке.

Чисто практические цели преследуют орфографические и орфоэпическиесловари, указывающие «правильное» (т. е. отвечающее принятой норме) написание слов и их форм или, соответ­ственно, их «правильное» произношение. Наибольшее распространение получили ***орфографические***словари, дающие сведения о правильном написании слов. ***Орфоэпические*** словари указывают правильное произношение слов. Например, словарь-справочник «Русское литературное произношение и ударение» Р.И. Аванесова и С.И. Ожегова.

Среди специальных лингвистических словарей большой интерес представляют различные ***фразеологические*** словари. Они бывают переводными (например, англо-русский фразеологический сло­варь А. В. Кунина) и одноязычными, дающими толкование значений фразеологизмов средствами того же самого языка. К этому послед­нему типу принадлежит, в частности, «Фразеологический словарь рус­ского языка» под ред. А. И. Молоткова (М., 1967), включающий 4000 словарных статей, а также более ста­рый, но не потерявший своей ценности словарь М. И. Михельсона, широко дающий и иноязычные параллели к русским фразеологиз­мам, а также справки о происхождении фразеологизмов, об их пер­воначальной мотивировке и т. д.. Материалом фразеологических словарей являются не слова, а фразеологические обороты. Такие словари есть во всех языках. В русском наиболее распространены: «Крылатые слова» С.В. Максимова (ряд изданий) и Н.С. и М.Г. Ашукиных (М., 1960) и ранее упоминавшийся «Фразеологический словарь русского языка».

Разновидностью фразеологических словарей являются словари «крылатых слов», т. е. хо­довых цитат из литературных произведений, афоризмов знаменитых людей и др. фразеологизмов, главным образом книжного употребле­ния, имеющих литературный источник. Обычно в словарях такого рода большое место занимают «крылатые фразы», вошедшие в культурный обиход многих народов, в том числе и такие, которые нередко цити­руются в иноязычной форме, на том языке, на котором они были впер­вые сформулированы. Наиболее известным словарем «крылатых слов», послужившим в значительной мере образцом для всех прочих, был вышедший еще в 1864 г. словарь Г. Бюхмана «Geflugelte Worte» (с тех пор десятки раз переиздававшийся). Наиболее удачным из русских словарей этого типа можно считать словарь Н. С. и М. Г. Ашукиных. Особую разновидность фразеологических словарей составляют словари народных пословиц и поговорок, например, «Пословицы русского народа», собранные В. И. Далем (1862 г., 1957 г.), или «Mudroslovi narodu slovanskeho ve pnslovich» Ф. Челаковского — сравнительный словарь пословиц всех славянских народов, с отдель­ными параллелями и из неславянских языков (1-е изд.: Прага, 1851).

Из других специальных лингвистических словарей упомянем сло­вари ***синонимов, антонимов, омонимов, иностранных слов, словари сокращений, различные словари имен собственных, словари рифм***. Среди двуязычных специальных словарей отметим словари так называемых «ложных друзей переводчика», т. е. слов, близких в каких-либо двух языках по звучанию и написанию, но рас­ходящихся по значению (например, в болгарском *гора* значит ‘лес’, а вовсе не ‘гора’, в английском *magazine —* ‘журнал’, а не ‘магазин’, в украинском *уродливый —* ‘красивый’, а не ‘уродливый’, или в немец­ком *kalt*— ‘холодный’, а похожее итальянское *caldo* значит ‘горячий, теплый’). Поподробнее остановимся на словарях синонимов. «Словарь синонимов рус­ского языка» 3. А. Александровой (1969 г.), содержащий около 9 тысяч синонимических рядов.

Институтом русского языка АН СССР выпущен в свет двухтомный «Словарь синонимов русского языка» под редак­цией А. П. Евгеньева (1970-1971 гг.). Словарь включает 4148 словарных статей. Синонимы объединены в одну словарную статью на основе единства выражаемого понятия. В словарной статье имеются указания на особенности сочетаемости синонимов, оттенки их значения, сферу употребления, стилистическую окраску.

Например:

*Вынуть, достать, извлечь, вытащить (разг.), вытянуть (разг.), выудить (разг.).*

Несов.: *вынимать, доставать, извлекать, вытаскивать, вытягивать, выуживать.*

*(из чего, из-под чего, из-за чего* ***Взять****) то, что находится внутри, в глубине чего-либо.* ***Вынуть*** *- взять изнутри чего-либо;* ***достать*** *- взять наружу откуда-либо;* ***извлечь*** *- слово литературной речи;* ***вытащить*** *употр. в обиходно-бытовой речи;* ***вытянуть*** *употр., когда надо указать, что действие совершается медленно, постепенно или с некоторым затрудне­нием; выудить — достать то, что нужно, из глубины чего-либо, слово имеет шутливый, иронический характер (стр. 200).*

Толкование синонимов сопровождается многочисленными примерами их употребления, взятыми из языка художествен­ной литературы от Пушкина до наших дней и из публицисти­ческих и научных произведений.

Такие словари имеют большое практическое значение при изучении как своего, так и иностранного языка. Наряду с большими специальными синонимическими словарями очень полезны краткие, типа учебных пособий, синонимиче­ские словари, подобные «Краткому словарю синонимов рус­ского языка» В. Н. Клюепон (1956 г. и 1961 г.); «Краткому словарю синонимов английского языка» И. А. Потаповой (1957), «Краткому словарю синонимов французского языка» Л. С. Андреевской-Левенстерн и О. М. Карлович (1959 г.) и другим.

Очень интересен «Словарь иноязычных выражений и слов, употребляющихся в русском языке без перевода» А. М. Баб­кина и В. В. Шендецова (М., «Наука», 1966). В двух книгах словаря не только приводятся иноязычные слова, употребляю­щиеся без перевода (*a propos*, франц. - кстати, к случаю; *all right*, англ. - все в порядке; *alma mater*, лат. - почти­тельное наименование студентами своего университета), но и даются многочисленные примеры их употребления.

Особую группу составляют ***лингвистические справоч­ные словари***, в которых дается не объяснение значения слова или особенностей его употребления и происхождения, а приводятся различного рода справки о слове как языковой единице. Собственно говоря, справочный характер имеют и другие словари, в первую очередь толковые, но в данном случае выделяются те словари, у которых справочная функ­ция является основной, для них важно не объяснить слово, а дать о нем ту или иную справку лингвистического характера.

Такие словари следует отличать от ***нелингвистических специальных справочных*** словарей типа «Словаря литературных терминов» и т. п., и которых объясняются не слова, а понятия, предметы, явле­ния, называемые этими словами, справки даются не о словах (происхождении, составе и т. п.), а о самих предметах, понятиях, явлениях. Лингвистические справочные словари могут быть различного типа в зависимости от характера справок.

Интересный материал о морфологическом составе слова дают и так называемые ***обрат­ные словари***, где слова расположены не в порядке началь­ных букв, а в порядке конечных, так, например, в «Обратном сло­варе современного русского языка» (1958 г.) X.X. Бильфельдта слова располагаются так: *а, ба, баба, жаба, лаба* и т. д. — по «обратному алфавиту», т. е. считая с конца слова, а не с его начала. Подобные словари очень полезны для подсчета словарного заполнения грам­матических моделей (например, слов с суффиксами*-ик-, -чик-, -щик-, -арь-, -ня-, -ба-* и т. д.), для фонетической статистики фина­лей, т. е. концов слов, а также и для поисков нужной рифмы, в чем эти обратные словари перекрещиваются со словарями рифм. Однако ограничение подачи слова только в основной форме (суще­ствительные в именительном падеже единственного числа, глаголы в инфинитиве и т. п.) сужает поиски рифмы, которая может быть связана с другими словоформами.

Словарь иностранных слов дает краткое объяснение значений и происхождения иноязычных слов, указывает язык-источник (последнее обстоятельство сближает словари иностранных слов с этимологическими). Начало созданию таких словарей было положено при Петре I, по указанию которого был составлен рукописный "Лексикон вокабулам новым по алфавиту". Этот словарик содержал 503 слова. В словнике слова из сферы военного искусства, мореплавания, дипломатии, администрации. При словах на буквы А, Б, В, Г сделаны собственноручные поправки Петра (1725 г.). Известные словари XIX в.: "30 000 иностранных слов" А. Д. Михельсона (М., 1866); "Толковый словарь иностранных слов" Н. Дубровского (М., 1866). Интересно, что в первом издании словаря А. Д. Михельсона - 30 000 слов, а через 20 лет (в издании 1885 г.) - уже 115000: сказалось слишком широкое введение в словарь специальной терминологии. Из современных словарей наиболее известен "Словарь иностранных слов" под редакцией И. В. Лехина, С. М. Локшина, Ф. Н. Петрова (гл. ред.) и Л. С. Шаумяна (6-е изд. М., 1964,. 23 000 слов). Его публикация началась в 1939 г. Словарь Л. П. Крысина (2-е изд., доп. М., 2000) содержит около 25 000 слов и словосочетаний, вошедших в русский язык главным образом в XVIII-XX вв. (некоторые - в более раннее время), а также образованных в русском языке от иноязычных основ. Он является первым филологическим словарем иноязычных слов, т. е. таким, в котором описываются свойства слова, а не обозначаемой им вещи: его происхождение, значение в современном русском языке, а также произношение, ударение, грамматические характеристики, смысловые связи с другими иноязычными словами, стилистические особенности, типичные примеры употребления в речи, способность образовывать родственные слова.

***Словари неологизмов*** описывают слова, значения слов или сочетания слов, появившиеся в определенный период времени или употребленные только один раз (окказионализмы). В развитых языках количество неологизмов, зафиксированных в газетах и журналах в течение одного года, составляет десятки тысяч. Еще в стародавние времена неологизмы привлекали к себе внимание ученых. Словари неологизмов создавались эпизодически. Только с начала 70-х гг. XX в., когда почти одновременно вышли в свет близкие по характеру и объему словари новых слов (неологические) русского, английского и французского языков, стало возможным говорить о появлении новой лексикографической специализации со своей теоретической базой.

