

Н. ПЕР.
 000000 00
 104E4B 00
 701131 00
 020105 00
 000100 00
 73211E 00
 5B3503 00
 000101 00
 330F6A 04
 410372 02
 7E1809 00
 0103DE 00
 000000 00
 505A09 00
 010117 00
 010305 00
 01010E 00
 3C1437 04
 000001 00
 7E1809 00
 4101DD 02
 000001 00
 105A99 00
 505D05 00
 000001 00
 5B7503 00
 0103DE 00
 000000 00
 033E67 00
 3C1437 00
 380124 05
 020117 00
 000001 00
 0B13B0 00
 0C13F8 00
 502406 00
 270890 00
 000001 00
 33261C 04
 470000 00
 5B3F2D 00
 731008 00
 505D05 00
 0103DE 00
 000000 00
 507709 00
 000000 00
 732009 00
 34288C 00
 5B393C 00
 000001 00
 3C1C05 04
 4101DD 02
 104E4B 00
 701131 00
 0103DE 00
 000000 00
 510DC8 00
 000100 00

FULL
 DUPLEX
 AND
 20
 MILLIAMPERE
 OPERATION
 ARE

ПОЛНЬЯ
 ДУПЛЕКС
 И
 20
 МИЛЛИАМПЕРНАЯ
 ОПЕРАЦИЯ

Ю.Н.Марчук

Проблемы машинного перевода

Издательство Наука

AN
 INTERNAL
 CURRENT
 SOURCE
 CAN
 BE
 USED
 IN
 HALF
 DUPLEX
 MODE

ВНУТРЕННИЙ
 ТЕКУЩИЙ
 ИСТОЧНИК
 МОЖЕТ
 БЫТЬ ИСПОЛЬЗОВАН
 В
 ПОЛОВИНЕ
 ДУПЛЕКСНОЙ
 РЕЖИМЕ

FOUR

ЧЕТЫРЕ

WIRE
 TWISTED
 PAIRS
 ARE
 RECOMMENDED
 FOR
 FULL
 DUPLEX

ПРОВОДНЫЕ
 СКРУЧЕННЫЕ
 ПАРЫ
 РЕКОМЕНДУЮТСЯ
 ДЛЯ
 ПОЛНОГО
 ДУПЛЕКСА

8

8

ГОСУДАРСТВЕННЫЙ КОМИТЕТ СССР
ПО НАУКЕ И ТЕХНИКЕ

АКАДЕМИЯ НАУК СССР

Всесоюзный центр переводов
научно-технической литературы и документации

Ю.Н.Марчук

Проблемы машинного перевода



Издательство «Наука»

Москва 1983

В монографии рассматривается современное состояние исследований в области машинного перевода, дается анализ существующих подходов к решению этой актуальной проблемы научно-технической информации и описываются принципы моделирования перевода на основе переводных соответствий. Построена система машинного перевода с английского языка на русский для достаточно широкого спектра научно-технических подязыков. Двухкомпонентная модель перевода на основе переводных соответствий, отражающая статику описания языка и динамику переводческого процесса, позволяет в режиме работы с человеком-редактором добиваться повышения качества собственно машинного перевода посредством пополнения словарей системы и совершенствования алгоритмов анализа и синтеза текстов. Лингвистическое обеспечение современных систем машинного перевода создается на основе теории «человеческого» перевода, принципов и особенностей научно-технического стиля речи, контрастивного языкознания, искусственного интеллекта.

Ответственный редактор

доктор филологических наук
Р. Г. ПИОТРОВСКИЙ

Юрий Николаевич Марчук
ПРОБЛЕМЫ МАШИННОГО ПЕРЕВОДА

Утверждено к печати
Научно-техническим Советом Всесоюзного центра переводов
научно-технической литературы и документации ГКНТ СССР и АН СССР

Редактор издательства Т. М. Дривинг
Художник А. А. Кущенко. Художественный редактор Т. П. Поленова
Технический редактор А. М. Сатарова. Корректор К. П. Лосева

ИБ № 26732

Сдано в набор 12.10.82. Подписано к печати 10.03.83. Формат 60×90^{1/8}. Бумага типографская № 1. Гарнитура литературная. Печать высокая. Усл. печ. л. 14,5. Усл. кр. отт. 14,7. Уч.-изд. л. 16,4. Тираж 2600 экз. Тип. зак. 4323. Цена 1р. 60 к.

Издательство «Наука» 117864 ГСП-7, Москва, В-485, Профсоюзная ул., 90
2-я типография издательства «Наука» 121099, Москва, Г-99, Шубинский пер., 10

М 4602020000—128
042(02)—83 379—83—II © Издательство «Наука»,
1983 г.

ПРЕДИСЛОВИЕ

Научно-технический прогресс в значительной мере зависит от распространения научно-технической информации. Важную ее часть составляют научно-технические переводы, объем которых в мире неуклонно возрастает в связи с расширением потребностей научно-технического общения. В начале 70-х годов продукция Всесоюзного Центра переводов научно-технической литературы и документации исчислялась в пять — шесть тыс. авторских листов переводов за год; в конце 70-х годов переводились уже несколько десятков тысяч, а в пределах СССР — сотни тысяч авторских листов, причем спрос на переводы примерно в два раза превосходит выполняемый объем. В Европейском экономическом сообществе шесть официальных языков¹; это означает, что каждый документ или научно-техническое описание должны быть переведены на пять других языков сообщества. Служба переводов этой организации в 1978 г. состояла из 1300 лингвистов, которые за год переводили примерно 600 000 страниц. В 1979 г. число переводчиков было увеличено до 2000 человек, однако и они не справляются с возросшим потоком переводов. Преодоление языковых барьеров становится одной из главных задач распространения научно-технической информации. Идея создания всемирного языка для общения в науке и технике оценивается руководителями информационных служб как неперспективная, поэтому в поисках средств ускорения и удешевления переводов все чаще обращаются к автоматизации процесса и созданию систем машинного перевода (МП).

В рамках общей теории перевода возникло и за последнее время достаточно четко определилось направление научно-технического перевода. Можно спорить о том, является ли научно-технический перевод самостоятельной областью или самостоятельной теорией в переводоведении, какова степень этой самостоятельности и пр. Однако бесспорно то, что перевод научно-технической литературы и документации характеризуется особенностями, существенно отличающими его от перевода других видов литературы. Как известно, общая теория перевода занимается базовыми закономерностями перевода как процесса преобразования текстов на одном языке в тексты на другом языке при сохранении неизменным содержания этих текстов; частные теории перевода рассматривают проблемы перевода между двумя конкретными языками, парами языков. Вопросы типологии видов перевода исследованы очень мало, хотя в условиях возрастаю-

¹ По состоянию на 1980 г.; вступление новых стран в ЕЭС увеличивает число официальных языков и усугубляет проблему перевода.

щего переводческого общения эта типология приобретает не только и не столько теоретическое, сколько практическое значение: вид перевода связывается, например, с его местом в информационном обслуживании, с оплатой и пр. В связи с этим вопросы научно-технического перевода также выделились за последнее время в самодовлеющий комплекс, который сейчас получает теоретическое обоснование. Непосредственное отношение к научно-техническому переводу имеет машинный перевод.

МП к настоящему времени насчитывает почти тридцать лет развития. Теоретические исследования в широко понимаемой проблеме МП, с одной стороны, а с другой — экспериментирование на ЭВМ по обработке больших массивов текста в рамках разнообразных лингвистических задач, а также собственно в целях получения МП как промышленно пригодного продукта способствовали формированию новой точки зрения на МП в ином социально-технологическом окружении. Потребность в ускорении и увеличении объема переводов, в условиях возросших и принципиально новых возможностей кибернетики и вычислительной техники, а также становление науки о научно-технической коммуникации породили принципиально иной и закономерно принадлежащий МП статус в первую очередь инженерно-лингвистической и прикладной науки, в которой приобретают особое значение инженерно-технологические решения и технико-экономическая эффективность, а также удобства работы в рамках человеко-машинной системы.

Полностью автоматизированный высококачественный машинный перевод следует, видимо, считать некоторым идеалом, подобно многим абстракциям науки (идеальный газ, абсолютно твердое тело и т. д.). Нет необходимости доказывать, что недостижимость такого идеала не должна служить препятствием к попыткам добиться реально пригодного МП. Ближайшей аналогией здесь является положение с «человеческим» переводом: неизбежность потерь информации при переводе, препятствующая получению абсолютно адекватного полностью эквивалентного перевода, не могла и не может служить основанием для отказа от перевода, и профессия переводчика является одной из древнейших.

Общие задачи исследования могут быть определены следующим образом:

— исследовать теоретические подходы к моделированию перевода и определить принципиальные позиции, с которых целесообразно подходить к такому моделированию в современных технологических условиях переработки информации;

— создать лингвистическую модель машинного перевода с использованием уровня переводных соответствий;

— построить действующую систему для конкретной пары языков, для чего определить конкретное «наполнение» теоретических компонентов модели;

— изучить основные особенности функционирования системы на текстах и закономерности, полученные в ходе работы;

— разработать принципы реализации технических решений в рамках общей концепции приближенных лингвистических вычислений;

— оценить технико-экономическую эффективность полученной модели в ее промышленной реализации в заданных конкретных условиях эксплуатации;

— определить таким образом, т. е. конкретным решением вопроса, основные этапы построения системы машинного перевода и круг связанных с этим теоретических и практических проблем.

Общая методика исследования характеризуется сочетанием дедуктивного и индуктивного подходов. Частные методики включают использование статистических характеристик, методов экспертных оценок и другие. Объекты исследования — английские тексты и их переводы на русский язык, словари, действия человека-переводчика, действующая система МП, результаты ее работы. Специальным объектом исследования выступает созданный в рамках принятой концепции англо-русский контекстологический словарь для машинного перевода многозначных слов с английского языка на русский. Используются также статистические характеристики распределения слов и словоформ, количественные данные и показатели частотности лингвистических явлений.

Основным теоретическим положением, идейной позицией автора служит точка зрения о том, что для построения системы МП, имеющей практическое значение, необходимы специальные модели, комплексный подход к самой проблеме МП как проблеме научно-технической информации и информационного обслуживания в современных технологических условиях. Проведенные исследования и построение системы показывают:

— машинный перевод приемлемого для информационного обслуживания качества может быть построен на основе модели машинного перевода по переводным соответствиям, состоящей из двух компонентов: предметного и динамического. Такая модель обеспечивает оптимальное с точки зрения лингвистической технологии решение проблемы независимости лингвистического описания и алгоритма. Возникает возможность совершенствовать систему по результатам обработки нового материала. Создание системы, позволяющей повышать качество перевода, является принципиально новым теоретическим и практическим моментом;

— система переводных соответствий, будучи реализована в рамках указанной модели, дает возможность обеспечить приемлемое качество МП еще до этапа редактирования;

— МП, допускающий постепенное улучшение качества и соответствующее уменьшение усилий редактора, осуществляется не только благодаря пополнению словарей, алгоритмов и этапов анализа и синтеза, но и качественной настройке лингвистического содержания алгоритмов на специфику подязыка, в результа-

те чего создается ориентированная на перевод совокупность семантических признаков и соответствующих алгоритмов для данного подязыка;

— в процессе адаптации системы МП к конкретному подязыку семантический аппарат, наличие которого в значительной мере определяет возможности системы в аспекте постепенного повышения качества перевода до редактирования, создается развитием от исходной системы семантических категорий, необходимых для получения релевантной собственно переводу совокупности лексических и грамматических характеристик, через расширение системы трансформационных преобразований посредством введения новых семантических параметров, признаков или отношений;

— при моделировании перевода с помощью переводных соответствий центр разрешения лексической неоднозначности — одной из важнейших проблем при МП — переносится на лексический уровень (не на семантический или синтактико-семантический, как это имеет место в других способах моделирования); при этом создается возможность оптимизировать алгоритмическое выявление контекстных зависимостей, с помощью которых автоматизируется перевод многозначных слов;

— машинный словарь приобретает особое значение при использовании приближенных методов в МП. Вследствие того, что применение машинных словарей характеризует не только МП, но и информатику (научно-техническую информацию) в целом, и того, что проблема МП является одной из сложнейших в автоматической обработке текстов, изучение характеристик машинных словарей имеет не только практическое, но и теоретическое значение. Представляется возможным говорить о новой языковедческой науке — вычислительной лексикографии.

Данная работа по своему научному направлению относится к прикладной лингвистике, области автоматической обработки текстов. Представляется, что в отличие от других наук, где четко определено инженерное направление и где сложились критерии технологической оценки научного и практического вклада, в технических приложениях прикладной лингвистики, в вычислительной (инженерной) лингвистике еще нет системы точного определения того, что именно относится к ней, а не к структурному языкознанию, например. Часто к работам такого содержания, т. е. направленным на конкретное практическое решение точно сформулированных проблем, предъявляются требования создания принципиально новых теорий, новых точек зрения на абстрактный предмет и т. п. Не отрицая значения теории, следует сказать, что именно в этой области прикладной лингвистики наиболее критическим оказывается отсутствие практики и широких машинных экспериментов. Хорошо известно, что язык как многообразное и всестороннее явление человеческой деятельности представляет неисчерпаемую почву для самых различных обобщений, и критерием истинности того или иного теоретического

обобщения становится практический эксперимент и пригодность для анализа достаточно широкого круга явлений. Поэтому приоритет в описываемой нами области принадлежит на данном этапе развития не самим новым теориям, а вопросам, связанным с технологией и инженерными решениями теоретических и практических задач.

В этом рассуждении можно провести некоторую аналогию. Работа инженера-строителя или конструктора оценивается не по тому, внесли ли они новые идеи в физику или сопротивление материалов, хотя не исключено, что их вклад может быть именно таковым. Главные параметры оценки работы инженеров заключаются в том, насколько разработанные ими конструкции технологичны, экономичны, устойчивы, обладают существенными для данной задачи или данной группы задач инженерными характеристиками. Работа по химической технологии, например, по способам получения новых (или старых, известных) химических соединений не связана с открытием новых химических веществ, однако она может иметь большое научное и практическое значение.