Иногда различают еще ***нормативные и ненормативные*** словари. К первым относят такие, которые уста­навливают определенные правила употребления слов, ко вторым — такие, где подобной задачи не ставится. Норма­тивным является большинство справочных словарей (орфо­эпические, орфографические), основная масса толковых словарей. К ненормативным относятся исторические, этимо­логические и т. п. словари. В последнее время в связи с усилением борьбы за культуру речи стали выпускать специальные словари, показывающие нормы словоупотребле­ния в особенно трудных случаях. Таков, например, изданный под редакцией С. И. Ожегова словарь-справочник «Правиль­ность русской речи» (М., 1962).

В заключение нашего обзора важнейших типов словарей отметим существование многочисленных промежуточных, переходных и смешан­ных типов. Так, переходными от лингвистических к нелингвистическим словарям являются ***словари терминов*** различных наук и отраслей техники. Эти словари бывают одноязычными, двуязыч­ными и многоязычными. Широкое распространение имеют терминологиче­ские словари, включающие специальные термины, употреб­ляемые в какой-либо научной области: химии, биологии, медицине, гидротехнике и т. п. Такие словари есть и для языкознания. Наиболее известны «Словарь лингвистических терминов» Ж. Марузо, изданный в русском переводе 1960 г., но во многом уже устаревший, и «Словарь лингвистических терминов» О. С. Ахмановой (М., «Советская энциклопедия», 1966), отражающий современную лингвистическую термино­логию. Словарь раскрывает содержание терминов и дает их иноязычные соответствия, что очень важно для чтения спе­циальной литературы на английском, немецком и других языках. Например,

*АНТРОПОНИМИКА англ. study of personal names, фр. anthroponymie, нем. Anthroponymie, исп. antro-ponimia. Раздел лексикологии, изучающий собст­венные имена людей (стр. 50).*

Терминологические словари могут содержать термины, употребительные в отдельных направлениях науки, научных школах. Таковы «Словарь американской лингвистической терминологии» Э. Хэмпа (М., «Прогресс», 1964) или «Лингви­стический словарь пражской школы» Й. Вахека (М., «Про­гресс», 1964).

Существует, наконец, тип универсаль­ных словарей, одновременно толковых и энциклопедических, включающих также этимологические и исторические справки, иногда важнейший материал иноязычных цитат, и снабженных в нужных случаях рисунками. Это различные «словари Лярусса» (по имени французского издателя, организовавшего выпуск таких словарей), в частности «Большой Лярусс», «Малый Лярусс» и т.д.; английские «словари Вебстера» (по имени первого составителя этих словарей), например, Webster’s Third New International Dictionary (Springfield. Mass., 1961), и другие издания и переработки, в том числе даже карманные; к этому типу примыкает в известном смысле и толковый словарь Хорнби.

**Составление словарей**

Важным вопросом при составлении словаря является вопрос о порядке расположения материала.

Чаще всего используется алфавитный порядок, иногда - в том или ином сочетании с иными принципами расположения. Напри­мер, во многих случаях применяется гнездование, т. е. объедине­ние в одно «гнездо» (в рамках одной словарной статьи) слов, связан­ных общностью корня, хотя бы это нарушало алфавитную последова­тельность. Фактически в этих случаях происходит отступлениеоталфавитного порядка слов в сторону алфавитного порядка корней. Это оказывается очень удобным для некоторых типов словарей, на­пример для словообразовательных и этимологических. Последова­тельное проведение гнездового принципа соответствует лексикогра­фической традиции многих языков. Так, например, арабские словари принято строить именно по алфавиту корней, помещая под каждым корнем все производные (в том числе и производные с приставками). Иногда и в словарях славянских языков глаголы с приставками вклю­чают в статью соответствующего бесприставочного глагола. Из рус­ских толковых словарей гнездовой принцип наиболее широко приме­няют первые издания словаря Даля (до переработки словаря Бодуэном де Куртенэ), однако для приставочных глаголов сделано исклю­чение — они идут на своем алфавитном месте.

Особое использование алфавитного принципа имеем в обрат­ных словарях. Слова в этих словарях располагаются по алфавиту не начальных, а конечных букв слова: *а, ба, баба, жаба, ...амёба, ...служба, ...изба, ...шайба, ...колба, ...дамба* и т. д. до последних слов, оканчивающихся на -яя: *передняя, ...безмужняя.* Среди неалфавитных принципов расположения материала важней­шим является принцип систематики (логической классификации) понятий, выражаемых лексическими единицами. Именно по этому принципу строятся упомянутые выше идеографические словари (называемые также «идеологическими» или «тематическими»). Вырабатывается та или иная логическая классификация понятий, и все, что подлежит включению в словарь, располагается по рубрикам этой классификации. Идеографические словари могут быть также двуязычными и многоязычными. Особую разновидность идеографических составляют «картинные словари» (по-французски «vocabulaires par l'image», по-немецки «Bildwörterbücher»), обычно двуязычные или мно­гоязычные. Они содержат рисунки, изображающие тот или иной «ку­сок действительности» (например, комнату с обстановкой, цех завода, угольную шахту, птицеводческую ферму, улицу и т. д.) и обильно снабженные номерами, под которыми на той же или соседней странице приводятся названия соответствующих предметов на одном, двух или нескольких языках. Из упомянутых выше словарей систематический принцип расположения применен в словарях пословиц Даля и Челаковского (ср. например, рубрики Челаковского: I. Бог. Религия. Черт. Грех... II. Добро - зло... III. Правда - ложь... и т.д.).

Составление словарей — очень сложная работа. Кроме обще­лингвистических положений о слове, его значениях и употребле­нии, грамматических и фонетических характеристик, надо знать технику составления словарей и понимать состав словаря.

Словарь состоит из: 1) словника, т. е. подбора со взаимными ссылками и отсылками, 2) филиации, т. е. расчлененной подачи значений той или иной вокабулы, 3) стилистических, грам­матических и фонетических ремарок или помет к словам и их значениям, 4) иллюстративных примеров, 5) идиоматических и фразео­логических сочетаний к данному слову и 6) перевода (в разноязыч­ных словарях) или толкования (объяснения — в одноязычных сло­варях).

Особо следует оговорить, что взаимно противоположные словари (допустим, русско-казахский и казахско-русский) никак нельзя мыслить просто как перестановку «правой колонки» (переводы) в «левую» (подлинники), и наоборот. Такие словари по словнику перекрывают друг друга лишь частично, так как каждый словарь «в подлинниках», т. е. в вокабулах, исходит из лексического состава своего языка, а, как известно, лексический состав разных языков (даже близкородственных) не совпадает. Поэтому любой перевод­ный словарь (есть ли уже «обратный словарь» или его нет) должен иметь свой идиоматичный данному языку словник, для чего лучше всего опиратьсяна одноязычный толковый словарь данного языка.

**Глава II. ЭЛЕКТРОННЫЕ СЛОВАРИ**

С появлением компьютерной техники, создатели программного обеспечения создали новый тип словарей - электронный словарь. Такой тип словаря - абсолютно новое слово в истории лексикографии, отметившее новую качественную ступень ее развития. Именно сейчас электронные словари вышли из тени бумажных и становятся самостоятельными игроками на языковой площадке, причем игроками, которые, похоже, в ближайшее время сделают остальных действующих лиц экспонатами Музея книги. Ведь электронные словари обладают рядом очевидных и существенных преимуществ по сравнению со словарями традиционными. Единственным же их недостатком является привязанность к персональному компьютеру и, следовательно, ограниченная доступность. Однако этот недостаток будет достаточно скоро устранен если не полностью, то, по крайней мере, большей частью, вследствие все возрастающих темпов компьютеризации, в том числе и растущей доступностью переносным компьютеров типа Laptop.

Электронных словарей сейчас выпущено довольно много, поэтому остановимся только на двуязычных англо-русских и русско-английских словарях. Для примера возьмем два самых известных: Lingvo компании Abbyy и МультиЛекс, разработанный фирмой МедиаЛингва. Эти словари любопытно сравнить, ведь создающие их команды, исповедуют разные взгляды на принципы электронной лексикографии.

**Электронные словари "МультиЛекс"**

Компания МедиаЛингва придерживается при создании словарей МультиЛекс довольно простой стратегии. Она создает цифровую копию известных книжных изданий. На сайте фирмы можно найти формулировку этого принципа: "В основу электронных словарей заложены словарные базы книжных изданий, уже завоевавших популярность и признание среди переводчиков, преподавателей иностранных языков, студентов и школьников". Некоторые эксперты считают, что такая политика покоится на эксклюзивном договоре МедиаЛингва с "естественным монополистом" рынка российских словарей, издательством "Русский язык". С точки зрения МедиаЛингва, задача электронной лексикографии - как можно точнее перевести традиционный словарь в электронную форму.

За основу словаря МультиЛекс взят "Новый большой англо-русский словарь" под редакцией А.Д. Апресяна. Есть и расширенная версия, где к основному словарю добавлены экономико-финансовый, юридический, строительный, политехнический словари и словарь по полиграфии и издательскому делу.

Конечно, словарь Апресяна - выдающееся достижение лексикографии, но подход МедиаЛингва имеет и недостатки. Первое, традиционные словари довольно серьезно отстают от языковой реальности. Обычно это не менее десяти лет. А электронные словари можно пополнять чуть ли не ежедневно. Второе, словари, содержащие сотни тысяч словарных статей, какими бы квалифицированными лексикографами они не составлялись, всегда содержат ошибки и неточности, не говоря уже о возникновении дополнительных значений слов. Жесткая привязка к бумажному прототипу не дает возможности исправлять и дополнять электронный, тем более изменять структуру построения словарной статьи.

**Электронные словари "Lingvo"**

По другому и, вероятно, более перспективному пути пошла компания Abbyy. Конечно, и в их большом электронном словаре Lingvo7.0 есть переведенные в цифровой вид лицензированные бумажные словари - это политехнический, юридический, экономический, финансовый, медицинский и - что очень своевременно - динамично пополняемый компьютерный словарь. Но основу Lingvo, по словам руководителя лингвистического отдела фирмы Владимира Селегея, составляет электронный словарь собственной разработки. Каждая новая версия Lingvo дополняется актуальной лексикой, и в ней исправляются найденные ошибки и неточности. Таким образом, благодаря лексикографическим исследованиям англо-русский словарь фирмы Abbyy близок к языковой практике.

Удачной находкой Abbyy выглядит приглашение всем желающим размещать на их Интернет-узле словари собственного изготовления http://www.lingvo.ru/dictionaries/index.htm. Такое вовлечение пользователей в лексикографическую работу вполне соответствует духу открытых Интернет-сообществ. Дополнительных словарей на сайте уже набралось 23 штуки. Причем всякий желающий может скачать их из Интернета и присоединить хоть все к уже имеющимся в базовой версии. Надо сказать, что базовая версия Lingvo-7.0 содержит миллион двести тысяч словарных статей. Причем основные статьи тщательно проработаны. Например, не самая большая статья о слове 'go' содержит более двадцати пяти тысяч знаков.

Первое, что бросается в глаза, когда мы говорим об электронных словарях - это резкое сокращение объема. На десятиграммовом компакт-диске помещается целая полка толстых словарей общим весом в двадцать пять килограмм. Но, естественно, не это главное. Важно, что электронный словарь принципиально может обойти ключевое противоречие книжной лексикографии: чем больше информации предлагает словарь, чем развитее его научный аппарат, тем сложнее им пользоваться. Поэтому классические словари разделяется на две категории. Первая - популярные, относительно удобные, но довольно простые. Вторая - обстоятельные академические издания, не позволяющие быстро получить искомую информацию.

Современные электронные словари не только значительно превосходят по объему книжные, но и находят искомое слово или словосочетание за несколько секунд. Причем искать можно в любой форме. Некоторые, например Lingvo, встраиваются во все основные офисные приложения и выделенное слово можно переводить нажатием нескольких клавиш.

**Преимущества электронных словарей**

При традиционном подходе минимальной единицей доступа является лексема (имя словарной статьи): нужно прочесть всю статью, чтобы определить, содержится ли в ней ответ на наш запрос. Для таких словарей, как оксфордский, это представляет серьезную проблему. Например, глагол 'set' имеет там 400 только основных значений (и у многих из них имеются подзначения).

Пользователь хотел бы, чтобы словарь максимально локализовал релевантную информацию. При этом речь не идет об автоматическом выборе переводного эквивалента (если мы говорим о переводном словаре). Специфика словарного ответа в том, что он дает весьма разнообразную информацию о слове или словосочетании, а не просто переводное соответствие, предполагает активный выбор пользователя из нескольких возможных хорошо обоснованных альтернатив.

Однако попытка решить проблему адекватной реакции словаря на запрос неизбежно наталкивается на сопротивление самого словарного материала, перенесенного из бумажного словаря.

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

Но, конечно, самое главное преимущество хороших электронных словарей - одновременный поиск не только по названию словарной статьи, но и по всему огромному объему словарей, что просто нереально в бумажном варианте. Такой поиск создает многомерный портрет слова, при этом извлекаются из глубин словарной статьи не только конкретные примеры его использования и устойчивые выражения, в которых слово встречается, но и обнажаются, становятся явными языковые законы, которым подчиняются правила словообразования. Даже мобильный электронный словарь не может отразить все сиюминутные движения языка, но он может дать ключ для расшифровки и понимания этих изменений, делая пользователя соавтором лексикографа. Что очень важно, когда требуется точный смысловой перевод, ведь это не задача подбора подходящего выражения, а в широком смысле отображение одной культуры с помощью языка другой. Поэтому в Lingvo можно строить и свой собственный словарь под общей оболочкой.

Язык — отражение реальной жизни. А жизнь не стоит на месте: появляются новые отрасли производства, науки, бизнеса, культуры. В обычную разговорную речь приходят новые слова, термины, устойчивые словосочетания. Можно ли представить в речи наших сограждан лет десять назад такие слова, как "холдинг", "транш"? Выражение "конечный пользователь" вызвало бы у них недоумение, и никто не мог бы предположить, что слово "мыло" будет означать на компьютерном жаргоне электронную почту (вольная русская транскрипция английского слова "e-mail").

Вся эта лексика не может быть адекватно отражена в "бумажных" словарях по той простой причине, что они слишком долго готовятся. Так, известный англо-русский словарь Мюллера, сочетающий относительную простоту пользования (один том, хотя и тяжелый!) и полноту содержания, был издан в 1960 году и с тех пор претерпел лишь косметические изменения в 1978 и 1994 годах.

Фактически многие словари, которые сформировались в языковой атмосфере середины века, сильно устарели. В них не указаны современные значения старых слов, а многие новые слова просто отсутствуют. Буквальное перенесение таких словарей на компьютеры бесперспективно. Это стало особенно очевидно в связи с развитием Интернета: большая часть Веб-страниц состоит из английских текстов, написанных живым современным языком, обильно использующим разговорную лексику и сленг. Вряд ли какой-либо из существующих англо-русских словарей может ответить на этот вызов. Решить данную задачу под силу лишь электронным словарям.

Большинство "бумажных" словарей ориентировано на человека, читающего на иностранном языке, то есть человека, который в непонятном ему тексте находит "опорные" слова, помогающие выстроить общую смысловую картину. Человек "пишущий", кроме знания всех используемых слов, должен четко представлять, как эти слова сочетаются друг с другом, какие предлоги при этом используются, есть ли устойчивые выражения, передающие необходимый смысл.

Увы, если "бумажный" словарь и удовлетворяет нужды Читателя, то уж интересы Писателя на неродном языке он чаще всего попросту игнорирует! А ведь в наш век электронных коммуникаций Писателем стал практически каждый пользователь Интернета!

И здесь электронный словарь оказывается намного полезнее "бумажного". Даже буквальное воспроизведение приличного "бумажного" словаря на компьютере дает возможность извлечь из него столь необходимую Писателю информацию, похороненную в глубинах словарных статей. Например, пользователь может открыть на экране сразу несколько словарных статей, характеризующих все значения слова "достать" (брать, получать, надоесть и т. д.) как на одном языке, так и на другом, и, таким образом, изучить все нюансы использования слова.

Однако более правильный путь — подумать о Писателе заранее, при составлении словаря. Чтобы учесть его интересы, надо уметь описывать способы образования сложных словосочетаний. Например, как передать по-английски смысл "подтасовать или фальсифицировать результаты выборов"? Это выражение не относится к идиоматическим, потому его не следует искать в словнике целиком. С другой стороны, оно не может быть правильно переведено и по частям. Логичнее всего искать это выражение в статье "election" (выборы). Однако, чтобы оно там оказалось, нужно желание разработчиков словаря его туда поместить.

Чтобы Писатель смог почувствовать оттенки слова, необходимо привести в словаре максимально возможное количество синонимов — слов, близких по смыслу. Например, английский глагол break означает, в частности: 1) ломать, разрушать, разбивать и 2) рвать, разрывать, отрывать. Для первого случая синонимами будут слова crush (давить, дробить) и smash (разбиваться вдребезги). Второму значению близки слова separate (отделять, разделять — более деликатный смысл) и tear off (отрывать, срывать). Через общую "карту" синонимов становится яснее, как перевести на иностранный язык слово с тем или иным смысловым оттенком. Очень полезны иллюстрирующие примеры, особенно на использование слов с предлогами или в устойчивых словосочетаниях. Например: "уехать отсюда", "уехать куда-то", "уехать за чем-то", "уехать" в значении "отсутствовать".

Информацию о синонимах, словосочетаниях и случаях употребления правильнее всего предоставить на родном языке пишущего: если Писатель русский - то в русско-английском словаре, если он англичанин — то в англо-русском. Ни для кого не секрет, насколько лучше помогают толковые английские словари при решении мучительной проблемы, какое слово употребить. А вот жесткая ориентация словаря на перевод, а не на ОПИСАНИЕ языка делает его использование Писателем непростым и неочевидным. Таким образом, в современных электронных словарях отражено пионерское достижение российской лексикографии - двуязычный словарь во многом становится толковым. Кроме того, такой электронный словарь, как Lingvo, строит нажатием нужной клавиши парадигму, то есть совокупность всех форм слова.

**Новые мехи требуют нового вина.**

Итак, мы видим новое противоречие: между новыми языковыми компьютерными технологиями и старым традиционным словарным содержанием, не позволяющим воспользоваться этими технологиями в полном объеме.  Иными словами, новые мехи требуют нового вина!

Источник этого противоречия тоже ясен: словарь представляет собой модель языка, устроенную на совершенно иных принципах, чем те формальные модели, которые лежат в основе этих технологий. И если в области морфологии противоречие еще не очень существенно, то в области синтаксиса и семантики оно становится почти непреодолимым.

Действительно, технология морфологического анализа всего лишь позволяет установить соответствие между исходной формой слова из текста и множеством лексем (словарных входов), для которой такая форма возможна. Синтаксический анализ позволяет сделать то же самое для словосочетаний, являющихся отдельными словарными входами.

Однако, для всех этих технологий само словарное содержание является "непрозрачным", полностью ими игнорируется. Заглянуть "внутрь" словарной статьи позволяет только полнотекстовый поиск. Однако, этот мощный инструмент работает со словарным содержимым как с текстом на естественном языке, что резко ограничивает его возможности.

Первый и очевидный шаг, на который уже идут создатели электронных словарей, это первичная разметка словарной статьи, формализация той внутренней структуры, которая в той или иной мере имеется в хороших бумажных словарях. В результате полнотекстовый поиск может различать, к примеру, переводы, примеры использования и комментарии, что принципиально усиливает его возможности с точки зрения пользователя.