Автор видел цель своей работы в описании опыта создания лингвистической технологии практически приемлемого решения сложнейшей прикладной лингвистической задачи, на пути к которому, естественно, пришлось произвести выбор теоретического обоснования исследования, однако лишь в пределах выбора рабочих (прикладной) модели. Подобно тому, как инженер-строитель в решении конкретной задачи — например, в возведении уникального моста — имеет дело с данными и методами целого ряда других наук, например с теорией механизмов и машин, гидравликой, геологией, теорией фильтрации, и пр., автор неизбежно касался сложных лингвистических проблем, таких как функционирование языкового знака в ЭВМ, проблема контекста, теория перевода, грамматические учения и т. д. По каждой из этих дисциплин в рамках поставленной проблемы возможны и необходимы самодовлеющие исследования глубокого и специфического для них плана. Однако в прикладных целях допустимо действовать подобно инженеру, который руководствуется данными справочников и новейшими данными смежных теоретических и прикладных наук, проверяя их только практическим расчетом и экспериментом, без углубления в теорию. Здесь приходится сожалеть, что лингвистическая теория в части автоматической обработки текстов не располагает на сегодняшний день такими выводами, которые могли бы составить содержание справочников.

Практическая часть работы содержит описание разработанной под руководством автора во Всесоюзном Центре переводов научно-технической литературы и документации ГКНТ и АН СССР (ВЦП) системы машинного перевода с английского языка на русский АМПАР (автоматизированный машинный перевод с английского на русский) текстов по вычислительной технике и

О СОВРЕМЕННОМ СОСТОЯНИИ МАШИННОГО ПЕРЕВОДА

программированию. Эта система создавалась и совершенствовалась в течение многих лет. В 1979 г., впервые в СССР, система была принята межведомственной комиссией и с 1980 г. находится в опытно-промышленной эксплуатации в ВЦП, снабжая потребителей отредактированным машинным переводом.

На ранних этапах создания большой вклад в работу осуществил Ю. А. Мстюрин, руководивший организационной и программной частью работ. В дальнейшем программная часть последовательно разрабатывалась под руководством канд. техн. наук Б. Д. Тихомирова и других товарищей. Активное участие в создании контекстологического словаря приняли Е. Е. Ловцкий и В. И. Щербинин. Большой вклад в разработку и отладку алгоритмов и словарей системы внесли А. Н. Киселевым, А. С. Ежовым, Л. В. Драмбян и другими сотрудниками лаборатории машинного перевода ВЦП.

Работа над созданием практически пригодных систем МП продолжается. В ВЦП вступила в строй первая очередь системы немецко-русского машинного перевода НЕРПА, которая создана по тем же принципам, что и АМПАР. Создаются усовершенствованные варианты АМПАРА для других подязыков и новых режимов работы. Автор ни в коей мере не претендует на исключительность и единственную правильность обоснованного в настоящем исследовании подхода. В ВЦП разрабатывается и уже сдана первая очередь системы французско-русского машинного перевода ФРАП, научным руководителем которой является Н. Н. Леонтьева. Эта система строится на принципиально другой основе, чем АМПАР, поэтому после ее завершения возникнут объективные основания для сравнения двух разных подходов по технико-экономической и качественной эффективности. Сейчас, в начале эры практической реализации МП, важно учесть все научные теоретические и практические достижения, весь накопленный опыт, чтобы двигаться оптимальным путем к эффективному решению этой актуальной проблемы.

1.1. Два основных подхода к решению проблемы

Развитие машинного перевода и его современную картину можно представить себе как диалектическое взаимодействие, борьбу двух основных направлений, двух главных подходов к проблеме. Эти два подхода определяются тем, что берется в основу системы описания входного языка для конечной цели перехода к выходному естественному языку в условиях сохранения содержания, смысла переводимого текста, какова система промежуточного между двумя естественными языками уровня преобразования знаков.

Первый подход (по фактическому преобладанию в нынешнем состоянии теории МП) характеризуется установкой на использование максимально мощного универсального языка смысла. Главное содержание этого подхода — примат абстрактной экспликации смысла над конкретными языковыми или речевыми выражениями этого смысла в естественных языках. Вообще говоря, термин «машинный перевод» применительно к этому подходу теряет собственное содержание и практически означает многочисленные работы, целью которых является изучение способов формального описания семантики и использование этого описания для различных прикладных целей. Применительно к переводу вопрос ставится так: надо сначала научиться выражать смысл в некотором искусственном языке, а затем, что уже дело техники, этот смысл можно представить в лексике и грамматике любого естественного языка, тогда проблема перевода по существу снимается. К этому подходу можно отнести метод «семантических множителей», модели с универсальным языком-посредником в виде разных «языков смысла», тезаурусный метод, метод фильтров, «анализ через синтез», модель «смысл—текст» и пр. Конечно, каждый из названных методов МП имеет свою специфику, однако общее в подходе преобладает над отдельными частными отличиями или приемами. Общим является принципиально дедуктивный подход к созданию теории МП, моделирование человеческого языка в целом, во всех его основных функциях, главным образом, в логической, стремление к универсальным решениям и применениям (машинный перевод, информационный поиск, искусственный интеллект). В том, что касается универсального языка—посредника или «языков смысла», нетрудно проследить непреходящее стремление человека избавиться от «нелогичности», «непоследовательности» и вообще от множест-

венности естественных языков и учредить единый для всего человечества язык логики и универсальных понятий.

В машинном переводе в собственном смысле этот подход приводит к так называемым «тотальным стратегиям» (удачный термин, введенный болгарским ученым А. Людскановым). Общей характеристикой тотальных стратегий является стремление получить полностью автоматизированный высококачественный машинный перевод максимальным использованием семантического уровня языка, причем достигается это благодаря последовательному осуществлению принципов независимости анализа и синтеза, четкого разделения других уровней (морфологический, синтаксический, семантический), многовариантности на всех уровнях, получения полной синтаксической структуры, перехода к высшему уровню через многовариантность нижних. Составление лингвистического описания — грамматики — есть дело лингвистов, составление алгоритма оперирования с этими данными для получения перевода есть дело математиков¹.

Если оценивать тотальные стратегии с методологической точки зрения, то очевидно, что созданию системы машинного перевода должна предшествовать разработка достаточно полной и детальной теории языка в целом (о чем, собственно, и говорили в явной форме сторонники этих стратегий). Не касаясь вопроса о принципиальной возможности такой работы, можно отметить крайнюю трудоемкость создания цельных систем МП на этом пути. Этим, видимо, и объясняется то обстоятельство, что несмотря на длительность существования этого подхода, до сих пор нет ни одной действующей системы МП, построенной по указанным принципам для достаточно широкого материала.

Оценивая этот подход в целом с точки зрения современного состояния, следует отметить, что он внес значительный вклад как в теорию, так и (в меньшей степени) в практику МП, главным образом своими детальными разработками многих вопросов анализа естественного языка, а также синтеза. Отсутствие действующих систем МП в рамках этого подхода можно, на наш взгляд, объяснить тем, что прямолинейное построение их на принципах тотальных стратегий не учитывает по крайней мере двух важных моментов:

— абсолютная независимость анализа от синтеза и многовариантность на всех уровнях не дают возможности смоделировать переводческий процесс удовлетворительным образом;

— поручение составления алгоритма перевода математикам вообще бессмысленно. Процесс перевода с одного естественного языка на другой с его селективным использованием семантики, отсеиванием вариантов на ранних стадиях, рекурсиями и проверками, наконец, своей связью с творческим мышлением не может

¹ Крайняя точка зрения сторонников этого направления: современный АП (автоматический перевод) есть «перевод без перевода, без машины, без алгоритмов» (Автоматический перевод 1967). Обоснованная критика экстремизма в этом направлении содержится в статье (Котов 1976).

быть игнорирован и заслуживает собственного моделирования именно как процесс.

Исследователи, придерживающиеся первого подхода, ввели в оборот представление о «поколениях» систем МП, каждое из которых, якобы, характеризуется именно наличием признаков абсолютной «независимости» анализа и синтеза, расчлененности, многовариантности на всех уровнях, использованием семантики. Однако аналогия с другими продуктами, имеющими поколения, например, поколения ЭВМ, здесь никак не выдерживается, поскольку никаких систем МП названных «поколений», кроме «первого» или «нулевого», просто нет, поэтому нет показателей, которые можно было бы сравнивать в терминах поколений. Некритическое отношение к «поколениям» систем МП может ввести читателя в заблуждение.

Второй подход хронологически возник раньше первого. Пионерами МП были математики и инженеры; теперь, когда мы оцениваем высказанные в то время соображения, мы должны признать, что интуиция и здравый смысл дали возможность уже в то время сформулировать важные индуктивные концепции. Этот подход базируется на таком представлении о промежуточном языке, которое близко подходит к идее переводных соответствий на чисто языковом уровне. Методическая суть этого подхода: изучение индивидуального поведения языковых единиц, особенно в языковом контексте; моделирование человеческого владения языком, особенно в процессе межъязыкового перевода; переход от простого, легко формализуемого, находящегося на поверхности, к более сложному, формализация которого достигается постепенно, последовательным движением от простого. Теоретические представления, образующие теорию МП, в рамках этого подхода создавались индуктивным путем, методом проб и ошибок. Как и в рамках первого направления, целый ряд исследований вышел за пределы МП: сформировались теоретико-информационные и статистические измерения языка, количественные методы изучения синтаксиса, смысла и семантики и пр. Однако в том, что касается собственно МП, этот подход привел к формулированию принципов так называемых селективных стратегий (термин А. Людсканова), идея которых заключается в том, что для перевода нужно брать только то, что для него нужно: излишнее «заглубление» в смысл, выход на все уровни анализа даже для простых случаев нерациональны.

Уже в 50-х годах на принципах селективных стратегий были построены простые, но достаточно эффективные системы МП, которые действуют и до сих пор (см. приложение). Отказ от попытки получить сразу полностью автоматизированный высококачественный машинный перевод (долгое время целью разработок и исследований по МП был именно такой перевод) и признание необходимости человека-редактора подняли практическое и теоретическое значение этих систем и вообще селективных стратегий МП. Однако даже в условиях взаимодействия с человеком

пре- или постредактором качество собственно МП должно быть достаточно высоким, иначе всякая экономическая эффективность (для чего и создается МП) исчезает. Поэтому в настоящее время принципы, по которым были созданы первые системы МП, должны быть пересмотрены и пополнены на основе достигнутого за эти десятилетия опыта и теоретических результатов.

Причины низкого качества МП в рамках систем второго направления заключаются, как нам представляется, в неоптимальной организации словаря и работы с ним, отсутствии достаточно эффективных алгоритмов семантико-синтаксического анализа, недостаточном использовании семантических сведений, неудобном программном обеспечении, которое не дает возможности, в ряде случаев, оперативно пополнять словари и алгоритмы, что необходимо для повышения качества в режиме работы с редактором.

Особо следует сказать об использовании семантики. Ученые первого направления критикуют сторонников селективных стратегий в МП за недостаточное использование семантики. На самом деле семантические сведения широко используются уже в ранних алгоритмах МП в виде семантико-синтаксических классов и подклассов слов, семантических кодов, признаков и пр. Другое дело, что этап использования семантики не выделяется в самостоятельный, как это предусмотрено в тотальных стратегиях. Однако на наш взгляд, вопрос о том, следует ли обязательно иметь четко выделенный семантический (а равно и синтаксический, морфологический и т. п.) уровень, остается открытым. Нет сомнений в том, что дифференциация по языковым уровням нужна для оперативности, четкости и эффективности организации пополнения, и классификация исходных и промежуточных данных по принадлежности к уровням (или этапам, стратам и т. п.) необходима. Тем не менее, исследования по искусственному интеллекту (по ситуационному управлению) показали, что четкое, раз навсегда заданное отделение исходных данных от алгоритма решения задачи не всегда достижимо, и, что главное, не всегда целесообразно. На каждом этапе могут быть собственные исходные данные². И в самом МП такая же точка зрения высказывалась достаточно рано.

Последовательно индуктивный подход в построении систем МП нам представляется также мало эффективным. Так, если мы зададимся целью использовать в МП только сведения, которые нам объективно дает языковая статистика, то мы рискуем оказаться в полной неопределенности, например, на стадии рассмотрения синтаксиса, поскольку нет (и, видимо, не может быть) достоверной статистики синтаксических явлений языка. Между

тем имеются достаточно логичные, объективно проверенные, подтвержденные практикой языкового общения синтаксические концепции, как разработанные в рамках дедуктивного подхода в МП, так и, что более важно, давно уже существующие в традиционной обычной грамматике и лингвистике.

В связи с вышесказанным нам представляется, что современное состояние МП характеризуется некоторым слиянием результатов двух подходов, но не механическим соединением результатов, достигнутых в рамках каждого из них, а слиянием их на базе новых моделей, созданных при основном внимании к собственно переводческому аспекту владения естественным языком. Короче говоря, центром исследований в современном МП становится моделирование действий человека — переводчика, особенно в части использования им двух- и многозначных переводных соотвествий при переводе с одного языка на другой.

МП, возникший вне лингвистики, вошел в лингвистику главным образом благодаря чрезвычайно важному аспекту, с которым он был с самого начала существенным образом связан, а именно — аспекту моделирования. Можно утверждать, что именно МП был первопричиной появления в языкознании этого плодотворного в дальнейшем представления. Теоретическое языкознание с некоторого времени уделяет большое внимание понятию модели и связанной с ним проблематике. Не касаясь здесь многочисленных трактовок этого понятия, мы лишь отметим, что представление о действующей модели сыграло огромную роль в автоматической обработке текстовой информации — оно привело к созданию новых областей языкознания, таких как вычислительная лингвистика, инженерная лингвистика, теории математической лингвистики, количественная и статистическая лингвистика и пр. При всей неисчерпаемости естественного языка как объекта изучения модель дает возможность исследовать его конкретные свойства в данной проблеме. Для МП нужны модели, реализующие оперативную обратную связь, т. е. модели нового для лингвистики типа. Такими моделями являются, например, воспроизводящие инженерно-лингвистические модели, описанные в (Питровский 1979).

В теоретическом языкознании моделирование ограничивалось построением идеальных аналогов лингвистического объекта. К лингвистическим моделям предъявлялось два требования — последовательность и непротиворечивость выделения существенных свойств лингвистического объекта, с одной стороны, а с другой — способность объяснять наблюдаемые факты и предсказывать ранее неизвестные свойства объекта. Дальнейшее развитие лингвистической теории, стимулированное лингвистической практикой, особенно в области автоматической обработки текстов, требует дополнить эти два требования также и третьим, а именно: лингвистическая модель должна воспроизводить языковые и речевые объекты. Модели типа, воспроизводящие объекты, представляют собой искусственно созданные формальные систе-

² «На данном этапе анализа есть все основания полагать, что разобшение материала задачи и операций по ее преобразованию является одной из причин наблюдающейся в ряде случаев неэффективности машинных программ по сравнению с мыслительной деятельностью человека». (Пушкин 1980, 17).

мы, построение или поведение которых имитирует микроструктуру или функционирование некоторого лингвистического объекта, в зависимости от того, строение или функционирование являются целью моделирования, причем строение также определяется функционально: строение модели имитирует строение объекта тогда и только тогда, когда воспроизводятся некоторые характеристики этого строения. Лингвистическое моделирование служит надежным способом реализации лингвистических теорий и одновременно средством проверки их строгости и жизнеспособности, чем замыкается связь между лингвистической теорией и практикой.

Возвращаясь к проблеме построения действующих систем МП, следует сказать, что именно в этом аспекте проявляется диалектическое противоречие между двумя указанными выше подходами. Ни один из них «в чистом виде» не может служить основанием для конкретной работы по построению системы МП. Механическое соединение теории, дедуктивно разработанной в рамках первого подхода, и существующих алгоритмов анализа, словарей, статистических данных и пр., созданных в рамках второго подхода, также не может породить действующую систему МП. Для построения последней необходимо создание специальных «воспроизводящих инженерно-лингвистических», или рабочих, переходных между теорией и языковыми фактами, промежуточных моделей, точно описывающих процесс перевода с одного естественного языка на другой. Промежуточными эти модели можно назвать потому, что они занимают среднее, промежуточное положение между тотальными теориями языковой деятельности и совокупностью языковых фактов, систематизация которых необходима для практического решения задачи. Представляется важным, чтобы эти модели имитировали не абстрактные свойства языка в целом, а конкретную языковую функцию, в частности, перевод с одного естественного языка на другой, перевод в данной паре языков в первую очередь, и затем уже переход к другим парам и обобщение полученных закономерностей.

Известно, что перевод как особый процесс межъязыковых преобразований затрагивает в комплексе разные уровни языка — морфологию, лексику, синтаксис, семантику. В переводе имеет место сложное взаимодействие этих уровней, такое, что, по существу, здесь появляются новые исходные единицы — «единицы перевода», «переводные соответствия», «закономерные соответствия» и т. д. Эти последние единицы, как правило, принадлежат разным языковым уровням. Модель перевода, таким образом, должна отражать иерархию языковых уровней, причем некоторым оптимальным для перевода образом. Наряду с чисто рутинными действиями (поиск по словарю, перевод фиксированных словосочетаний, перевод конструкций, имеющих однозначные соответствия и пр.) процесс перевода может затрагивать принципиальные сущности человеческого владения языком, мышления через язык. Таким образом, одной из подлежащих разреше-

нию проблем является взаимодействие языковых уровней в переводе. Другой, не менее важной для машинной реализации проблемой является соотношение уровней формализации. Современное развитие соответствующих областей науки — вычислительной лингвистики, формальной семантики, прикладного языкознания — дает возможность обеспечить формальное описание единиц и систем единиц разных языковых уровней. Однако уровни формализации в комплексе, обеспечивающем перевод, должны определяться, исходя из общей концепции. Поэтому особой проблемой в МП является отношение между традиционно выделяемыми языковыми уровнями и уровнями формализации.

В духе теории воспроизводящих инженерно-лингвистических моделей при создании модели системы МП мы в первую очередь обращаем внимание на собственно переводческие действия человека. Любой перевод есть перевод по переводным соответствиям, если мы подходим к нему с достаточно широкой меркой. Мы говорим, что текст перевода соответствует тексту оригинала в том случае, если для каждого отрезка текста оригинала есть некоторый отрезок текста перевода, «соответствующий» или эквивалентный ему «по содержанию», причем границы и рамки соответствия понимаются иногда достаточно широко. В теории перевода есть концепция «уровней эквивалентности», где специально рассматриваются вопросы достижения текстовой эквивалентности для разных по протяженности отрезков текста. Поэтому само по себе использование переводных соответствий просто неизбежно, если мы говорим о переводе как таковом, а не о реферировании, аннотировании и пр. Новизна в моделировании с использованием переводных соответствий заключается именно в том, что они ставятся в центр модели и всего моделирования.

Особое значение имеет принцип ориентированности на цельную систему МП. На протяжении всей истории развития МП и в настоящее время существуют скептики, которые утверждают, что машинный перевод невозможен вследствие того, что машина не может разрешить тот или иной тип неоднозначности, не может учесть контекста в полном объеме, необходимых экстралингвистических сведений, не может сделать логического вывода и т. д. Однако некоторый парадокс заключается в том, что практически для всех приводимых в таких случаях примеров неразрешимости на самом деле существуют способы формализации и достаточно алгоритмизуемые последовательности действий, в результате которых данная проблема удовлетворительно разрешается. Дело в том, что такие алгоритмизуемые последовательности действий на достаточно формализованной исходной информации должны быть объединены в непротиворечивую систему, если мы хотим построить систему для перевода не одного предложения, а совокупности предложений — текстов. Механическое соединение частных рецептов, естественно, не обеспечит такого построения.

Таким образом, прикладная модель, необходимость которой постулируется, должна упорядочивать языковые факты в двух отношениях: во первых, в соотношении языковых уровней и уровней формализации, и, во-вторых, в соотношении подлежащих учету и разрешению языковых закономерностей в их текстовом проявлении. Ситуации, в которых полностью адекватный и эквивалентный перевод невозможен, существуют и для человека-переводчика. Трудно ожидать, что в таких случаях станет возможным перевод машинный. Однако и для тех ситуаций, в которых автоматическое, «машинное» решение возможно и существует в виде конечной последовательности формальных операций, это решение не должно входить в противоречие с общей концепцией модели, рассчитанной на перевод некоторого большинства слушателей.

Большое значение имеет пригодность работы системы в режиме общения с человеком. Здесь следует отметить два аспекта: технологичность собственно интер- и постредактирования (диалоговый режим), которая может базироваться только на специально заложенных при разработке системы принципах, и такое лингвистическое строение системы, которое позволяет вносить необходимые коррективы в словарь и алгоритмы системы. Точнее, здесь важны: а) простота и удобство корректировки продукта, производимого системой МП, и б) доступность и технологичность корректировки частей самой системы и системы в целом. Современный МП не может быть реализован как процесс постоянной работы некоторого раз навсегда сделанного, завершеного и некорректируемого устройства. Если различать процессы генерации и эксплуатации системы, то восприятие изменений должно быть частью не только процесса генерации, но также и характеристикой эксплуатационного применения системы к переводу. Такая характеристика системы может быть получена только при соответствующих принципиальных качествах лингвистического обеспечения.

В начале работ по МП в вихре новых идей относительно формализации естественного языка была достаточно убедительно показана «несостоятельность» естественного языка и его «традиционных» грамматических и лексических описаний с точки зрения его математизации и структурализации. При этом, как сейчас видно, безосновательно смешивались проблемы формализации (математизации) языка и собственно машинного перевода. Многовековые традиции пользования языком в человеческом обществе и описания языка в целях обучения, нашедшие выражение в грамматических системах, иногда незаслуженно игнорировались. Поэтому в построении лингвистического обеспечения систем МП бывает полезно вернуться к результатам ранее проведенных исследований в поисках ответа на сегодняшние актуальные проблемы.

1.2. Возникновение идеи и периодизация истории машинного перевода

Возможность механизации такого казавшегося рутинным умственного труда, как перевод, давно привлекала внимание. В 1933 г. в СССР было запатентовано механизированное переводное устройство П. П. Смирнова-Троянского (Панов, Ляпунов, Мухин 1956). Однако на МП это устройство не оказало никакого влияния, поскольку в нем было реализовано слишком примитивное представление о сущности переводческого процесса. Собственно машинный перевод начинается с применения электронных вычислительных машин (ЭВМ) к обработке текстов на естественных языках. Эта идея основывалась на успехах, достигнутых в применении ЭВМ к криптографическим задачам. В дискуссии между У. Уивером и А. Д. Бутом в 1946 г. было высказано предположение, что «основные элементы языка могут быть обнаружены с помощью устройств, разработанных во время второй мировой войны и предназначенных для расшифровки секретных кодов противника» (Бут, Локк 1957, 16). После публикации работ К. Шеннона по математической теории связи, оказавших большое влияние на развитие математической теории информации, были сделаны попытки применить к анализу языка инструментальных средств, пригодных для обнаружения общих свойств кодов, и машинный перевод рассматривался как еще одна область применения техники дешифрования. При этом важно, что предполагалось осуществить обнаружение на первых порах самых общих свойств «кода». В меморандуме У. Уивера — одном из исходных документов, от которых началось развитие МП, — приводится пример дешифрования и отмечается важное качество языковых сообщений, а именно инвариантность языковых характеристик на разных уровнях, позволяющая найти хотя бы самые общие свойства кода. «...Можно было догадаться, что при расшифровке использовалась частотность букв и комбинаций букв, ... которые почти не зависят от того, каким языком написан зашифрованный текст. Сразу же возникает предположение, что на том этапе, когда человек изобретал и развивал языки, возникли некоторые их инвариантные свойства, которые опять-таки неточно, но до некоторой статистически полезной степени присущи всем языкам» (Уивер 1957, 34).

Порядок установления иерархии языковых уровней в общем случае реконструкции знаковой системы может быть иной — не от мельчайших единиц (низших уровней) к высшим, а от общих уровней к частным. «При любой дешифровке, как и при всех других видах реконструкции текстов на естественных языках, следует исходить из того, что дешифруемый текст (как и любой языковой текст) может быть представлен как результат последовательных преобразований, осуществляющихся в соответствии с порядком иерархии языковых уровней от наиболее общих смысловых уровней через последовательность более дробных уровней,

оперирующих единицами все меньшей величины, вплоть до фонологических уровней (фонем и различительных признаков)» (Иванов, Топоров 1966, 5).

Можно считать, что на становление и развитие МП повлияли следующие этапы криптоанализа, касающиеся различных уровней языкового описания:

- исследование чисто внешних характеристик сообщения, таких как частота отдельных элементов и сочетаний элементов;
- исследование дистрибуции элементов, составляющих текст;
- анализ билингв с целью выявления совпадающих элементов и использования некоторых ключевых элементов (собственных имен и пр.) с учетом статистических и дистрибутивных характеристик;

— нахождение ключа и применение его ко всей последовательности закодированных сообщений (Kahn 1968).

Процесс криптоанализа подразумевает совместную работу человека и ЭВМ; работа аналитика состоит из рутинной обработки больших массивов текста в непрерывной связи с тонкими умственными догадками, требующими лингвистической интуиции³.

Основные вехи дальнейшего развития МП следующие: в 1947 г. А. Д. Бут и Д. Г. В. Бриттен разработали подробный «код» для пословного МП. В 1948 г. Р. Г. Риченс предложил правила разбиения словоформ на основу и окончание. Первая конференция по МП состоялась в Массачусетском технологическом институте в 1952 г. В 1954 г. состоялся первый публичный эксперимент по МП — Джорджтаунский эксперимент, когда на машине ИБМ 701 в Нью-Йорке были переведены несколько специально подготовленных русских предложений.

Истории машинного перевода в настоящее время посвящено достаточно много работ, в том числе и обзорного библиографического характера. Интерес к истории этой проблемы связан с тем, что в ходе направленных на ее разрешение исследований возникли многие получившие дальнейшее развитие гипотезы, предположения и идеи. Настоящий краткий экскурс в историю МП, ни в коей мере не претендующий на полноту, осуществлен лишь с целью обратить внимание на отдельный аспект проблемы МП, носящий скорее методический, нежели принципиальный характер. Этот аспект связан с постепенным обнаружением свойств релевантной переводу языковой структуры. Именно он позволяет прийти к концепции приближенных лингвистических вычислений применительно к проблеме МП. Представляется важным обратить внимание на этот аспект вследствие того, что дальнейшее развитие МП пошло главным образом в сторону развития лингвистических теорий, которым самой проблемой был дан весьма ощутимый импульс. Само по себе положительное явление, это развитие затемнило собой технологические стороны проблемы.

³ См. также: Техника молодежи, 1971, № 8, с. 36—39.

Ход исследований представляется возможным осветить, пользуясь принципом периодизации.

Историю машинного перевода можно разделить на три периода, каждый из которых охватывает примерно десять лет: 1946—1957 гг., 1957—1967 гг., 1967 г. — настоящее время. Общие характеристики каждого из периодов следующие.

Первый период включает появление идеи и связанный с этим энтузиазм. Оптимистические оценки ближайших перспектив машинного перевода базировались в значительной мере на таких же оптимистических воззрениях на перспективы развития электронной вычислительной техники вообще, поскольку природа ЭВМ и их ограничений была еще не вполне ясна. Еще менее ясности было в вопросе о возможностях и границах формализации применительно к такому сложному общественному интеллектуальному феномену, каким является естественный язык. Возникли многочисленные научные коллективы, многие из которых ставили своей целью быстрое получение машинного перевода с разных языков на многие другие. Статистические и теоретико-информационные измерения языка привели к созданию концепции языка как кода. Внимание к МП было привлечено первыми машинными опытами, такими как Джорджтаунский эксперимент. Особое внимание отводилось построению машинных словарей, хотя уже складывались представления об алгоритмах морфологического, синтаксического, лексического анализа для МП. Если кратко подвести основные итоги этого периода, то они выглядели бы следующим образом:

- выдвинута концепция языка как кода;
- созданы первые машинные словари для МП экспериментального характера;
- разработано общее представление об МП, включающее машинный словарь, алгоритмы анализа и синтеза, программное обеспечение;
- созданы и опробованы первые экспериментальные системы МП, которые, хотя и носили рекламный характер, обладали рядом существенных для систем МП черт.