Однако все эти меры являются поверхностными. Ясно, что задача состоит в том, чтобы единицей описания было отдельное лексическое значение, и технологии анализа могли бы устанавливать соответствие между исходным запросом и теми лексическими значениями, которые релевантны для этого запроса по синтаксическим и семантическим критериям.

В качестве иллюстрирующего примера можно привести практически любой глагол, принадлежащий ядру языка. Например, глагол "*развести*" может встретиться в таких контекстах:

*разводить руками;*  
*разводить спирт водой;*  
*разводить супругов;*  
*разводить мосты;*  
*разводить баранов;*  
*разводить дерущихся;*  
*разводить пилу;*  
*разводить/разбивать сады*

*(английские эквиваленты: bring; conduct; part, separate;  mix;  dissolve; divorce; breed; plant, etc...)*

Задача создания такого словарного содержания, которое позволило бы сделать единицей анализа отдельное лексическое значение, а не морфологическую лексему, является, вероятно, наиболее перспективным направлением в компьютерной лексикографии. Ясно, что для ее решения требуется "синхронизация" словарных описаний и формальных моделей, используемых технологиями анализа. В пределе это должно быть единое интегральное лексико-синтактико-семантическое описание.

**Читатели и писатели**

Интегральный подход к лексическим описаниям позволяет также решить и проблему "монофункциональности" бумажных словарей.

К примеру, особенностью большинства бумажных переводных словарей является ориентация описания структуры лексического значения в исходном языке на лексическую систему языка перевода и на реализацию ровно одной функции - собственно перевода с языка А на язык Б в предположении, что язык А является иностранным, а язык Б - родным. Нечего и говорить, что такое ограничение делает словарь исключительно неудобным при необходимости перехода от пользовательской модели Читатель к модели Писатель.

Фактически сегодня такие модели реализуются разными типами словарей, что достаточно неудобно читателю. Поэтому интегральный подход к лексическим описаниям оправдан не только методически (и, что немаловажно, экономически), но и с точки зрения учета интересов пользователя.

**Проблема актуальности**

Коснемся проблемы актуальности словарного содержания. Как уже указывалось, фундаментальные (лучшие!) бумажные словари -  неизбежно словари устаревшие.

Особенно это характерно для  разговорной лексики, в частности, ненормативной. В этой области отечественные классические словари предстают не только устаревшими, но попросту ханжескими.

Функции фиксации текущего состояния языка берут на себя растущие, как грибы после дождя, небольшие словарики, обычно весьма конъюнктурные и поверхностные. Новые значения в них оторваны от своих языковых корней, плохо или произвольно объяснены.

Для массовых программных продуктов, каковыми являются электронные словари,  характерны частая смена версий и  наличие постоянной обратной связи с тысячами пользователями. Поэтому компьютерная лексикография - это неизбежно актуальная лексикография.

Жизнь электронного словаря должна быть похожа на нелегкую жизнь других программных систем: с маниакальным стремлением особо вредных пользователей обнаружить очередную ошибку или лакуну, и, с другой стороны, с возможностью и необходимостью поправить дело сейчас, а не через десятилетия.

Такой подход всего лишь фиксирует естественное положение дел: коллективное авторство на словарное содержание принадлежит всем носителям языка; задача лексикографа - фиксация языковых фактов и их методически правильное описание.

**Соответствие уровню достижений лингвистической науки**

Отрыв лексикографической теории от  лексикографической практики велик. Это должно быть особенно обидно для российской лингвистической науки, в которой лексическая семантика занимает особое место. Достаточно назвать такие имена, как Мельчук, Апресян, Падучева  и многие другие.

Разумеется, существуют особые "концептуальные" словари, в которых лексика представлена интегрально и систематически. Например толковые и синонимические словари группы Апресяна.

При этом в массовых бумажных словарях эти идеи не воплощены в жизнь. А ведь именно в развитии этих идей будущее практической компьютерной лексикографии. Не углубляясь в подробный анализ теоретических концепций, являющихся одновременно и практически полезными, укажем следующие из них:

- Понятие "лексической функции", позволяющее систематически описывать несвободную сочетаемость слов. Например, то, что "войну ведут", а "экзамен - держат", что "теории выдвигают", а "мысли подают" и т.п.

- Описание семантики и практической реализации грамматического словоизменения и словообразования.  Каждый язык имеет свои собственные способы грамматического кодирования смысла. И эти способы никогда не описывается в массовых словарях систематически. Например, как передать по-английски смысл "довыпендриваться", даже если знаешь как передать "выпендриваться"?

- Синтаксические описания. Здесь ситуация наиболее печальна, поскольку в массовых словарях не существует даже системы понятий, с помощью которой синтаксическая информация могла бы быть доведена до обычного читателя. Идея, что за составление предложения ответственна грамматика, изложенная в справочнике, а словарь обеспечивает перевод отдельных слов, не выдерживает критики с точки зрения современных представлений о центральной роли слова в синтаксисе.

Выход из этой печальной ситуации уже указан. Будущее лексикографии за интегральными словарными описаниями, основанными на формальных моделях, учитывающих упомянутые научные результаты. На этих же моделях будут основываться технологии доступа к словарному содержанию.

**Глава III. МАШИННЫЙ ПЕРЕВОД**

**Немного истории.**

Осуществление перевода компьютером – сложная, но интересная научная задача. Основная ее сложность состоит в том, что естественные языки плохо поддаются формализации. Отсюда и невысокое качество получаемого с помощью систем МП текста, содержание и форма которого служит неизменным объектом шуток. Однако идея машинного перевода уходит корнями далеко в прошлое. Впервые мысль о возможности машинного перевода высказал Чарльз Бэббидж, разработавший в 1836-1848 гг. проект цифровой аналитической машины. Идея Ч. Бэббиджа состояла в том, что память объемом 1000 50-разрядных десятичных чисел (по 50 зубчатых колес в каждом регистре) можно использовать для хранения словарей. Ч. Бэббидж привел эту идею в качестве обоснования для запроса у английского правительства средств, необходимых для физического воплощения аналитической машины, которую ему так и не удалось построить.

А через 100 лет, в 1947 году, У. Уивер (директор отделения естественных наук Рокфеллеровского фонда) написал письмо Норберту Винеру. В этом письме он предлагал использовать технику дешифрования для перевода текстов. Этот год считается годом рождения машинного перевода. В этом же году был разработан алгоритм осуществления пословного перевода, а в 1948 году Р. Риченс предложил правило разбиения слова на основу и окончание. В последующие два десятилетия системы машинного перевода бурно развивались. В январе 1954 года на машине IBM – 701 была продемонстрирована первая система машинного перевода IBM Mark II. Но в 1967 году специально созданная Комиссия Национальной Академии Наук США признала машинный перевод нерентабельным, что существенно затормозило исследования в этой области. Новый подъем машинный перевод переживает в 70-е годы, а в 80-е становится экономически выгодным за счет сравнительной дешевизны машинного времени.

Однако в СССР исследования в области машинного перевода продолжались. После демонстрации системы IBM Mark II  группа ученых ВИНИТИ начала разработку системы машинного перевода для машины БЭСМ. Первый образец перевода с английского на русский язык был получен к концу 1955 года.

Другое направление работ возникло в Отделении прикладной математики Математического института АН СССР (ныне ИПМ им. М. В. Келдыша РАН) по инициативе А. А. Ляпунова. Первые программы машинного перевода, разработанные этим коллективом, были реализованы на машине "Стрела". Благодаря работе над созданием систем МП оформилось такое направление, как прикладная лингвистика.

В 70-е годы над созданием систем МП трудилась группа разработчиков ВИНИТИ РАН под руководством проф. Г.Г. Белоногова. Первая их система МП была разработана в 1993 году, а в 1996 году после ряда доработок была зарегистрирована в РОСАПО под названием Retrans. Эта система использовалась министерствами обороны, путей сообщения, науки и технологии.

Параллельные исследования велись в лаборатории Инженерной Лингвистики ЛГПИ им. А. И. Герцена (ныне Педагогический Университет). Именно они и легли в основу наиболее популярной сейчас системы МП “PROMT”. Последние версии этого программного продукта используют наукоемкие технологии и построены на основе технологии расширенных сетей переходов и формализма нейронных сетей.

**Классификация систем машинного перевода по Лари Чайлду**

Новые члены форума по иностранным языкам компании CompuServe зачастую задают вопрос о том, не мог ли бы кто-нибудь посоветовать им хорошую программу машинного перевода за умеренную цену.

Ответом на этот вопрос неизменно является "нет". В зависимости от отвечающего, ответ может содержать два основных аргумента: либо о том, что машинам перевод не под силу, либо, что машинный перевод стоит слишком дорого.

Оба эти аргумента в определенной степени справедливы. Однако ответ далеко не так прост. Изучая проблему машинного перевода (МП), следует рассмотреть отдельно различные подразделы этой проблемы. Следующее разделение основано на лекциях Лари Чайлдса, проведенных в рамках Международной Конференции по Техническим Коммуникациям 1990 года:

*- полностью автоматический перевод;*

*- автоматизированный машинный перевод при участии человека;*

*- перевод, осуществляемый человеком с использованием компьютера.*

**Полностью автоматизированный машинный перевод.** Этот вид машинного перевода и подразумевается большинством людей, когда они говорят о машинном переводе. Смысл здесь прост: 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.

Основной проблемой является сложность языка как такового. Возьмем, к примеру, значения слова "can". Помимо основного значения модального вспомогательного глагола, у слова "can" имеется несколько официальных и жаргонных значений в качестве существительного: "банка", "отхожее место", "тюрьма". Кроме этого, существует архаичное значение этого слова - "знать или понимать". Если предположить, что у выходного языка для каждого из этих значений имеется отдельное слово, каким образом может компьютер их различить?

Как оказалось, определенные успехи были достигнуты в сфере разработки программ перевода, различающих смысл основываясь на контексте. Более поздние исследования при анализе текстов опираются больше на теории вероятности. Тем не менее, полностью автоматизированный машинный перевод текстов с обширной тематикой все еще является невыполнимой задачей.