Наиболее плодотворным можно считать следующий, второй период в истории МП. Он закончился в некотором смысле декларированным отходом от МП как проблемы, однако в целом характеризовался повышенным вниманием к МП как теории. Это несколько парадоксальная ситуация объясняется достаточно частой в истории науки реакцией на первичный энтузиазм. Более пристальное внимание к разработке компонентов систем МП показало, что слишком многие аспекты языковой деятельности чрезвычайно далеки от формализации и тем более алгоритмизации. Лингвистические исследования этого периода, проводимые в русле МП, чрезвычайно разнообразны. Можно утверждать, что именно в хронологических рамках второго периода началась реальная дивергенция научных интересов, впоследствии приведшая к тому, что МП стал представлять собой конгломерат разных

наук, и не только лингвистического цикла, а также связанных с ЭВМ другими аспектами. Особенно следует отметить возникновение и подробную разработку идеи языка-посредника. От общих положений, таких, как выгодность перевода с одного из языков на язык-посредник и только с него на любой другой, противопоставляемая неэкономичности разработки алгоритмов перевода с каждого на каждый язык из группы в n языков, до конкретных представлений и попыток создать такой язык шло развитие этой идеи. Хотя она и не оказала сколько-нибудь существенного влияния на создание промышленно ориентированных систем МП, но в значительной мере стимулировала развитие теоретических представлений о методах и средствах анализа и синтеза. Кульминацией периода можно считать доклад комиссии Национального научного фонда США, специально посвященный экономической целесообразности МП и перспективам научных исследований в этой области. (Язык и машины 1968). В этом докладе, исходя из реальной ситуации с переводами в США и показателей стоимости перевода, делался вывод о невыгодности машинного перевода, недостижимости целей, поставленных в МП, в течение сколько-нибудь разумного периода времени, содержались рекомендации изучать иностранные языки вместо того, чтобы поручать дело перевода машине. Пессимизм доклада в отношении МП не получил поддержки многих ученых в США и за рубежом; однако доклад оказал серьезное влияние на оценку МП, начиная с того времени.

Таким образом, в течение второго периода:

- осознана недостаточность исходных представлений о языке для его полного формального описания, необходимого для создания алгоритмов и систем МП;

- в развитие концепции языка как кода разработаны статистические и теоретико-информационные методы измерения речи;

- возникла идея языка-посредника;

- возникло и утвердилось представление о теоретической важности и целесообразности отделения собственно лингвистического описания от алгоритма. Можно считать, что соответствующие представления явились некоторым следствием разработки идеи языка-посредника;

- центр тяжести исследований переместился от лексики к синтаксису;

- сформулированы основные идеи семантического анализа для МП и принципиального использования смысла для МП;

- созданы и опробованы в эксперименте разнообразные алгоритмы синтаксического и семантико-синтаксического анализа;

- выдвинута концепция синтеза (анализ через синтез);

- автоматизирован морфологический анализ для многих языков. Точнее, были достаточно подробно разработаны способы анализа словоформ аналитических языков и созданы соответствующие машинные алгоритмы и программы;

- начато составление обширных словарей для МП с разнообразной информацией;

- продолжена эксплуатация экспериментальных систем МП и накоплен важный опыт их программной реализации;

- уточнены требования к математическому и программному обеспечению систем МП и отдельных алгоритмов анализа и синтеза;

- составлено большое количество частотных словарей, разработана методика анализа с использованием статистических и теоретико-информационных характеристик, сложились основы количественной лингвистики, статистической лексикографии и других наук статистического цикла;

- сделаны успешные попытки промышленной эксплуатации систем МП в режиме реального информационного обслуживания;

- отработана методика контекстного анализа для функционального разрешения основных лингвистических проблем.

Параллельно с развитием МП в этот период возникла идея создания автоматического словаря в помощь человеку-переводчику. Автоматический словарь получил практическое применение.

В течение этого десятилетия ускоренно развивалась вычислительная техника, появлялись новые возможности машинной реализации лингвистических формализмов, которые, может быть, не всегда были в соответствующей мере осознаны создателями лингвистических теорий прикладного характера. Второй период характерен усилением внимания к теоретическим проблемам МП. Возникли и получили развитие такие лингвистические идеи, как гипотеза глубины, представление о проективности синтаксических структур, были разработаны алгоритмы синтаксического анализа, способы изображения, представления и обнаружения синтаксических структур. Развитие представлений о порождающих механизмах языка привело к созданию ряда моделей, главным преимуществом которых можно считать семантические формализмы. Ряд действующих систем МП был внедрен в информационную практику именно в этот период: результаты работы этих систем позволили по-новому взглянуть на возможности и перспективу развития МП.

Третий период, начавшийся после 1967 г., характерен возникновением нового социального статуса МП. В это время сформировалась наука о научно-технической информации (информатика), фундировавшая возросшие потребности в научных коммуникациях века научно-технической революции и резкое увеличение активности в этой области. Если во втором периоде развитие научной мысли в МП шло в некотором отрыве и с учетом новых возможностей вычислительной техники, то третий период, как можно считать, соединил новые достижения технической кибернетики с новыми результатами теории и практики лингвистических исследований в автоматической обра-

ботке текстов. Это соединение произошло на базе осознания социальных потребностей в области МП — преодоления языковых барьеров. В научном аспекте этот период характерен возвратом к построению цельных систем МП. Однако этот возврат происходил достаточно медленно, и только к 1977 г. активизировалась работа в этом направлении. Показательным событием этого периода является конференция «Преодоление языковых барьеров» 1977 г. в Люксембурге. Председатель секции машинного перевода на этой конференции следующим образом описал современное состояние МП: «Поскольку теоретические основы в достаточной степени созданы, перед нами теперь стоят проблемы совершенствования систем и включения человека-переводчика в автоматические процессы, особенно в диалоговом режиме. Становится очевидным соединение автоматического перевода и устройств помощи переводчику. Безусловно, возникнут новые проблемы: проблемы соревнования, даже соперничества, плюрализма и направлений развития. Конференция в Люксембурге может стать важной вехой в развитии, подобно тому, какой стал знаменитый доклад десять лет тому назад, однако в противоположном направлении» (Zemb 1978, 73).

Можно считать, таким образом, что развитие МП происходило согласно гегелевской триаде, каждому из элементов которой соответствовал свой период — тезис, антитезис, синтез. Современный период в такой концепции соответствует синтезу. Практически состояние дел с МП подтверждает такую точку зрения.

Мы уже подчеркивали эскизность настоящего экскурса в историю МП и его направленность на освещение одной стороны развития МП, а именно технологического аспекта, основанного на методической стороне проблемы. Назовем наиболее известные публикации, которые могут служить основой для более тщательного изучения истории развития МП, чего она, безусловно, давно уже заслуживает, не столько сама по себе, сколько как некоторый «склад» различных, часто весьма интересных, лингвистических и кибернетических идей, лишь некоторые из которых получили впоследствии развитие и среди которых безусловно есть достойные не быть забытыми в дальнейшем развитии науки. Сборник, относящийся к первым годам появления МП: (Машинный перевод 1957). Довольно полные библиографические обзоры: (Автоматический перевод 1967, Общее и прикладное языкознание 1972). Последующие сборники (Автоматический перевод 1971). Обзорные работы (Кулагина 1979, Bruderer 1978), (Dostert 1973) и ряд других.

1.3. Действующие системы машинного перевода

В научной литературе, особенно в советской, понятие «действующие» системы машинного перевода часто применялось по отношению к исследованиям и разработкам по МП, которые

имели вид алгоритма, а не описания. Именно в таком смысле это понятие использовалось, например, в отношении моделей «смысл — текст» (Кулагина 1979). По существу, таким образом это название приписывалось моделям, которые не действовали, но в принципе могли бы быть сделаны таковыми по утверждению их создателей. Нам представляется, что здесь слово «действующие» было несколько преждевременным. В нашем рассуждении мы сохраним за этим словом его первоначальный смысл, т. е. под «действующими» системами будем понимать системы МП, включенные в реальное переводческое обслуживание, дающие пригодный для использования информационный продукт, который имеет товарный вид, т. е. продается на информационном рынке за некоторую плату (при этом, естественно, может иметь место экономическая неэффективность, действуют законы конкуренции и пр.).

Конкретное описание некоторых действующих систем МП дано в Приложении I. Следует отметить, что в этом аспекте положение меняется очень быстро и соответствующая информация устаревает растущими темпами. Поэтому в настоящем разделе мы обратим внимание лишь на некоторые аспекты использования действующих систем МП.

Первая характерная черта действующих систем МП заключается в том, что все они являются человеко-машинными, т. е. требуют существенного участия человека-редактора или переводчика, корректирующего машинный выход.

Из этого обстоятельства вытекают многие практические следствия.

В отличие, например, от автоматических словарей в помощь переводчику, здесь заменяется сам основной деятель в процессе перевода, т. е. переводчик. Он становится редактором. Выигрыш во времени, который почти всегда существен, компенсируется в отрицательную сторону гораздо большими затратами редакторского труда, в сопоставлении с теми, которые имеют место при редактировании перевода, выполненного человеком. Кроме того, сам характер редакторского труда становится иным. Вместо анализа и корректировки некоторого «гладкого» текста приходится иметь дело с повторяющимися досадными и глупыми ошибками машины; преимущество в том, что машина не может скрыть за гладкостью текста своего непонимания сути, как может это сделать переводчик, зато иногда машина раздражает редактора тем именно, что она «не понимает» простых для него вещей.

Далее, бесспорно, что машинному переводу целесообразно подвергать тексты определенного вида. В первую очередь это научно-технические тексты, имеющие вид спецификаций, инвентарных списков, стандартизированных описаний и пр. Сколько-нибудь сложный в синтаксическом и семантическом отношении текст при переводе на ЭВМ требует такого большого объема постредактирования, что становится экономически невыгодным

его машинный перевод. Можно предполагать, что в дальнейшем требования МП заставят разработчиков употреблять более стандартизованный язык, что сделает возможным расширение круга документов, которые могут быть переведены на ЭВМ; не исключено, что такая стандартизация будет стимулироваться желанием распространения. Пока что приходится ориентироваться на существующие источники и виды литературы.

Существенным моментом, который должен быть учтен, является то, что в этом качестве МП ломает традиционную сферу переводческого обслуживания. Потребитель склонен мириться с тем, что переводы делаются долго, лишь бы они были высокого качества и походили на написанное «человеческим языком». Внедрение МП в переводческое обслуживание связано с дальнейшей дифференциацией процесса распространения и использования переводов, с «настройкой» потребителя на машинный перевод, может быть, на первых порах в виде сигнальной информации с последующей тщательной отделкой только действительно нужных потребителю вещей. Такая дифференциация уже давно вводится в переводческое обслуживание. С 1973 г. Всесоюзный Центр переводов, например, ввел так называемые два потока для научно-технических переводов. Первый поток характеризуется тем, что поступающие переводы после работы над ними переводчика не проходят обязательного научного редактирования; последнее делается только для работ «второго» потока, для случаев особо трудных текстов или языков. Первый поток в процентном отношении занимает преобладающее место в продукции ВЦП, а именно около 90%. Отмена научного редактирования не привела к возрастанию жалоб потребителей.

О дифференциации переводческого обслуживания говорит также опыт зарубежных многоязычных сообществ, например, Европейского экономического сообщества. Тексты технического характера типа спецификаций и пр. в рамках этого сообщества переводятся на ЭВМ системой СИСТРАН уже в течение достаточно долгого времени.

1.4. Требования технологии

МП как проблема научно-технической информации требует прежде всего оценки трудоемкости создания начального цикла обработки текста, основных параметров системы, отправляясь от которых можно постепенно уточнять получаемую информацию в порядке реализации приближенных вычислений.

Неоднократно отмечалось многими авторами с разных точек зрения особое свойство лингвистических объектов. Как правило, эти объекты образуют множества, границы которых часто определены весьма расплывчато (Шрейдер 1971, Пиотровский 1975). Математический взгляд на проблему приводит к концепции нечетких множеств, лингвистический — к идее нормы, употребления и пр. Однако, если в инженерных и технических нау-

ках для определения точно не вычисляемых величин установилось понятие приближенных вычислений, то в лингвистике, особенно в прикладной, где такие представления необходимы, они еще, по существу, не сформировались.

Для реализации идеи приближенных вычислений в лингвистике необходимы:

— рабочая теория (модель), точно соответствующая данной задаче;

— «поправочные коэффициенты» (по аналогии с физикой), которые учитывали бы факты, отклоняемые чистой теорией;

— определение допустимой и необходимой точности вычислений, величины допустимой ошибки и цены, которую придется платить за требуемую точность вычислений;

— метод приближенных вычислений. Метод имеет особо важное значение, поскольку именно он определяет технологию и все ее основные, в том числе и экономические, характеристики. Теории, моделирующие язык в целом, могут вообще не касаться методов получения соответствующих описаний. Рабочие теории не могут абстрагироваться от метода;

— оптимальным образом организованное получение массовой информации. Массовость может пониматься и в статистическом смысле, и как наиболее естественное и простое в языке. Характеризирующее норму;

— способ исключения прочного (замкнутого) круга. Для ряда прикладных задач, особенно в автоматической обработке текстов и в МП, требуются специальные приемы для обеспечения получения информации, которая может быть получена, вообще говоря, только при наличии некоторой другой;

— алгоритм, т. е. последовательность однозначных действий с проверкой результатов после некоторого количества шагов. Для совершенствования алгоритма применяется принцип обратной связи.