**Автоматизированный машинный перевод при участии человека.** Этот вид машинного перевода теперь вполне осуществим. Говоря о машинном переводе при участии человека, обычно подразумевают редактирование текстов как до, так и после их обработки компьютером. Люди-переводчики изменяют тексты так, чтобы они были понятны машинам. После того, как компьютер сделал перевод, люди опять-таки редактируют грубый машинный перевод, делая текст на выходном языке правильным. Помимо такого порядка работы, существуют системы МП, во время 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 узко-ограниченной тематики.

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, да и работа эта не из приятных. Создание и поддержание в рабочем состоянии баз данных слов - 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.

**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.

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

Вот список наиболее известных систем ТМ:

- Transit швейцарской фирмы Star,

- Trados (США),

- Translation Manager от IBM,

- Eurolang Optimizer французской фирмы LANT,

- DejaVu от ATRIL (США),

- WordFisher (Венгрия).

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

В настоящее время ведутся разработки по усовершенствованию систем ТМ. Например, ядро системы Transit фирмы Star реализовано на основе технологии нейронных сетей.

Несмотря на широкий ассортимент систем TM, они имеют несколько общих функций:

- Функция сопоставления (Alignment). Одно из преимуществ систем ТМ – это возможность использования уже переведенных материалов по данной тематике. База данных ТМ может быть получена путем посегментного сопоставления файлов оригинала и перевода.

- Наличие фильтров импорта – экспорта. Это свойство обеспечивает совместимость систем ТМ с множеством текстовых процессоров и издательских систем и дает переводчику относительную независимость от заказчика.

- Механизм поиска нечетких или полных совпадений. Именно этот механизм и представляет собой основное достоинство систем ТМ. Если при переводе текста система встречает сегмент, идентичный или близкий к переведенному ранее, то уже переведенный сегмент предлагается переводчику как вариант перевода текущего сегмента, который может быть подкорректирован. Степень нечеткого совпадения задается пользователем.

- Поддержка тематических словарей. Эта функция помогает переводчику придерживаться глоссария. Как правило, если в переводимом сегменте встречается слово или словосочетание из тематического словаря, то оно выделяется цветом  и предлагается его перевод, который можно вставить в переводимый текст автоматически.

- Средства поиска фрагментов текста. Этот инструмент очень удобен при редактировании перевода. Если в процессе работы был найден более удачный вариант перевода какого-либо фрагмента текста, то этот фрагмент может быть найден во всех сегментах ТМ, после чего в сегменты ТМ последовательно вносятся необходимые изменения.

Конечно, как и любой программный продукт, системы ТМ имеют свои достоинства и недостатки, и свою область применения. Однако в отношении систем TM, основным недостатком является их дороговизна.

Особенно удобно использовать системы ТМ при переводе таких документов, как руководства пользователя, инструкции по эксплуатации, конструкторская и деловая документация, каталоги продукции и другой однотипной документации с большим количеством совпадений.

**Системы машинного перевода (МП)**

В соответствии с указанной выше классификацией, целью данной работы является исследование и анализ систем МП второй группы, поскольку систем МП первой группы еще не существует в природе, а системы третьей группы в сущности не являются системами МП, а более напоминают электронные словари.

Системы МП осуществляют автоматизированный перевод текста. Единицами перевода при этом служат слова или словосочетания, причем последние разработки позволяют учитывать морфологию переводимого слова. Развитые системы МП осуществляют перевод по заданным разработчиком и/или корректируемым пользователем алгоритмам перевода.

Для осуществления машинного перевода в компьютер вводится специальная программа, реализующая алгоритм перевода, под которым понимается последовательность однозначно и строго определенных действий над текстом для нахождения переводных соответствий в данной паре языков L1 – L2 при заданном направлении перевода (с одного конкретного языка на другой). Система машинного перевода включает в себя двуязычные словари, снабженные необходимой грамматической информацией (морфологической, синтаксической и семантической) для обеспечения передачи эквивалентных, вариантных и трансформационных переводных соответствий, а также алгоритмические средства грамматического анализа, реализующие какую-либо из принятых для автоматической переработки текста формальных грамматик. Имеются также отдельные системы машинного перевода, рассчитанные на перевод в рамках трех и более языков, но они в настоящее время являются экспериментальными.

Наиболее распространенной является следующая последовательность формальных операций, обеспечивающих анализ и синтез в системе машинного перевода:

1. На первом этапе осуществляется ввод текста и поиск входных словоформ (слов в конкретной грамматической форме, например дательного падежа множественного числа) во входном словаре (словаре языка, с которого производится перевод) с сопутствующим морфологическим анализом, в ходе которого устанавливается принадлежность данной словоформы к определенной лексеме (слову как единице словаря). В процессе анализа из формы слова могут быть получены также сведения, относящиеся к другим уровням организации языковой системы.

2. Следующий этап включает в себя перевод идиоматических словосочетаний, фразеологических единств или штампов данной предметной области (например, при англо-русском переводе обороты типа *in case of*, *in accordance with* получают единый цифровой эквивалент и исключаются из дальнейшего грамматического анализа); определение основных грамматических (морфологических, синтаксических, семантических и лексических) характеристик элементов входного текста (например, числа существительных, времени глагола, синтаксических функций словоформ в данном тексте и пр.), производимое в рамках входного языка; разрешение омографии (конверсионной омонимии словоформ – скажем, англ. *round* может быть существительным, прилагательным, наречием, глаголом или же предлогом); лексический анализ и перевод лексем. Обычно на этом этапе однозначные слова отделяются от многозначных (имеющих более одного переводного эквивалента в выходном языке), после чего однозначные слова переводятся по спискам эквивалентов, а для перевода многозначных слов используются так называемые контекстологические словари, словарные статьи которых представляют собой алгоритмы запроса к контексту на наличие/отсутствие контекстных определителей значения.

3. Окончательный грамматический анализ, в ходе которого доопределяется необходимая грамматическая информация с учетом данных выходного языка (например, при русских существительных типа *сани*, *ножницы* глагол должен стоять в форме множественного числа, несмотря на то, что в оригинале может быть и единственное число).

4. Синтез выходных словоформ и предложения в целом на выходном языке.

В зависимости от особенностей морфологии, синтаксиса и семантики конкретной языковой пары, а также направления перевода общий алгоритм перевода может включать и другие этапы, а также модификации названных этапов или порядка их следования, но вариации такого рода в современных системах, как правило, незначительны. Анализ и синтез могут производиться как пофразно, так и для всего текста, введенного в память компьютера; в последнем случае алгоритм перевода предусматривает определение так называемых анафорических связей (такова, например, связь местоимения с замещаемым им существительным – скажем, местоимения им со словом местоимения в самом этом пояснении в скобках).

В настоящее время существует две концепции развития систем МП:

1.      Модель «большого словаря со сложной структурой», которая заложена в большинство современных программ-переводчиков;

2.      Модель «смысл-текст», впервые сформулированная А.А. Ляпуновым, но пока что не реализована ни в одном коммерческом продукте.

На сегодняшний день наиболее известны такие системы машинного перевода, как

- PROMT 2000/XT компании PROMT;

- Retrans Vista  компаний Vista и Advantis;

- Сократ – набор программ компании Арсеналъ.

В настоящее время качество машинного перевода оставляет желать много лучшего, и само наличие таких систем пока правильнее воспринимать как предмет научных исследований. В большинстве случаев при работе над проектом применение систем МП не оправдано, поскольку:

- Системы МП не дают приемлемого качества выходного текста. Более высокого качества можно добиться с помощью предварительной настройки системы (продукты серии PROMT XT предоставляют пользователю множество возможностей для этого), что совершенно неприемлемо при небольших объемах переводимого текста, и/или путем последующего редактирования, а это только замедляет работу, если переводчик использует слепой метод печати.

- Системы МП не гарантируют соблюдения единства терминологии, особенно при работе коллектива переводчиков над большим проектом. Вернее, могут гарантировать при условии очень внимательного обращения с пользовательскими словарями, а на это не всегда стоит рассчитывать.

Однако в некоторых случаях использование систем МП все же помогает сократить временные затраты. Это происходит, если текст достаточно объемный и содержит однообразную терминологию, что позволяет сравнительно быстро настроить под него систему МП. Тогда редактирование текста не займет слишком много времени. Однако в этом случае следует особенно внимательно отнестись к стилю текста перевода. Машинный перевод формален, поэтому высока вероятность калькирования синтаксических структур языка оригинала, которое характерно для перевода вообще, а потому вполне может быть пропущено при редактировании.

Вообще говоря, системы  МП вполне могут применяться там, где используется максимально стандартизованный язык с простой грамматикой и сравнительно небольшим запасом слов. Довольно успешным проектом системы МП считается немецкая программа Meteo, выполняющая перевод метеопрогнозов с французского языка на английский и обратно. Для облегчения работы переводчиков и технических писателей компанией Boeing в свое время был разработан стандарт языка для написания технической документации, который известен как Boeing English.

**Система МП Retrans Vista.** Системы машинного перевода текстов с одних естественных языков на другие моделируют работу человека-переводчика. Их эффективность зависит, прежде всего, от того, в какой степени в них учитываются объективные законы функционирования языка и мышления. К сожалению, эти законы пока еще недостаточно изучены. Решая проблему машинного перевода, необходимо учитывать богатый опыт межнационального общения и опыт переводческой деятельности, накопленный человечеством. А этот опыт свидетельствует о том, что в процессе перевода в качестве основных единиц смысла рассматриваются, прежде всего, фразеологические словосочетания, выражающие целостные понятия, а не отдельные слова. Именно понятия являются теми элементарными мыслительными образами, используя которые можно строить более сложные мыслительные образы, соответствующие переводимому тексту.

Условимся называть системы машинного перевода, в которых в качестве основных минимальных единиц смысла рассматриваются не отдельные слова, а фразеологические словосочетания, системами фразеологического машинного перевода. В этих системах отдельные слова также могут использоваться, но они рассматриваются как вспомогательные единицы смысла, к которым приходится прибегать за неимением лучших.

Система фразеологического машинного перевода должна включать в свой состав базу знаний, содержащую переводные эквиваленты для наиболее часто встречающихся фраз, фразеологических сочетаний и отдельных слов, и программные средства для морфологического и синтаксического анализа и синтеза текстов и для их редактирования человеком. В процессе перевода текстов система использует хранящиеся в ее базе знаний переводные эквиваленты в следующем порядке: сначала делается попытка перевести всю фразу как целостную единицу; далее, в случае неудачи, входящие в ее состав словосочетания; и, наконец, осуществляется пословный перевод тех фрагментов текста, которые не удалось перевести первыми двумя способами. Фрагменты выходного текста, полученные всеми тремя способами, должны грамматически согласовываться друг с другом (с помощью процедур морфологического и синтаксического синтеза).

Принципы построения систем фразеологического машинного перевода текстов были впервые сформулированы в 1975 году в предисловии к книге Д. Жукова "Мы переводчики". В более полном виде они были изложены в 1983 году в книге Г. Г. Белоногова и Б. А. Кузнецова "Языковые средства автоматизированных информационных систем". Наконец, в 1993 году были опубликованы две статьи, в которых были описаны система машинного перевода, построенная на этих принципах, и методы автоматизированного составления двуязычных словарей по параллельным (русских и английским) текстам. Важнейшими среди этих принципов являются следующие:

1. Основными единицами языка и речи, которые, прежде всего, следует включать в машинный словарь, должны быть фразеологические единицы (словосочетания, фразы). Отдельные слова также могут включаться в словарь, но они должны использоваться только в тех случаях, когда не удается осуществить перевод, опираясь только на фразеологические единицы.

2. Наряду с фразеологическими единицами, состоящими из непрерывных последовательностей слов, в системах машинного перевода следует использовать и так называемые "речевые модели" - фразеологические единицы с "пустыми местами", которые могут заполняться различными словами и словосочетаниями, порождая осмысленные отрезки речи.

3. Реальные тексты, независимо от их принадлежности к той или иной тематической области, обычно бывают политематическими, если они имеют достаточно большой объем. Поэтому машинный словарь, предназначенный для перевода текстов даже только из одной тематической области, должен быть политематическим, а для перевода текстов из различных предметных областей - тем более. Он должен создаваться, прежде всего, на основе автоматизированной обработки двуязычных текстов, являющихся переводами друг друга, и в процессе функционирования систем перевода.

4. Наряду с основным политематическим словарем большого объема, в системах фразеологического машинного перевода целесообразно использовать также набор небольших по объему дополнительных тематических словарей. Дополнительные словари должны содержать только ту информацию, которая отсутствует в основном словаре (например, информацию о приоритетных переводных эквивалентах словосочетаний и слов для различных предметных областей).

На основе описанных принципов в ВИНИТИ РАН (см. выше) были построены две системы фразеологического машинного перевода:   
1) система русско-английского перевода (RETRANS)   
2) система англо-русского перевода (ERTRANS).

Обе системы имеют одинаковую структуру и примерно одинаковые объемы машинных словарей. Поэтому мы рассмотрим только первую систему.

Система RETRANS имеет следующие характеристики:

1. Область применения, назначение, функциональные возможности. Система предназначена для автоматизированного перевода научно-технических текстов с русского языка на английский. Русско-английский политематический машинный словарь системы содержит терминологию по естественным и техническим наукам, экономике, бизнесу, политике, законодательству и военному делу. В частности, он содержит термины и фразеологические единицы по следующим тематическим областям: Машиностроение, Электротехника, Энергетика, Транспорт, Аэронавтика. Космонавтика, Робототехника, Автоматика и Радиоэлектроника, Вычислительная Техника, Связь, Математика, Физика, Химия, Биология, Медицина, Экология, Сельское Хозяйство, Строительство и Архитектура, Астрономия, География, Геология, Геофизика, Горное Дело, Металлургия и др.

Перевод текстов может осуществляться в автоматическом и в диалоговом режимах.

2. Объем политематического машинного словаря: более 1.300.000 словарных статей; 77 процентов из них составляют словосочетания длиной от двух до семнадцати слов. Объем дополнительных машинных словарей (для настройки системы на различные тематические области) - более 200.000 словарных статей.

**Система МП PROMT XT.** В основу программных продуктов компании PROMT поставлено решение следующих фундаментальных проблем:

Во-первых, всем ясно, что чем больше словарь, тем лучше перевод, значит, первая проблема - проблема создания больших словарей для систем.

Во-вторых, ясно, что система должна переводить такие предложения: ПРИВЕТ, КАК ДЕЛА? Значит, еще одна проблема - научить систему распознавать устойчивые обороты.

В-третьих, понятно, что предложение для перевода пишется по определенным правилам, по определенным правилам переводится, а значит есть еще одна проблема: записать все эти правила в виде программы. Вот, собственно, и все.

Самое интересное, что эти проблемы действительно являются основными при разработке систем машинного перевода, другое дело, что методы их решения известны далеко не всем и отнюдь не так просты, как может показаться.

Методы организации больших баз данных достаточно хорошо разработаны, но для перевода не менее, а может быть, и более важно правильно структурировать информацию, которая приписывается элементу базы, правильно выбрать этот самый элемент. Сколько, например, записей в словаре должно соответствовать обыкновенному русскому слову "программа"? И, вообще, большой словарь - это словарь, который содержит много словарных статей, или словарь, который позволяет распознать много слов из текста? Очевидно, более верно второе. Поэтому для описания и входного, и выходного языка в системе должен существовать некоторый формальный метод описания морфологии, на котором основывается выбор единицы словаря.

Практически во всех системах, которые претендуют на то, чтобы считаться системами перевода, проблема представления морфологических моделей так или иначе решается. Но одни системы могут распознать миллион словоформ при объеме словаря в пятьдесят тысяч словарных статей, а другие при объеме словаря в сто тысяч словарных статей могут распознать именно эти сто тысяч.

В системах семейства PROMT разработано практически уникальное по полноте морфологическое описание для всех языков, с которыми системы умеют обращаться. Оно содержит 800 типов словоизменений для русского языка, более 300 типов, как для немецкого, так и для французского языка, и даже для английского, который не принадлежит к флективным языкам, выделено более 250 типов словоизменений. Множество окончаний для каждого языка хранится в виде древесных структур, что обеспечивает не только эффективный способ хранения, но и эффективный алгоритм морфологического анализа.

Кроме того, используемая модель морфологии позволила разработать экспертную систему для пользователя - создателя словаря. Эта система фактически автоматизирует процедуру выделения основы и определения типа словоизменения при вводе новых словарных статей.

Такой возможности нет ни в одной из существующих систем машинного перевода, даже в таких распространенных системах как Power Translator (Globalink, США), Language Assistant (MicroTac, США), TRANSEND (Intergaph,США), где пользователям приходится вручную спрягать и склонять слова для задания морфологической модели.

Однако разработка описания морфологии позволяет решить только проблему того, что является заголовком словарной статьи, по которому происходит идентификация единицы текста и единицы словаря. Но ведь идентификация слова из текста со словарной статьей происходит не ради идентификации, как это требуется в спеллерах или электронных словарях, она необходима для выполнения программой собственно процедур перевода. Какая же нужна информация в словарной статье и как должны быть описаны правила перевода для того, чтобы программа переводила?

Во многих системах МП в прошлом (как, впрочем, и сейчас) словарное описание и описание алгоритмов рассматривались как стороны одной проблемы, но решение, как правило, искалось в ограничении рассматриваемого мира, либо грамматического, либо семантического. Например, на основе признака "принадлежность к части речи" описывалась грамматика такого типа:

именная группа - это существительное

именная группа - это прилагательное + именная группа

глагольная группа - это глагол + именная группа

предложение - это именная группа + глагольная группа

Понятно, что некоторая часть предложений естественного языка описывается такой грамматикой, но эта часть очень незначительна, и на ее основе нельзя правильно анализировать и переводить хоть сколько-нибудь реальный текст. Но зато можно использовать эффективные методы построения преобразователя по заданной грамматике или, на худой конец, написать программу, которая путем перебора построит древа зависимостей для ограниченного множества предложений. Такие системы точно так же получали определения "экспериментальные".

Так или иначе, но именно из таких проектов появились системы перевода, которые сейчас предлагаются конечному пользователю. Это и Power Translator (компания Globalink) и Language Assistant (компания MicroTac) и TRANSEND (компания Intergraph).

Системы семейств STYLUS и PROMT - не исключение, поскольку многие специалисты компании PROMT имели опыт работы в такого типа проектах. Однако при разработке систем PROMT впервые был применен фактически революционный подход, который и позволил получить впечатляющие результаты. Системы перевода семейства PROMT - это системы, спроектированные на основе не лингвистических, а кибернетических методов.

Оказалось, что очень продуктивно рассматривать систему перевода не как транслятор, задачей которого является перевод текста, допустимого с точки зрения входной грамматики, а как некоторую сложную систему, задачей которой является получение результата при произвольных входных данных, в том числе и для текстов, которые не являются правильными для грамматики, с которой работает система.

Вместо принятого лингвистического подхода, предполагающего выделение последовательных процессов анализа и синтеза предложения, в основу архитектуры систем было положено представление процесса перевода как процесса с "объектно-ориентированной" организацией, основанной на иерархии обрабатываемых компонентов предложения. Это позволило сделать системы PROMT устойчивыми и открытыми.

Кроме того, такой подход дал возможность применения различных формализмов для описания перевода разных уровней. В системах работают и сетевые грамматики, близкие по типу к расширенным сетям переходов, и процедурные алгоритмы заполнения и трансформаций фреймовых структур для анализа сложных предикатов.