В рамках первого подхода к МП возникают технологические трудности, связанные в основном с двумя проблемами: проблемой отделения описания от алгоритма работы с ним и проблемой использования глубинного языка смысла для МП.

Основой системы МП при таком понимании является аппарат, позволяющий от поверхностной структуры перейти к глубинной — от текста перейти к его смыслу⁴. Лингвистическое обеспечение такой системы строится на следующих пяти принципах: 1) отделенность лингвистического обеспечения от алго-

⁴ «При движении от текста к смыслу формальные представления предложений на каждом шаге утрачивают специфические черты исходного естественного языка и приобретают черты универсализма; семантический язык, предназначенный для записи смысла предложений на самом глубоком уровне, является, по замыслу, универсальным» (Апресян, Богуславский, Йомдин и др., 1978, 5). Нетрудно заметить аналогию между такого рода универсальным языком смысла и проектами международных универсальных языков смысла — см. подробное исследование П. Н. Денисова о принципах моделирования языка (Денисов 1965).

ритмов; 2) обратимость лингвистических правил и их принципиальная неупорядоченность; 3) стандартизованность форм записи языковой информации и связанная с этим независимость описания каждого из языков; 4) приоритет словарей (при весьма сложной грамматике); 5) включение в систему глубинно-синтаксического уровня. Идейной основой такого подхода является стремление избежать неконгруэнтностей двух языков в переводе посредством обращения к смыслу: если мы будем знать смысл, то при наличии определенного языкового инвентаря, сопоставляемого единицам смысла, мы всегда сможем различными, если нужно, способами выразить заданный смысл в любом выходном языке. Концепция такого рода родилась в машинном переводе в конце пятидесятых годов под влиянием трудностей, связанных с собственно языковым описанием для МП и была отражена в ряде публикаций в СССР и за рубежом (Автоматический перевод 1967). Можно проследить этапы развития этой концепции: принцип семантических множителей — анализ через синтез — идея толково-комбинаторного словаря с обширной лексико-семантической и синтаксической информацией — разделение поверхностной и глубинной структур.

Идея разделения лингвистического описания и алгоритма работы с ним была впервые предложена С. Лембом и В. Ингве. В их работах (Лемб 1964) и (Ингве 1965) говорится о том, что целесообразно отделять языковые уровни лингвистического описания от машинных программ (в частности, полезен словарь основ по сравнению со словарем словоформ) и что грамматическое описание должно быть отделено от соответствующих блоков алгоритма (в частности, особый блок должен заниматься межъязыковыми преобразованиями). Однако разделение описания и алгоритма рассматривается ими лишь как средство повышения эффективности алгоритма. В названных работах везде подчеркивается своеобразие переводческого процесса как сложного взаимодействия различных уровней языка. Известно, что и безотносительно к переводу язык не может быть описан однозначным и непересекающимся набором единиц разных уровней: имеют место сложные и разнообразие взаимосвязи и зависимости между единицами описания. Если бы этого не было, язык не отличался бы от кода. Перевод же, как конкретную задачу особого оперирования с языковыми единицами, трудно себе представить без описания одних единиц через другие, рекурсий, возвращений к пройденному пути и т. п. Поэтому часто оказывалось, что стремление отделить описание от алгоритма было трудно реализовать в конкретных системах МП⁵.

⁵ «По-видимому, и не следует стремиться к тому, чтобы непосредственное задание алгоритма было полностью исключено из методов работы; полезно сохранять его там, где разделение на отдельные шаги носит содержательный характер. Именно так обстоит дело с делением всего процесса перевода на последовательные этапы, порядок которых задан. В пределах каждого из этапов могут быть применены различные подходы к построению алгоритма этого этапа» (Лейкина, Никитина, Откупщикова и др., 1966, 47).

Отметим также, что в отношении разделения описания и алгоритма МП оказывается в положении, отличном от других лингвистических задач. Так, можно себе представить, что в задаче обучения иностранному языку можно четко отделить описание языковых данных от процесса оперирования с ними: можно четко и однозначно наметить последовательность вывода в действие новых порций данных и т. п. Однако на этом основании никто не предлагал поручить составление алгоритма обучения иностранным языкам математикам. Специфика переводческого процесса, игнорируемая в стремлении полностью отделить описание от механизма работы с ним, существенно определяет направление и порядок алгоритмизации, в такой же мере, как и специфика других процессов в различных прикладных задачах.

Не требует особых доказательств и положение о том, что слияние алгоритма с описанием значительно затрудняет разработку алгоритма перевода в целом, постоянную корректировку и использование обратной связи для совершенствования системы перевода.

При таком подходе речь идет не о собственно переводе, а об автоматическом реферировании, толковании, развертывании смысла от некоторого исходного к такому, словесное выражение которого можно было бы принять за эквивалентный перевод заданного исходного текста. Главной при этом становится проблема осуществления логических преобразований над некоторой начальной информацией и вопрос о собственно переводных соответствиях уходит на самый последний (поверхностный) план. «Переводиться» (точнее, истолковываться) здесь должны не отдельные предложения, а весь текст целиком. Два сложных теоретических и практических вопроса затрудняют использование такой модели в качестве рабочей для МП: эквивалентность перифраз и множественность вариантов анализа (синтеза), требующая средств выбора одного из них в качестве основы для перевода. Утверждать, что два синтаксически и лексически разных предложения имеют один и тот же смысл, можно только весьма приблизительно⁶.

Многовариантность в сочетании с отходом от поверхностной структуры к глубинной приводят к тому, что иногда утрачивается исходное представление о структуре и ее приходится заново восстанавливать, вместо того, чтобы просто ее сохранить (Boitet 1978, 90). Чрезвычайно трудно строить всеобъемлющую систему фильтров, базируясь на представлениях типа «такое-то выражение более идиоматично, чем другое», «так говорят — так не говорят», и т. п. Локальные указания в статьях толково-

⁶ «Подстановка в дискурсе вместо предложения какого-либо его преобразования равносильна замене в слове одной морфемы другой, равнозначной, в результате чего получались бы формы вроде «козляника» и «куряника» рядом с «клубнятина» и «ежевятина». Из всего этого следует, что собственно смысл протвinitся преобразованиям предложений». (Звегинцев 1976, 197).

комбинаторных словарей, которые достаточно конкретны, отличаются от общих оценок, применимых в системе в целом, поэтому в настоящее время многие сторонники смыслового (глубинного) языка-посредника начинают занимать критическую позицию в отношении многовариантности. Связь МП в данной концепции со смысловым анализом всего текста в целом не раз подчеркивалась сторонниками этого подхода (см., напр., Леонтьева, Никогосов 1977); это означает, что задача собственно перевода заменяется задачей преобразований смысла текста без учета собственно языковых средств передачи заданного содержания на уровне языковых (речевых) переводных соответствий.

Второй подход к МП базируется на представлении о постепенном получении из текста информации, на базе которой можно строить перевод. «Лингвистика текста», в которой наиболее последовательно отражен данный подход, представляет МП как продукт следующих стадий переработки⁷: создание автоматического словаря, с помощью которого можно получить лексический перевод, являющийся основой всякого перевода, поскольку основная часть синтаксической (структурной) и смысловой информации заложена в лексике и фразеологии текста; построение системы семантического перевода, который опирается на алгоритмы устранения многозначности, позволяющие автоматически распознавать актуальные значения лексических единиц в перерабатываемом тексте; составление грамматических алгоритмов. Этот подход базируется на особом рассмотрении текста: лингвистика текста рассматривает последний как избыточную многоуровневую систему, на верхних ярусах которой находятся наиболее информативные лексические единицы, на нижних — малоинформативные элементы. Ориентация делается только на бинарный перевод, при построении которого структуры входного и выходного языков объединяются в суперструктуру, которая называется двуязычной ситуацией. Все алгоритмы МП предполагается строить на основе идеи формального распознавания смыслового образа входного текста. Необходимо отказаться от стапроцентной переработки текста. При этом предусматривается последовательное развертывание системы МП от простых алгоритмов, работающих на верхних ярусах двуязычной ситуации, и обеспечивающих извлечение из входного текста максимума основной информации, к более сложным алгоритмам, которые ориентированы на глубинные уровни этой ситуации и с помощью которых из входного текста извлекаются малые дозы более тонких и детальных сведений (Пиотровский 1975, 270—271).

⁷ «...Конечная цель ЛТ (лингвистики текста.— Ю. М.)... состоит в том, чтобы установить инвариантные схемы построения текста, или, иными словами, сформулировать правила текстообразования. Однако в отличие от дедуктивно-интроспективного подхода порождающей грамматики ЛТ стремится решить эту задачу индуктивным путем, рассматривая большие массивы текста» (Пиотровский 1975, 55).

Этот подход привел к созданию ряда алгоритмов МП, работающих по принципу пословного перевода. «Имплицитная статистика» первого подхода заменяется здесь широким эксплицитным применением статистических методов.

Однако и здесь возникают технологические трудности, которые можно суммировать следующим образом:

— в рамках последовательного проведения статистического подхода трудно смоделировать собственно перевод, поскольку последний касается многоуровневой информации, а статистическое описание таких уровней, как синтаксический, и особенно семантический, существенно затруднено, вследствие чего трудно создать комплексное и вместе с тем последовательно статистическое описание процесса. Можно утверждать, что статистическое описание предполагает иную формализацию уровней и содержит другие уровни, чем естественно-логическое описание процесса межъязыкового перевода;

— последовательно индуктивная методика заставляет заново открывать многие важные аспекты языковой деятельности, существенные для перевода и уже подтвержденные многолетней практикой.

Отсюда следует, что затруднительно строить систему МП на принципах каждого из подходов в чистом виде. Поэтому представляется целесообразным еще раз рассмотреть основные принципы моделирования, причем именно процесса перевода как такового. В последнее время достигнуты значительные результаты в психолингвистическом моделировании процессов языкового понимания. Важно также и то, что в переводе, по-видимому, отражаются некоторые существенные особенности мышления в целом.

Попробуем с этой точки зрения разложить процесс перевода на некоторые мыслительные элементы. Будем иметь, во-первых, необходимость работы со списками — ассоциативная память человека. Во-вторых, способность к разрешению неоднозначности языковых единиц в зависимости от некоторых конкретных языковых же условий. Здесь можно провести аналогию с лабиринтным методом решения задач, или с эвристическим программированием. Наконец, приходится иметь дело с общей проблемой осмысленности, где нужно различать всю задачу целиком, привлекать ассоциативные связи разных уровней мышления. Очевидно также, что при моделировании для перевода целесообразно, описывая процесс целиком, сохранить деление на статику и динамику, разумную, не абсолютизированную независимость между описанием и алгоритмом, разделение на сравнительно автономные участки процесса, с тем, чтобы можно было сделать последний обозримым и доступным для корректировки и пополнения.

ЛИНГВИСТИЧЕСКИЕ ОСНОВАНИЯ МОДЕЛИРОВАНИЯ ПЕРЕВОДА

2.1. Представление о переводе

Отвлекаясь от разных определений перевода¹, примем в качестве некоторого рабочего определения следующее: перевод есть вид человеческой языковой деятельности, в результате которой некоторый текст на одном языке ставится в соответствие тексту на другом языке, при этом обеспечивается их смысловая эквивалентность. Говоря о переводе вообще, мы также будем различать, в тех случаях, когда это необходимо, перевод как процесс и перевод как предмет, т. е. результат процесса.

Теория перевода как особая лингвистическая дисциплина возникла сравнительно недавно, примерно в начале 50-х годов. До сих пор многие лингвисты, в том числе и занимающиеся фундаментальными проблемами перевода, скептически относятся к положению о том, что теория перевода есть самостоятельная лингвистическая наука (см., например, Jäger, 1977). Тем не менее, в значительной мере трудами советских лингвистов обоснована стройная теория перевода, которая в настоящее время получила широкое признание и успешно развивается².

Мы полагаем, что имело место следующее важное взаимодействие наук. В первый период развития машинного перевода сама идея МП произвела чрезвычайно большое впечатление на лингвистов и немало способствовала развитию как языкознание в целом, так и особенно отдельных направлений языкознания — структурной и прикладной лингвистики, математической лингвистики в количественном и неколичественном понимании, сопоставительного языкознания и т. п. Концепт МП убедительно показал неполноту наших представлений о строении и функционировании языка, выявил потребность в разносторонней и полной формализации и объективизации наших знаний о языке, дал яркую перспективу того, для чего такие знания нужны. В дальнейшем, однако, развитие МП в период спада пошло разными путями, но стимул, данный им языкознанию, продолжал действовать. В числе лингвистических наук, стимулированных таким образом, была и теория перевода, которая в это время пережила период становления и ускоренного развития. Достаточно напомнить, например, что в это время было сформулиро-

вано представление о переводе на уровне языковых средств и переводе-интерпретации, вошли в научный обиход концепции глубинных структур, смысловых инвариантов, языка-посредника и т. п. В настоящее время мы являемся свидетелями того, как этот стимул работает в обратном направлении, а именно: результаты, полученные в теории традиционного, немашинного перевода, помогают вновь и более эффективно ставить проблемы перевода машинного.

Чрезвычайно важно, что в теорию перевода вошло и укрепилось представление о моделировании. Теоретики «человеческого» перевода в большинстве случаев принципиально отмежевываются от машинного перевода; однако тем интереснее выводы, к которым они приходят, когда оказывается, что многие положения общей теории перевода могут быть применены в машинном переводе. Принцип моделирования используется в теории перевода не для построения модели как таковой, а для осознания и анализа детальной методикой происходящих в человеческом мозгу явлений, конструирующих процесс перевода.

Совершенствуя качество перевода в рамках систем второго, т. е. индуктивного, подхода, представляется необходимым и целесообразным обращать внимание на то, как именно перевод производит человек, какие последовательности мыслительных операций складывают этот процесс и приводят к желаемому результату.

Чрезвычайно важным является вопрос о том, является ли перевод творческим процессом или он носит некоторый рутинный характер. При возникновении идеи МП разработчики существенным образом исходили из последнего, т. е. считали процесс перевода нетворческой деятельностью, в значительной мере характеризующейся поиском по словарю. Современные теоретики перевода склонны признавать принципиально творческий характер процесса перевода, а результат — творчеством. Безусловно, типология перевода имеет к этому самое прямое отношение. Так, никто не будет спорить с тем, что перевод художественного произведения есть творчество, а перевод инвентарных списков творчеством вряд ли является. Однако относительно к этой типологии, а также к разделению на перевод и интерпретацию представляется бесспорным, что в переводе действуют как элементы творчества (иногда обязательные и для перевода инвентарных списков), так и элементы рутинные, присутствующие, может быть, и в переводе поэзии (например, поиск рифм).

Обращаясь к общей теории перевода в поисках более эффективных путей моделирования процесса перевода и изучения последовательности действий человека, прежде всего определим объект и цели моделирования.

¹ См. например, определения перевода в (Бархударов 1975).

² О современном состоянии теории см. (Radó 1979).

2.2. Объект и цели моделирования

В определениях перевода подчеркивается, что важно не только то, «что» выражено, но и «как» выражено³. Перевод принципиально отличается от пересказа, реферата, других способов выражения «той же самой» мысли на другом языке. Эта идея является практически главной в работах теоретиков и практиков «общечеловеческого» перевода, из которых мы рассмотрим только относящиеся к последнему времени⁴. Тот факт, что в подавляющем большинстве работ вопросы машинного перевода принципиально не рассматриваются, делает высказывания о переводе как бы более объективными. Следует также отметить, что хотя разбор сложных случаев трансформаций и неконгруэнтностей занимает большое место в теории перевода, чрезвычайно редко используется понятие глубинной структуры и концепция перехода от поверхностной к глубинной и наоборот. Поэтому первый вопрос: как выражено содержание в переводе? Является ли язык кодовой системой? Действительно, способ (метод) описания знаковой системы в целом и ее отдельных уровней может быть задан с позиции взгляда на язык как на разновидность кодовой системы. Теория перевода пользуется представлением о двуязычной коммуникации, когда процесс перевода описывается как коммуникативный акт. Отметим, что и теоретическая лингвистика, в тех случаях, когда речь идет о способах изучения семантического аспекта языка, обращается к представлениям теории связи⁵. Однако принципиальная разница между языком и кодом заключается в том, что в плане выражения языковой код представляет собой сложную систему разнородных единиц, связанных между собой разнообразными отношениями и способных соединяться в высказывания только по определенным сложным правилам, а в плане содержания информация, составляющая содержание кодовых единиц, не дана заранее, а должна быть получена наблюдением за использованием знака в процессе общения и соотнесения этой информации с реальной действительностью и содержанием других

знаков (Комиссаров 1973, 14). Схема двуязычного коммуникативного акта имеет следующий вид (см. рис. 1).

Согласно этой схеме, процесс перевода распадается на три фазы: 1) коммуникация между отправителем (O) и переводчиком, выступающим в функции получателя (П) исходного сообщения; 2) мена кода исходный язык — переводящий язык, осуществляемая переводчиком, выступающим на этот раз в функции перекодировщика. Эта фаза специфична для перевода; декодируя исходное сообщение, т. е. подвергнув его анализу с помощью правил исходного языка (первого кода), переводчик

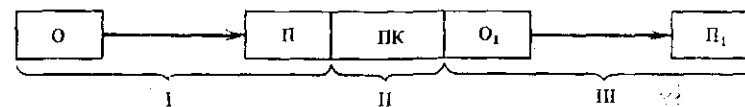


Рис. 1. Схема двуязычной коммуникации при переводе

преобразует информацию в знаки другого кода (переводящего языка); 3) коммуникация между переводчиком, выступающим в роли отправителя (O₁) конечного сообщения и получателем этого сообщения (П₁) (Швейцер 1973).

Кодовые правила, поскольку таковые существуют, здесь нельзя понимать как-нибудь иначе, чем набор переводящих соответствий для данной языковой пары.

Таким образом, хотя язык кодом не является, коммуникативная функция языка и элементы языковой структуры, ответственные за нее, могут быть описаны в терминах теории связи. Для МП это открывает возможность постепенного обнаружения единиц речевой структуры в коммуникации (в письменном тексте).

Следующее важное для МП положение теории перевода заключается в разделении перевода на процесс и предмет.

Речь идет не о тривиальном разделении процесса и предмета (процесс — сама переводческая деятельность, последовательность действий, предмет — результат деятельности, перевод как текст, представленный, например, на бумаге). Для того, чтобы процесс мог быть описан алгоритмом, требуется выделение однозначной последовательности операций, причем нужно, чтобы часть, относящаяся к алгоритму в собственном смысле, т. е. к действиям, была отделена от части, относящейся к описанию, т. е. к некоторой статической информации, а затем последняя была бы дифференцирована, с тем, чтобы на каждой ступени алгоритма было бы четко известно, какая именно информация и в каком объеме используется.

Нельзя сказать, что в современной теории перевода проведено такое разделение. Система соответствий при переводе, как правило, не отделяется от описания действий переводчика. Тем не менее создано достаточно четкое представление о после-

³ «Любой перевод всегда регламентируется тем, что заложено в тексте оригинала и, следовательно, в конечном счете определяется отношениями между исходным языком и языком перевода» (Швейцер 1973, 8). «В отличие от пересказа, перевод должен передавать не только то, что выражено подлинником, но и так, как это выражено в нем. Это требование относится как ко всему переводу данного текста в целом, так и к отдельным его частям» (Рейкер 1974, 7).

⁴ Бархударов 1975; Комиссаров 1973; Швейцер 1973; Рейкер 1974; Ревзин—Розендвейг 1964; Catford 1965; Jäger 1975; Ljudskanov 1972; Kade 1968; Nida 1969 и др.

⁵ «...собственно лингвистика не может судить о том, что происходит в голове человека при говорении и слушании. Для нее ... существенно лишь строение высказывания и той знаковой системы (естественного языка), которая представлена в этом высказывании. В связи с этим вполне естественно воспользоваться понятиями общей теории связи» (Вардуль 1979, 11).

довательности действий, соотнесенных с восприятием содержания сообщения. Так, в концепции уровней эквивалентности В. Н. Комиссарова⁶ существует привязка действий к статической информации следующим образом: порядок перечисления содержательных уровней в основном соответствует этапам анализа. «Начиная с идентификации языковых знаков, Рецептор последовательно уясняет содержание анализируемого текста на уровне высказывания, сообщения, описания ситуации и приходит к пониманию содержания целого на уровне цели коммуникации. Разумеется, подобный поэтапный процесс представляет собой лишь условную схему. Он может реально существовать лишь при «расшифровке» текста, составленного при помощи мало знакомого кода. В обычном процессе коммуникации Рецептор, по-видимому, воспринимает передаваемую ему информацию комплексно, не расчленяя ее по уровням содержания» (Комиссаров 1973, 67).

Естественно, что в процессе машинного анализа допустима лишь последовательность действий, начинающаяся с некоторого начального исходного уровня и постепенно усложняющаяся. Кроме того, поэтапность здесь важна не только для усвоения содержания, но и для анализа формы, что уже относится в большей степени к собственно МП.

Всякий перевод есть процесс, так или иначе связанный с некоторой потерей информации: это вытекает из общей схемы перевода как коммуникативного процесса. При межъязыковом преобразовании, как и при всяком другом виде языковых преобразований, неизбежны потери, т. е. происходит неполная передача значений, выражаемых текстом подлинника (Бархударов 1975, 11). Однако полная передача информации необязательна для осуществления процесса языковой коммуникации. Часто несущественная для понимания информация оказывается связанной с выбором определенных элементов языкового кода; кроме того, неполное понимание части информации является привычным и не вызывает трудностей в общении. Эти особенности перевода можно сопоставить с общими условиями обмена информацией в теории, которая рассматривает обмен информацией как взаимные изменения, происходящие в тезаурусах отправителя и получателя по сравнению с некоторым исходным состоянием этих тезаурусов: абсолютно полного понимания, согласно этой теории, произойти не может, поскольку два тезауруса, участвующие в общении, никогда не могут быть полностью эквивалентными, и речь может идти только о некоторой мере близости.

⁶ «Таким образом, в передаваемом от Источника к Рецептору тексте (отрезке устной или письменной речи) цепочка звуков или букв репрезентирует сложный смысловой комплекс, в котором можно выделить несколько основных уровней плана содержания, а именно: 1) уровень языковых знаков (слов); 2) уровень высказывания; 3) уровень (структуры) сообщения; 4) уровень описания ситуации; 5) уровень цели коммуникации» (Комиссаров 1973, 66).

Таким образом, для МП целесообразно моделировать следующие общие закономерности, которыми описывается перевод с одного естественного языка на другой:

— в переводе должна быть обеспечена эквивалентность передачи не только содержания, но и выражения;

— относительная ценность единиц языкового выражения не дается переводчику заранее, а выясняется, раскрывается по тексту методами, имеющими много общего с декодированием, реконструкцией знаковой системы по результатам ее действия;

— в расшифровке языкового «кода» важно разделение на предмет и процесс.

Неизбежные потери информации при переводе не влекут за собой неосуществимости акта коммуникации.

Концепция моделирования перевода с помощью переводных соответствий должна строиться с учетом этих основных закономерностей.

2.3. Переводные соответствия и «единица перевода»

Несмотря на то, что в теории перевода обобщения производятся к некоторым моделям, описывающим универсальные закономерности двуязычной коммуникации, практически в центре прикладной и иллюстрирующей стороны в подавляющем большинстве случаев выносится вопрос о двуязычных переводных соответствиях. Это и понятно, поскольку трудно себе представить процесс языкового перевода без установления в качестве окончательного результата четких текстовых соответствий для данной пары языков и для данного текста⁷. Вопрос о речевых соответствиях является также неизбежным для любой теории и для любого способа моделирования этой теории в целях перевода, поскольку этот вопрос затрагивает уровень поверхностных структур, миновать которые, если поставлена задача на перевод, а не на другие преобразования над текстами, попросту невозможно.

Переводные соответствия суть соответствия особого рода, принципиально отличающиеся от других сопоставлений (например, в контрастивной лингвистике) тем, что в них сопоставляются не явления, занимающие сходное место в структуре языка, а явления, находящиеся в данный момент в речевом соответствии или вообще регулярно сопоставимые, эквивалентные по выражению содержания данного текста. В результате этого моделирование перевода по принципу установления соответ-

⁷ «Установление отношений эквивалентности в процессе перевода предполагает существование в ИЯ и ПЯ эквивалентных (более или менее равнозначных) единиц. Сам процесс перевода может быть представлен как процесс последовательной подстановки вместо каких-то единиц оригинала эквивалентных единиц ПЯ, т. е. соответствий (Комиссаров 1973, 168).

вий должно базироваться, во-первых, на типологии соответствий, во-вторых — на решении проблемы взаимодействия уровней языка в переводе.

Функциональная классификация переводных соответствий производится в теории перевода по двум основаниям: по степени регулярности и по содержанию собственно переводческих лингвистических действий над текстом.

Примером классификации по степени регулярности может служить типология, введенная В. Н. Комиссаровым. При переводе определенной единицы можно различать постоянные, варианты и окказиональные соответствия. При наличии постоянного соответствия задача переводчика сводится к подстановке такого соответствия в выходном языке вместо соответствующего отрезка оригинала. Для единиц, имеющих варианты соответствия в ПЯ, переводчик выбирает вариант, наиболее точно передающий значение такой единицы в данных конкретных условиях ее употребления. Наконец, в случае отсутствия у данной единицы каких-либо соответствий в ПЯ или невозможности использования их при построении текста перевода, переводчик, учитывая всю дополнительную информацию, передает содержание единицы оригинала совершенно новым способом, пригодным лишь для данного случая.

В классификации по этому основанию имплицитно считается заданной так называемая «единица перевода». Единица перевода представляет собой одно из самых сложных понятий теории перевода. Не будучи в состоянии произвести обзор хотя бы самых общих положений относительно этого понятия, примем к сведению свидетельства теоретиков перевода о том, что отсутствие четкого и признаваемого большинством исследователей определения этого понятия не мешает практическому осуществлению перевода⁸; можно полагать, что есть аналогия с этим положением и с отсутствием общепринятого определения слова, которое (отсутствие) не очень затрудняет практическую работу лексикографа.

Исследователи, считающие возможным рассмотреть понятие единицы перевода, отмечают прежде всего ее варьированность и вследствие этого неопределенность понятия. «Думается, что само понятие «единица перевода» представляет собой противоречие в терминах прежде всего потому, что любая единица — это постоянная величина, тогда как так называемая «единица перевода» по определению является величиной переменной. ... Величина минимальных переводимых отрезков (а последние действительно заслуживают внимательного рассмотрения в теории перевода, хотя и никак не подпадают под категорию единиц) определяется сложным взаимодействием внутрилингвистиче-

ских и внелингвистических факторов, а порой может зависеть и от чисто случайных обстоятельств (например, от расстояния между данным словом и другой единицей, уточняющей его значение)» (Швейцер 1973, 71—72). Л. С. Бархударов отмечает разноуровневость единиц перевода: на каждом уровне — фонем, морфем, слов, словосочетаний, предложений, текста — могут быть свои единицы перевода и в соответствии с этим можно определять перевод как перевод на уровне фонем, слов, словосочетаний, текста и пр. (Бархударов 1975, 174—176). Д. Радо, постулируя творческий характер всякого перевода, в качестве единицы перевода вводит «логему» (Radó 1979), в результате чего возникает некоторое противоречие, ибо логика вроде бы противопоставлена творчеству.

Очевидно, что, с одной стороны, нужна некоторая единица, в терминах которой можно было бы так или иначе описывать переводные соответствия, некоторая мера. С другой стороны, признается, что многоуровневость процесса перевода, гибкость и подвижность переводных соответствий не дают возможности определить указанную величину достаточно точно для того, чтобы можно было пользоваться понятием «единица».