Описание лексической единицы в словарной статье, которое фактически не ограничено по размерам и может содержать множество различных признаков, тесно взаимосвязано со структурой алгоритмов системы и структурировано не на основе извечной антитезы синтаксис - семантика, а на основе уровней компонентов текста.

При этом системы могут работать и с не полностью описанными словарными статьями, что является важным моментом при открытии словарей для пользователя, от которого нельзя требовать тонкого обращения с лингвистическим материалом.

Первая система машинного перевода, выпущенная компанией PROMT в 1991 году, переводила с английского языка на русский специализированные тексты по программному обеспечению. Она использовала небольшой словарь - около 17 тыс. слов и выражений, работала в среде ДОС и не имела средств настройки для пользователя. Но уже эта первая система была правильно устроена, и нынешняя технология разработки алгоритмов машинного перевода, применяемая в компании PROMT, не претерпела значительных изменений. Напротив, найденный тогда подход оказался очень плодотворным для самых разных языков.

Сначала поясним некоторые определения: вместе с развитием машинного перевода как области прикладной лингвистики появились и классификации систем, и стало принято делить системы перевода на системы типа TRANSFER и системы типа INTERLINGUA. Это разделение основано на особенностях архитектурных решений для лингвистических алгоритмов.

Алгоритмы перевода для систем типа TRANSFER строятся как композиция трех процессов: анализ входного предложения в терминах структур входного языка, преобразование этой структуры в аналогичную структуру выходного языка (TRANSFER) и затем синтез выходного предложения по полученной структуре.

Системы типа INTERLINGUA предполагают априори наличие некоторого метаязыка структур (INTERLINGUA), на котором можно описать все структуры как входного, так и выходного языков в общем случае; поэтому алгоритм перевода в системе типа INTERLINGUA предполагается как более простой: анализ входного предложения в терминах метаязыка и затем синтез из метаструктуры соответствующего предложения выходного языка. "Единственная" сложность в этом случае - разработать сам метаязык и описать естественный язык в соответствующих терминах.

Несмотря на то, что эта классификация существует, и в среде разработчиков машинного перевода считается хорошим тоном спросить, к какому типу относится система PROMT, не было разработано еще не одной реальной системы, основанной на принципе INTERLINGUA.

Система PROMT не является исключением, и на этот вопрос мы отвечаем: наша система выполняет перевод типа TRANSFER. Но это очень простой ответ, он практически не отражает особенностей архитектуры системы PROMT. А особенности состоят в том, что этот метод (TRANSFER) применен в системе не в соответствии с лингвистическим стандартным подходом.

Дело в том, что система перевода, как правило, работает в условиях не полностью описанных данных, ведь в язык - это живая система, которая развивается очень быстро: постоянно появляются новые слова, новые функции старых слов, и, вместе с новыми сущностями, новые значения. В этих условиях определяющим структурным свойством алгоритмов перевода становится их устойчивость к произвольным входным данным, и в основу алгоритмов, выполняющих перевод в системе PROMT, вместо последовательного TRANSFER'а был заложен иерархический подход, разделяющий процесс перевода на взаимосвязанные TRANSFER'ы для разных единиц анализа.

В системе выделяется уровень лексических единиц, уровень групп, уровень простых предложений и уровень сложных предложений. Все эти процессы связаны и взаимодействуют иерархически в соответствии с иерархией текстовых единиц, обмениваясь синтезируемыми и наследуемыми признаками. Такое устройство алгоритмов позволяет использовать разные формальные методы для описания алгоритмов разных уровней.

Рассмотрим уровень лексических единиц: лексическая единица - это слово или словосочетание, которое является единицей самого низкого уровня. И в случае входного, и в случае выходного языка слово описывается как совокупность основы и окончания. Это обеспечивает возможность, с одной стороны, распознавания входных слов и анализа входной морфологии и, с другой стороны,  удобного синтеза выходных слов по их морфологической информации (основа, тип словоизменения и адрес окончания в массиве окончаний этого типа). Таким образом, если ввести правила преобразования входной морфологической информации в выходную морфологическую информацию, осуществляется TRANSFER на морфологическом уровне.

Уровень групп рассматривает структуры более сложные: группы существительных, прилагательных, наречий и сложные глагольные формы. Этот уровень при анализе, основываясь на формальных сетевых грамматиках, умеет соединять группы в синтаксические единицы, каждая из которых характеризуется синтезированной структурной информацией и главным элементом группы. По входной структуре, полученной в терминах непосредственных составляющих, вместе с синтезированными признаками формируется выходная группа как набор лексических единиц со значениями морфологических признаков, которые могут наследоваться исходя из результатов анализа группы. Таким образом, реализуется TRANSFER на уровне групп.

Анализ простых предложений как структур, состоящих из синтаксических единиц, выполняется на основе фреймовых предикатных структур, которые позволяют эффективно выполнять преобразования. Глагол считается для простых предложений главным элементом и его валентности определяют заполнение соответствующего фрейма. Для каждого типа фреймов существует некоторый закон преобразования в выходной фрейм и оформление актантов. Таким образом, осуществляется TRANSFER на уровне предложений. Анализ сложных предложений требуется в случае формирования согласования времен и правильного перевода союзов.

**Глава IV. ПРАКТИЧЕСКАЯ ЧАСТЬ**

Чтобы лучше понять принципы действия систем МП и их методы использования словарей и анализа грамматики, равно как и синтеза структур на выходном языке, следует на практике перевести несколько текстов (желательно различных по функциональному стилю и тематике), используя одну из вышеописанных систем машинного перевода. Наиболее целесообразным представляется использование системы МП PROMT XT, поскольку она является самой последней на данный момент версией ряда продуктов PROMT и объективно лучшей из доступных.

Возьмем в качестве первого примера следующую логическую задачку:

*"You are given 12 identical-looking coins, one of which is counterfeit and weighs slightly more or less (you don't know which) than the others. You are given a beam balance which lets you put the same number of coins on each side and observe which side (if either) is heavier. How can you identify the counterfeit and tell whether it is heavy or light, in 3 weighings?"*

Вот ее 'вольный' перевод на русский язык, который мог бы быть сделан человеком - переводчиком:

*"У вас есть 12 одинаковых по виду монет, одна из которых - фальшивая и весит немного больше или меньше, чем остальные (вы не знаете, какая именно). Имеются рычажные весы, на чаши которых вы можете класть равное число монет и смотреть, какая из чаш перевесила (или весы остались в равновесии). Как за 3 взвешивания определить фальшивую монету и узнать, легче она или тяжелее остальных?"*

При переводе переводчику пришлось поменять порядок слов в нескольких предложениях. А вот как переводит этот текст система автоматического перевода семейства PROMT (использовалась PROMT XT):

*"Вам дают 12 идентично-выглядящих монет, одна из которых - подделка и весит немного более или менее (Вы не знаете который) чем другие. Вам дают баланс луча, который позволяет Вам помещать то же самое число {номер} монет на каждой стороне и наблюдать {соблюдать}, какая сторона (если любой) более тяжел. Как Вы можете идентифицировать подделку и сказать, тяжело ли это или легко, в 3 взвешиваниях?"*

Обратим внимание на "баланс луча". Эта ошибка вызвана, как легко понять, отсутствием в словаре словосочетания "beam balance", означающего "рычажные весы". Очевидно, что варианты перевода слов (данные в фигурных скобках) - вторые значения соответственных слов во входном языке. Таким образом, отбирая, исходя из контекста, наиболее вероятные значения слов, программа PROMT иногда затрудняется выдать однозначный вариант, оставляя, таким образом, право выбора за редактором-переводчиком. Однако справедливости ради, нужно отметить, что в обоих случаях двоякого толкования слова программой в данном тексте, в качестве наиболее вероятных были выбраны именно правильные значения.

Немного смущает фраза "если любой", употребленная в тексте в скобках. Во-первых, очевидно, что грамматически текст данный в скобках никак программой не связывается с текстом за скобками - отсюда и разница в роде: сущ. *сторона* - ж.р. ед. ч.; и мест. *любой* - м.р. ед. ч. Во-вторых, вероятно фраза "if any" представляет собой одно единое семантическое целое - в своем роде клише официального стиля. Доказательством тому служит присутствие выражения "*if any*" в словаре ABBY Lingvo7.0 в качестве отдельной словарной статьи:

"***if any*** - **если это имеет место**

At the start of every month I have to send him an account of my earnings, if any. — В начале каждого месяца я должен посылать ему отчет о моих заработках, если таковые имелись."

Поэтому упущением создателей системы PROMT является пословный перевод этой фразеологической единицы. Следующим существенным недостатком является очевидное 'нежелание' системы менять порядок слов в предложении - иногда это жизненно необходимо. То есть в выходном языке порядок слов почти всегда такой же, как и во входном.

Общеизвестно, что в английские существительные утратили грамматическую категорию рода. Поэтому почти все они согласуются с местоимением 3 лица ед. числа "it" - среднего рода. В русском же языке категория рода у существительных присутствует и в 3-м лице ед. числе они согласуются с тремя разными местоимениями в зависимости от грамматического рода. Так вот, система PROMT, синтезируя текст на выходном языке не учитывает возможность согласования существительного и местоимения, его заменяющего в роде. Примером может служить последнее предложение текста: "How can you identify the counterfeit and tell whether it is heavy or light, in 3 weighings?". Переведено оно следующим образом: "Как Вы можете идентифицировать подделку и сказать, тяжело ли это или легко, в 3 взвешиваниях?" Перевод слов "It" и "counterfeit" не согласуются в роде: подделка - женского рода, это - среднего рода; более того - не видя связи между двумя этими словами, программа-переводчик вместо личного местоимения подставляет указательное.