Представляется, что есть некоторый способ упорядочить понятия «единица перевода» и «переводное соответствие», явно находящиеся между собой в тесной связи. Будем рассматривать статику и динамику переводческого процесса. Примем, что есть некоторые исходные данные, описанные в терминах теории перевода, так сказать, переводные исходные данные, и есть некоторый процесс перевода, или алгоритм, который определен над этими данными. Тогда мы можем сказать, что некоторое переводческое описание относится к статике, или к исходным данным, а другое — к процессу, или к динамике. В этом случае можно считать «единицей перевода» (= «минимальным переводимым отрезком») некоторый языковой элемент, участвующий в переводе в качестве исходной информации. Так, мы можем рассматривать единицы перевода, принадлежащие разным языковым уровням и т. п. согласно концепции Л. С. Бархударова. Единицы перевода будут выступать в новом аспекте, когда они участвуют в переводе как процессе. Этот процесс заключается в установлении переводных соответствий в рамках данной пары языков. Переводное соответствие, вообще говоря, элемент непрочный, возникающий только в данном тексте в рамках данной конкретной задачи на перевод. Если единицу перевода можно задать в некотором словаре, или в устройстве, принципиально от словаря не отличающемся, то переводное соответствие есть текстовое образование, его нет смысла задавать в словаре, поскольку от текста к тексту оно может меняться.

Переводное соответствие, таким образом, можно понимать как единицу перевода, взятую в динамике переводческого процесса, реализуемую в условиях конкретного перевода конкретного текста в рамках данной пары языков.

⁸ «Проблемой установления «единицы перевода» продолжают заниматься многие теоретики, но сопоставление различных точек зрения убеждает в том, что для письменного перевода такой проблемы не существует» (Рецкер 1974, 24).

Если единицу перевода можно понимать как некоторую обобщенную переводческую сущность, отнесенную, например, к единице перевода некоторого эталонного языка (рассматривать единицу перевода в рамках одного языка бессмысленно, ведь речь идет о переводе, что подразумевает по крайней мере два некоторых языка, например, один из них язык-посредник или искусственный язык), то переводное соответствие есть единство в рамках двух естественных языков. Как и единица перевода, оно может быть многоуровневым, т. е. затрагивать разные языковые уровни. Кроме того, в последнем случае может не быть уровня соответствия, т. е. например, соотношение уровней в переводном соответствии может быть разнообразным. Например, в трансформационном соответствии единица перевода на синтаксическом уровне может затрагивать также и семантические или морфологические преобразования в выходном языке и т. п.

2.4. Типология переводных соответствий

Я. И. Рецкер выделяет три вида соответствий по следующему признаку: 1) эквиваленты, установившиеся в силу тождества обозначаемого, а также отложившиеся в традиции языковых контактов; 2) вариантные и контекстуальные соответствия; 3) все виды переводческих трансформаций. Между первым видом и остальными есть принципиальное отличие: первые соответствия относятся к сфере языка, остальные — к речи. Основа установления равноценности языковых средств может быть только функциональная, а не формальная. В комплексный процесс перевода вовлечено слишком много разнородных факторов, чтобы можно было устанавливать формальные соответствия на уровне речи. И хотя переводчик вынужден прибегать к помощи логики, психологии, литературоведения, все же единственная опора его работы — текст, и основа лингвистического подхода к нему — функциональные соответствия (Рецкер, 1974, 10).

Эта классификация включает также элементы «окказиональность—регулярность». Эквивалентное соответствие «эквивалентно» всегда⁹. Контекстуальные же и трансформационные соответствия должны специально уточняться на регулярность.

Критерий окказиональности—регулярности может быть реализован не только с помощью некоторых достаточно общих статистических данных, но и с использованием хорошо разработанных методов статистической лингвистики и лексикографии (в том, что касается частот отдельных слов, к которым можно «привязывать» те или иные грамматические явления). Поэтому

⁹ «Там, где между языками установилось традиционное эквивалентное соответствие, переводчик фактически лишен возможности выбора. Отказ от использования эквивалента в исключительных случаях должен быть оправдан особыми условиями контекста или обстановки» (Рецкер 1974, 9).

есть смысл не рассматривать такие соответствия, которые могли бы быть отнесены в разряд окказиональных. Их практическая реализация могла бы быть делом человека-редактора. Гораздо важнее установить классификацию соответствий регулярных, задачу нахождения которых целесообразно возложить на алгоритм МП.

Уточнение объема преобразований при переводе можно начать с разграничения обращений к собственно языковым средствам и к экстралингвистической реальности. Такое разграничение было сделано И. И. Ревзиным и В. Ю. Розенцвейгом при введении понятия собственно перевода и интерпретации. Были рассмотрены два случая. Случай первый: переводчик воспринял некоторую речевую последовательность, от этой последовательности он переходит к ситуации, рассматривает эту ситуацию, затем, полностью абстрагируясь от сообщения, которое было ему передано, а только имея в виду данную ситуацию, переводчик сообщает об этой ситуации другому лицу. Второй случай: переводчик, получив сообщение, переходит от него не к ситуации в действительности, а к языку-посреднику, т. е. к системе соответствий между исходным языком и языком перевода¹⁰. Затем от языка-посредника он переходит к системе ПЯ, с помощью ПЯ он формирует сообщение, которое и информирует адресата о ситуации в действительности.

Первый случай был назван интерпретацией, второй — собственно переводом. Не касаясь представления о языке-посреднике, заметим, что в данном случае перевод понимался как нечто, не выходящее за пределы чисто языковых соответствий.

Деление на перевод и интерпретацию подверглось критике на том основании, что «представление о переводе, осуществляемом человеком как о порождении текста на языке перевода, ... соответствующем тексту на исходном языке... по заданным соответствиям без обращения к действительности или к предшествующему опыту переводчика, явно идет вразрез со всем, что нам известно о реальных процессах обычного, т. е. немашинного перевода. Даже в области синхронного перевода, где, по мнению сторонников рассматриваемой концепции, передача смысла сообщения осуществляется почти без всякого обращения к действительности, на самом деле перевод едва ли может быть осуществлен без того, что в обиходной речи именуется знакомством с предметом» (Швейцер, 1973, 31). Однако следует заметить, что эта критика слишком сурова для некоторого идеализированного случая, который был принят авторами деления на перевод и интерпретацию за исходный. Они тогда же подчеркивали, что в реальном процессе перевода собственно перевод и интерпретация взаимно связаны. На замечание

¹⁰ Система соответствий в виде языка-посредника предполагает более абстрактную постановку задачи, чем в модели МПС. Переводные соответствия в нашей концепции суть лишь основные составляющие модели перевода.

о том, что для переводчика существенно обращение к действительности, «знакомство с предметом» или его предшествующий опыт, можно возразить, что этот предшествующий опыт (даже многолетний и коллективный), а также знание предмета, уже отражены в сложившихся переводных соответствиях, поскольку синхронный переводчик не выдумывает и не подбирает их заново, а лишь извлекает из хранящегося у него в памяти списка. Тем не менее ясно, что если использовать понятия «перевод» и «интерпретация» и далее применительно к переводческому процессу, то требуется их значительное уточнение. Так, по-английски нельзя сказать «муха сидит на потолке», а надо «муха стоит на потолке» (пример взят из Комиссаров 1973, 92), однако представляется, что отсутствие равнозначности при эквивалентном описании денотата не следует называть интерпретацией, а иногда интерпретация понимается именно так. Для моделирования перевода важно, что в данном случае происходит хотя и идеализированное, но размежевание чисто переводческих действий и действий логических, связанных с самостоятельным (не переводческим) анализом и толкованием ситуации.

Типологию переводных соответствий для моделирования можно строить следующим образом.

Эквивалентные соответствия в данной паре языков и для данного подязыка суть такие, которые всегда неизменны в любых контекстных ситуациях.

Выбор таких соответствий иллюстрируем примером. Пусть имеются следующие слова для перевода на английский с русского языка: 'информировать', 'реакция', 'попугай'. На большом объеме исходных текстов отмечено, что нет случаев, когда бы слово 'информировать' переводилось каким-либо другим словом, чем *inform*. При этом не возникает каких-либо отенков значений, которые требовали бы перевода исходного слова другими словами. Перевод слова 'реакция' зависит от подязыка; так, в политической жизни 'реакция' иная, чем в химии или физиологии. Однако в той мере, в какой переводом этого слова будет служить английское слово *reaction*, переводное соответствие здесь можно считать эквивалентным. Слово 'попугай' может означать птицу или болтливого человека: в обоих случаях перевод будет один и тот же, что дает возможность считать и это соответствие эквивалентным. Таким образом, мы можем сказать, что классификацию соответствий целесообразно проводить на основе билингв методом сравнения с эталоном (считая значениями слова его переводы на выходной — эталонный язык), и при этом эквивалентность (неоказиональность) соответствия выступает как некоторый функциональный признак.

Второй тип — варианты соответствия — варианты по разным причинам: стиля, зависимости от синтаксических конструкций, предварительного выбора лексики, прагматических аспектов. Общей характеристикой вариантных соответствий мож-

но считать следующую: любой из вариантов может быть принят, однако при этом будет разница в потере информации. Рассмотрим пример. Пусть дана английская фраза: *Practical training was poor, and the men received little practice in shooting*. Сравнивая два ее наиболее вероятных перевода, А. Д. Швейцер приходит к оптимальному, а именно 'Практические занятия были организованы плохо, и солдаты не получили достаточной стрелковой подготовки' (антонимический перевод для подчеркивания логической связи двух событий, предпочтительный другому: 'Практические занятия были организованы плохо, и солдаты получили недостаточную стрелковую подготовку') (Швейцер 1973, 106). Отметим, что, вообще говоря, 'были организованы плохо' может и не быть точным эквивалентом (например, занятия были организованы хорошо, но их было недостаточно). Машина перевела бы это предложение следующим образом: 'Практическая тренировка была плохой, и солдаты получили мало практики в стрельбе'. Несовершенный стиль не мешает стопроцентно правильной передаче содержания. Отметим, однако, что вариантность здесь не ведет к потере информации.

Третий тип — такие соответствия, которые реализуются с помощью самых разнообразных трансформаций, любая из которых необходима для правильной передачи содержания. Рассмотрим примеры¹¹. *If the funeral had been yesterday, I could not recollect it better* 'Я помню эти похороны так, как будто они были вчера'; *Well, this is a beautiful day, if there ever was one* 'Да, день сегодня действительно великолепный' и пр. Примеры такого рода можно продолжить. Во всех подобных случаях требуется для сохранения содержания сообщения радикально изменить синтаксическую структуру, причем изменения структуры могут быть самые различные. Заметим, однако, что в целом ряде случаев необходимые изменения могут быть поставлены в соответствие определенным грамматическим и лексическим признакам, а также описываться определенными правилами, например, подобно тому, как это происходит при передаче английской пассивной конструкции на русский язык. Многие изменения структуры могут также быть локализованы при лексике.

Окказиональность, регулярность и необходимость обращения к экстралингвистическим факторам в случае трансформационных соответствий может определить информант на этапе составления модели.

Приведенная общая схема классификации переводных соответствий может служить основой для построения прикладной модели.

В теории перевода уделяется мало внимания типам переводимых текстов и типам перевода. Рассмотренные работы теоретиков перевода содержат лишь упоминания об особых видах перевода, например, переводе дипломатических текстов или «ин-

¹¹ Примеры заимствованы из (Комиссаров 1973).

формативном» переводе научно-технических текстов. Между тем, например, соотношение между переводом и интерпретацией, как было отмечено, зависит от типа текста: в художественном тексте интерпретация преобладает, в научно-технической она минимальна. Однако это весьма общее соображение. Видимо, проблемы типологии переводимых текстов должны входить в частные теории перевода; однако последние, по большому числу определений, занимаются вопросами перевода для конкретных пар языков. Машинный перевод существенно зависит от типов соответствий, поэтому типология соответствий весьма важна для моделирования МП. Соотношение между переводом и интерпретацией есть лишь начальный элемент такой типологии: известно, что и в техническом тексте есть элементы интерпретации, а в художественном может быть собственно перевод. **Особое отношение к переводу** проявляется, в частности, в том, что в научно-техническом переводе, в отличие от художественного, выражение собственных мнений, предпочтений и толкований со стороны переводчика является делом запрещенным, как, видимо, и при других видах перевода, например, дипломатических документов. В художественном, поэтическом и пр. видах перевода творческое участие переводчика неизбежно. Таким образом, типология переводных соответствий должна разрабатываться и далее; однако в ее развитии может помочь опыт машинного моделирования, которое возможно на основе уже имеющихся предпосылок.

2.5. Переводные соответствия и передача смысла

Переводные соответствия должны обеспечить передачу содержания текста на одном языке в текст на другом языке так, чтобы при этом сохранился некоторый смысловой инвариант. Можно ли обеспечить передачу смысла на уровне речевых соответствий?

Соответственно концепции Г. Фреге будем различать сообщение (иния), референт (денотат) и смысл. Человек-переводчик в процессе перевода, как можно предположить, в состоянии воспринять и передать все три компонента. Кроме того, он использует также и то, что Фреге называет «представлением». Разница между оригинальным предложением и его переводом на другой язык не должна превышать представление¹² (см. рис. 2).

Не касаясь вопроса о том, может ли машина мыслить, т. е. может ли она также использовать все три компонента при переводе с одного языка на другой, отметим, что наибольшую труд-

ность представляет формализация двух элементов треугольника, а именно «сообщения» и «смысла». Аргументы о невозможности машинного перевода базируются на том, что эти элементы нельзя в принципе формализовать, или это невозможно сделать до создания искусственного интеллекта. Так, М. Таубе отрицает теоретическую возможность машинного перевода, связывая это с невозможностью установления формальных соответствий между единицами всех уровней двух знаковых систем. При этом он утверждает, что нельзя заставить машину, как человека, переводить аппроксимативно, из-за того, что возможности машинной памяти и ее быстроедействие ограничены.