В качестве второго примера возьмем начало Декларации Независимости США. Здесь мы увидим перевод программой PROMT XT текста, написанного в официально-деловом стиле в конце XVIII века. Вот оригинал текста:

"*When in the course of human events, it becomes necessary for one people to dissolve the political bands which have connected them with another, and to assume among the powers of the earth, the separate and equal station to which the laws of nature and of nature's God entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the separation. We hold these truths to be self-evident, that all men are created equal, that they are endowed by their creator with certain unalienable rights,  that among these are life, liberty and the pursuit of happiness, that to secure these rights, governments are instituted among men, deriving their just powers from the consent of the governed...*"

Как мы видим, этот текст далеко не прост с точки зрения восприятия: предложения длинные, много придаточных предложений и однородных членов. С другой стороны, несмотря на то, что тексту более 225 лет, в нем нет слов или грамматических конструкций, которые были бы непонятны человеку, знающему английский язык хотя бы на уровне носителя. В тесте не используются слова, значения, которых устарели бы и поэтому единственной сложностью является большой объем предложений и относительная сложность их конструкции. Далее следует перевод этого текста, сделанный человеком переводчиком:

"*Когда в ходе человеческой истории для одного народа становится необходимым порвать политические узы, связывающие его с другим, и занять среди наций мира самостоятельное и равное положение, присвоенное ему в силу естественного права и законов Божественной Природы, — уважение к мнениям человечества требует декларации причин, вынуждающих его к отделению. Мы считаем самоочевидными следующие истины: что все люди созданы равными, что они наделены Создателем определенными неотъемлемыми правами, среди которых имеется право на жизнь, свободу и на стремление к счастью; что для обеспечения этих прав существуют среди людей правительства, осуществляющие свою власть с согласия тех, кем они управляют…*"

А вот перевод этого текста, осуществленный системой PROMT XT:

"*Когда в ходе человеческих событий, это становится необходимым для одного людей расторгнуть {распустить} политические полосы{оркестры}, которые соединили их с другим, и принимать среди полномочий земли, отдельная и равная станция, на которую законы характера{природы} и Бога характера{природы} дают право им, приличное уважение{отношение} к мнениям относительно человечества требует, чтобы они объявили причины, которые побуждают их к разделению. Мы считаем эти истины быть самоочевидными, что все мужчины созданы равными, что они обеспечены их создателем с некоторыми неотъемлемыми правами, которые среди них являются жизнью, свободой и преследованием счастья, что, чтобы обеспечить эти права, правительства назначены{установлены} среди мужчин, получая их справедливые полномочия от согласия управляемых..."*

Как мы видим, с таким текстом у системы машинного перевода возникает больше проблем. Здесь можно даже не упоминать выбора неправильных значений слов, таких как *People, bands, station*. Основываясь на данном образце перевода текста, можно воочию, так сказать, увидеть все несовершенство механизмов грамматического анализа и синтеза.

В отношении грамматического синтеза, с другой стороны, сказать можно не так много. Ну, например, выходной вариант "*дают право им*" не соответствует норме языка с точки зрения порядка слов. Таким образом, при синтезе грамматических структур на выходном языке, мы опять-таки видим жесткую привязку к порядку слов в тексте на входном языке. Затем, такая фраза как "*мы считаем эти истины быть самоочевидными…*" сильно напоминает речь иностранца, 'изучившего' русский язык посредством разговорника. А ведь программа писалась русскими специалистами! Однако в данном случае, лингвистов, участвовавших в создании систем PROMT в незнании грамматики родного языка обвинить нельзя. Ведь проблема здесь не в синтезе грамматической структуры, а все в том же непонимании структуры входного языка - то есть в грамматическом анализе.

Грамматический анализ, как мы видим, находится на самом примитивном уровне. Да, простые предложения система переводит (да и то, как мы убедились - не все) почти без ошибок. Сложносочиненные, да и классические примеры сложноподчиненных предложений тоже даются системе с относительной легкостью. Однако как только возникает нестандартная ситуация (например, одно придаточное предложение усложняется другим (или даже элементарное вводной или пояснительной конструкцией) и, как следствие, разрывается) и программа не находит подходящего алгоритма грамматического анализа - она сразу забывает о синтаксисе и начинает элементарный пословный перевод, формально (посредством флексий) пытаясь связать хотя бы рядом стоящие слова. Эта попытка связать грамматически рядом стоящие слова вкупе неправильным выбором значений некоторых слов еще более запутывает выходной вариант.

Язык - это живая структура, которая не поддается полной алгоритмизации и, следовательно, посредством одних лишь алгоритмов проблему машинного перевода не решить. Машина не понимает текст, она лишь преобразовывает его посредством различных алгоритмов и правил. И не важно, сколько будет этих правил, без хотя бы общего понимания входного текста не может быть сколько-нибудь связного и стабильного процесса перевода. На уровне простых предложений и в рамках строго определенной тематики машинный перевод в принципе возможен, но не более.

**Заключение.**

До уровня полной автоматизации перевода человечество еще не дошло, да и дойдет, вероятно, не скоро. Причиной этому является, вероятно, недостаточный уровень развития наук, затронутых в создании подобных систем. Слишком сложно сказать, как человек переводит - а тем более сложно смоделировать этот процесс с помощью компьютерной программы. Тем более сложно сделать это, если учесть, что человек мыслит образами, а научить этому компьютер - невозможно в принципе (по крайней мере, на настоящем уровне развития ЭВМ).

Возьмем, к примеру, неопределенный артикль "a". Если человеку с определенным багажом лингвистических знаний сказать фразу "*неопределенный артикль "а*", в его сознании моментально возникает несколько образов - начиная от звуковой формы данного артикля и заканчивая образом неопределенности, каким бы у данного человека этот образ ни был. Однако даже для самой современной компьютерной системы фраза "*неопределенный артикль "а*" обозначает лишь последовательность из двухсот-восьми единиц и нулей, составляющих бинарный эквивалент литерной величины "*неопределенный артикль "а*". Поэтому-то научить ЭВМ самостоятельно осуществлять адекватный перевод текстов в принципе невозможно на данном этапе развития. Язык образен и не поддается полной алгоритмизации, а посему проблема полной автоматизации перевода сводится следующей проблеме: научить машину мыслить и оперировать образами - а эта проблема уже из области проблем искусственного интеллекта, создание которого все еще является чем-то из области фантастики.

Другое дело, что уже теперь мы можем использовать достижения науки и техники для облегчения работы человека во всех сферах его деятельности. Конечно, применимость ЭВМ может быть где-то более актуальна, а где-то - менее. Тем не менее, ЭВМ применимы везде, более того - уровень их применимости постоянно растет. Справедливо это и для автоматизации процесса перевода. Если машины и не могут пока осуществлять адекватный перевод самостоятельно, то служить серьезным подспорьем для переводчика они вполне в состоянии. При их грамотном использовании эффективность перевода может возрасти в несколько раз, причем качество перевода не снизится, а наоборот - повысится (взять, к примеру, те же системы Translation Memory).

Таким образом, говоря о наиболее перспективных путях развития систем автоматизации перевода, следует, вероятно, сосредоточиться на том, что выполнимо на данный момент, то есть на создании более эффективных электронных словарей с как можно более эффективным механизмом поиска и индексации, с как можно более интегрированной системой словарных статей. Если же брать во внимание развитие систем Машинного Перевода, то наиболее перспективным направлением здесь окажется совершенствование подсистем грамматического анализа и синтеза, а также увеличение объема контекстуального охвата текста и совершенствование семантических цепочек с целью более точного подбора значений слов.

**Библиография**

1. http://courier.com.ru/nauka/diction1.htm#up ("Что внутри электронного словаря?");
2. http://mcbsys.com/;
3. http://osp.admin.tomsk.ru/school/1999/2/06.htm;
4. http://school.ort.spb.ru/library/informatica/compmarket/internet/transl.htm ("Интернет: кое-что в помощь переводчику", Ильдар Кутыев);
5. http://www.a-z.ru/person/belonogov/index.htm#I0
6. http://www.computerra.ru/ ("Что могут словари?" Денис Зельцер);
7. http://www.krugosvet.ru/articles/82/1008256/1008256a1.htm;
8. http://www.lingvoda.ru/transforum/articles/pdf/selegey\_a1.pdf ("Электронные словари и компьютерная лексикография", Владимир Селегей, компания ABBY);
9. http://www.mtoday.com/article\_m.htm (Анна Марченко, центр переводов "Гильдия");
10. http://www.promt.ru/mtw/articles/article\_Sokolova.phtml ("Как переводит компьютер", Автор:Соколова Светлана);
11. http://www.promt.ru:8000/mtw/class.phtml;
12. http://www.promt.ru:8000/mtw/developer.phtml;
13. http://www.svoboda.org/programs/sc/2001/sc.050101.asp (Александр Костинский);
14. www.multilex.ru;
15. А. САВИНА, Т. ТИПИКИНА («Наука и жизнь» 1999, N 9);
16. Владимир Самусенко, "Электронный словарь - друг человека" 24.02.1999, Компьютер в школе, #2/1999;
17. Г. Г. Белоногов, Ю. Г. Зеленков,"Б. А. Кузнецов, А. П. Новоселов, Александрдр А. Хорошилов, Алексей А. Хорошилов. Автоматизация составления и ведения словарей для систем фразеологического машинного перевода текстов с русского языка на английский и с английского на русский. Сб. "Научно-техническая информация", Сер. 2, | 12, ВИНИТИ, 1993;
18. Г. Г. Белоногов, Ю. Г. Зеленков,"Б. А. Кузнецов, А. П. Новоселов, Н. А. Пащенко, Александрдр А. Хорошилов, Алексей А. Хорошилов. Интерактивная система русско-английского и англо-русского машинного перевода политематических научно-технических текстов. Сб. "Научно-техническая информация", Серия 2, | 3, ВИНИТИ, 1993.
19. З.М.Шаляпина "АВТОМАТИЧЕСКИЙ ПЕРЕВОД: ЭВОЛЮЦИЯ И СОВРЕМЕННЫЕ ТЕНДЕНЦИИ" (Вопросы языкознания, 1996, 2, с.105-117);
20. Сергеев В.Н. Словари – наши друзья и помощники. – М.: Просвещение, 1984.