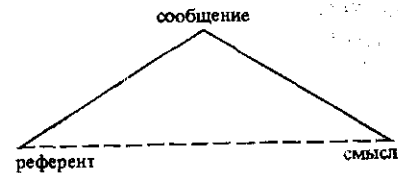


Рис. 2. Отношение между смыслом, референтом, сообщением

Положительную аргументацию о возможности моделирования перевода на ЭВМ и передачи компонентов языкового значения с помощью речевых соответствий можно строить на отрицательных доводах противников МП; наиболее четко сформулированные доводы были выдвинуты именно 10—15 лет назад.

«Допустим, что способности человека переводить интуитивно соответствует способность машины переводить аппроксимативно. Здесь аппроксимация — это строгий математический процесс приближения к пределу, который в принципе может никогда не достигаться. Например, если смысл определяется контекстом, можно построить автоматический словарь, в котором запоминаются не отдельные слова, а сочетания из двух-трех слов, фразы, целые предложения и т. д. Если, как и считает большинство современных логиков, основным носителем содержания служит предложение, а блоки памяти современных вычислительных машин имеют объем на 45 порядков ниже требуемого для составления словаря из предложений и если в общем память вычислительных машин растет со скоростью один порядок в десятилетие, то вычислительную машину с требуемым объемом памяти можно ожидать через 450 лет» (Таубе 1964, 43).

Человек действительно переводит в значительной мере интуитивно, однако при этом творческий процесс соотносится с рутинным; никто не будет спорить с тем, что эквивалентные соответствия не так уж редки в речи. Смысл действительно определяется контекстом, и поэтому можно построить словарь, содержащий большое количество исходных языковых элементов, а также достаточное количество подробных правил, позволяющих по формальным признакам контекста выявлять значение (или значимость) каждой языковой единицы. Тот факт, что основным носителем содержания служит предложение, можно оценивать

¹² «...Можно усматривать три степени различия между выражениями (словами, словосочетаниями и целыми предложениями): различие затрагивает либо только представление, либо смысл, но не денотат, либо, наконец и смысл, и денотат... Разница между переводом и оригиналом не должна, вообще говоря, выходить за пределы первой степени различия» (Фреге 1977, 187).

по-разному, в зависимости от того, как мы определим слово «содержание». Слово также является основным носителем, или составным элементом «содержания», поэтому наряду с теоретически и практически возможным «словарем предложений» целесообразно иметь также и словарь слов. Возможности памяти и быстроедействие современных ЭВМ таковы, что можно иметь довольно обширные «словари предложений», включающие единицы разной длины. Таким образом, мы видим, что в этой части аргументации М. Таубе не остается убедительных доводов против «аппроксимативной» методики осуществления МП.

Рассмотрим следующий содержательный аргумент М. Таубе. «Другой «практический» подход к проблеме машинного перевода состоит в попытке составить программу вычислительной машины, способную определять синтаксическую структуру предложения путем анализа не контекста, а таких характеристик предложения, как порядок слов в нем, окончания слов, структура фразы и т. п. Стоит заметить, что этот подход содержит порочный круг, так как более разумно предположить, что части речи, порядок слов, окончания и вообще синтаксический характер частей предложения определяется содержанием предложения, а не наоборот — не содержание предложения определяется синтаксическими свойствами его частей. В обычных учебниках грамматики анализ предложения проводится для того, чтобы показать, как выражен смысл, а не для установления того, каково содержание. Без предварительного понимания содержания предложения такого грамматического анализа проводить нельзя. Например, мы исходим из понимания смысла данного слова, когда говорим, что оно является подлежащим предложения. Мы не смогли бы узнать, какой символ служит подлежащим предложения, если бы не знали значений символов» (там же, 44).

Заметим, что такие признаки, как порядок слов в предложении, окончания слов, структура фразы, тоже могут пониматься как составляющие контекст, и в таком случае противопоставление этих средств контекстным будет неправомерно. Правильной является мысль о том, что если противопоставлять «синтаксический характер частей предложения» его содержанию, то возникает порочный круг. Тогда действительно без предварительного понимания содержания предложения грамматического анализа проводить нельзя. Однако если мы не будем противопоставлять синтаксис и семантику в определении частей речи и членов предложения, а будем рассматривать эти два аспекта в тесной связи и в их манифестации в контексте, то тогда такого порочного круга возникнуть не будет. Попутно заметим, что содержание предложения и его синтаксические составляющие М. Таубе представляет себе в традиционных членах предложения.

Порочный круг возникает не в этой ситуации, а в том случае,

если мы попытаемся содержательно определить, что представляет собой смысл. Если пытаться определить, что есть смысл, то приходим к тому, что смысл описывается через перевод¹³. Смысл предложения можно описать как то, что бывает усвоено, когда понято предложение, или как то, что имеют общего два предложения в различных языках, если они правильно переводят друг друга» (Чёрч, цит. по Ревзин, Розенцвейг 1964, 68). Перевод определяется как такая операция над двумя предложениями разных языков, в результате которой неизменным остается некоторый смысл каждого из этих предложений, такой, что смысл первого предложения равен смыслу второго. С другой стороны, смысл есть то, что остается инвариантным при переводе, т. е. как мера эквивалентности двух предложений, если известно, что одно является переводом другого. Понятие языка-посредника при этом не меняет ничего, так как языком-посредником может быть естественный язык.

Раз смысл определяется через перевод, то тогда, зная, что два предложения переводят друг друга, мы с помощью выявления переводных соответствий можем достаточно точно описать способы передачи смысла. Точнее, способы переводных соответствий тогда и есть способы передачи смыслового содержания. Трудность машинного перевода в этом случае заключается в том, чтобы найти каждый раз достаточно формальное соответствие.

Из положения о том, что смысл определяется через перевод, следует и некоторое методическое заключение. Методически целесообразно при этом исходить от параллельных текстов. Сравнением предложений и их частей можно обеспечить постепенное выявление смысла этих частей и затем смысла всего предложения. Машина в принципе может обращаться и к смыслу, и к референту, однако первое осуществимо в той степени, в какой формализована семантика, а второе — настолько, насколько предусмотрено использование экстралингвистических сведений. Видимо, в принципе деление на перевод и интерпретацию и не имеет здесь решающего значения: просто обращение к смыслу без использования референта может быть осуществлено более простыми средствами.

¹³ «Интуитивное понятие смысла лучше всего определять именно через перевод... Как же отразить инвариантность смысла в теории перевода? Мы уже говорили, что каждому выражению ставится в соответствие некоторая совокупность единиц языка-посредника. Отсюда основное требование к языку-посреднику: его единицы должны соответствовать смысловым единицам, выделяемым в обоих языках (как в ИЯ, так и в ПЯ). Таким образом, язык-посредник и инвариантность смысла должны определяться одновременно и взаимно уточнять друг друга по мере привлечения нового языкового материала. В соответствии со сказанным выше мы будем понимать под смыслом некоторого выражения совокупность элементарных смысловых единиц языка-посредника, поставленных в соответствие с данным выражением» (Ревзин, Розенцвейг 1964, 68).

Выше приведена цитата из М. Таубе относительно отсутствия взаимно-однозначных соответствий между единицами естественных языков. Действительно, трудно утверждать, что всегда есть формальное однозначное соответствие между единицами двух языков хотя бы на одном из языковых уровней. Если бы такое соответствие было, то язык не отличался бы от кода. Дело, однако, в том, что отличия языка от кода известны, и типы неоднозначности могут быть более или менее полно описаны. Так, редко какое слово естественного языка полностью соответствует по объему и содержанию значений слову другого языка (хотя есть и такие случаи соответствия, например, термины). Однако для каждого многозначного слова мы можем задать более или менее полный перечень его сопоставимых с другим языком значений и дать правила выяснения соответствий в зависимости от контекста. Мы не можем привести перечень однозначно совпадающих предложений, однако совпадение по смыслу предложений может быть достигнуто в некоторых пределах совпадением по смыслу отдельных частей предложения.

Таким образом, с некоторой методической и рабочей точки зрения замкнутый круг определения смысла через перевод с помощью ряда допущений на самом деле дает возможность разрешить проблему. Здесь следует подчеркнуть, что целью настоящего раздела является не исследование фундаментальных проблем смысла, значения и референции, а поиск рабочей гипотезы о передаче смысла.

2.6. Языковой знак и перевод

В общей теории перевода, если судить по рассмотренным работам, не возникает потребности изучать языковой знак более подробно, чем в его двух сущностях, материальной и идеальной, и в аспекте конституирования двух планов: плана выражения и плана содержания. Однако в целях моделирования перевода на ЭВМ целесообразно рассмотреть основные характеристики машинного семиозиса.

Наиболее сложной формой знака является символ, служащий основным звеном семиозиса во всех аспектах сознательной деятельности человека и в различных формах его социального общения. Наиболее типичными примерами символов являются значимые единицы естественного языка — слова, их формы, словосочетания, предложения. Лингвистический смысл включает четыре компонента: имя (материальный носитель информации), денотат (предмет или явление действительности, обозначаемое именем), десигнат (смысл, понятие о предмете или явлении), а также коннотат, охватывающий дополнительные экспрессивно-оценочные значения. В искусственных информационно-семиотических системах возникает и используется искусственный знак (ИЗ), представляющий собой соединение имени (означающего) с его десигнатом. Переработка текста естест-

венного языка на ЭВМ затрагивает соотношение лингвистического знака и искусственного знака.

Согласно концепции, развиваемой Р. Г. Пиотровским, четырехкомпонентная структура символа в ИЗ свертывается в двухчленное построение «имя — десигнат». Роль имени (означающего) здесь выполняет материальный носитель информации (цифра, буква или другой символ), а в качестве десигната выступает значимость этой буквы, цифры или символа в данной математической системе или какой-либо другой искусственной семиотике. В машинном знаке роль означающего выполняет закодированная цепочка букв, а функцию означаемого — лексико-грамматическая информация, записанная в сопровождающем



Рис. 3. Передача сообщения в информационной схеме «человек—машина—человек»

цепочку машинного документа. Этот десигнат, записываемый вручную или программно, соотносит машинное имя (означаемое) с той искусственной семиотической системой лингвистических знаний, которая была введена в ЭВМ. Таким образом, номинативная функция в искусственном машинном языке ограничивается десигнативными отношениями между машинным означаемым и означающим. Лингвистический символ, являющийся продуктом ассоциативно-эвристического мышления человека, оказывается несравненно богаче математического знака.

Передачу сообщения в схеме «человек—машина—человек» можно рассматривать в виде комбинации двух информационных процессов. Первый процесс предусматривает передачу текстового сообщения от человека к машине, второй — от машины к человеку (см. рис. 3).

Каждый из двух процессов включает по четыре блока. Блоки первого процесса следующие: 1.1) порождение текста отправителем сообщения. Человек порождает некоторые лингвистические символы, каждый из которых включает все четыре компонента: имя, денотат, десигнат и коннотат; 1.2) разрушение каждого из этих символов: путем многократного перекодирования в память ЭВМ вводится лишь цепочка импульсов, соответствующая означающему (имени) текстового символа, что же касается денотата, коннотата и десигната этого символа, то они, естественно, в компьютер не попадают; 1.3) машинный семиозис, в результате которого происходит как бы восстановление разрушенного знака. Принятая машиной последовательность импульсов сравнивается с записанными в машинных словарях цепочками, пока не произойдет полного совпадения текстовой цепочки со словарной цепочкой. При таком совпадении

лексико-грамматическая информация словарной цепочки передается текстовой цепочке, которая тем самым получает свой десигнат и превращается в машинный ИЗ. Восстановленный таким путем ИЗ является обедненной моделью исходного речевого символа, поскольку машинный документ не может вместить все дополнительные лексико-грамматические значения и коннотативные оттенки; 1.4) преобразование внутри ЭВМ введенного текста, например, машинный перевод, автоматическое индексирование, реферирование и т. п. В этом блоке происходит новое разрушение знаков текста, которые превращаются в цепочку импульсов.

Блоки второго процесса следующие: 2.1) и 2.2) — выдача компьютером машинных результатов и чтение этих результатов адресатом Ч₂. Здесь с ИЗ происходит то же, что и в блоке 1.4. В блоке 2.3) — семиозис машинного сообщения — человек — приемник сообщения, опираясь на свои знания предмета и лингвистические навыки, восстанавливает у полученных от ЭВМ цепочек их десигнативные и денотативные значения. Оживающие знаки обрастают также коннотативными оттенками. Блок 2.4.) реакция человека₂ на полученное сообщение.

Таким образом, согласно данной концепции, в ходе переработки текста в ЭВМ дважды происходит разрушение и восстановление знака: первый раз при передаче сообщения от человека к машине и второй раз при передаче машинных результатов от компьютера к человеку. Эффект работы системы зависит, во-первых, от объема содержательной информации исходного сообщения, который удастся восстановить живому приемнику, и, во-вторых, от степени готовности массового читателя пользоваться услугами системы. Минимизации информационных потерь можно добиться лишь тогда, когда в десигнаты машинных знаков вкладываются те лексические и грамматические значения, которые с достаточно большой вероятностью могут считаться существенно необходимыми для правильной формализации и переработки входного сообщения (Пиотровский, 1975).

Данное представление о машинном семиозисе может быть взято за основу для моделирования процесса перевода в теории переводных соответствий. Представляется, однако, что нужна некоторая модификация этой концепции, связанная с тем, что в рамках приведенного описания в течение какого-то времени (в блоке 1.2) машинный знак выступает без озаглаемого. Видимо, более соответствующей реальному процессу была бы модель, в которой знак естественного языка преобразуется в искусственный знак таким образом, что план содержания первого не исчезает, а совмещается с планом выражения. Если новое разрушение знаков произойдет в блоке 1.4, то вряд ли станет возможной машинная переработка текстов, например, перевод, для знаков, у которых утрачены даже десигнаты. Видимо, основные компоненты языкового знака не утрачиваются необратимо. Бедность десигната ИЗ может быть в значительной мере

устранена благодаря увеличению объема памяти, отведенного под словарь, и лучшей организации информации, хранимой в словаре: здесь нет каких-либо принципиальных ограничений. Преимущества данной схемы перед другими концепциями машинного знака заключаются в ее технологичности и в том, что эксплицируются основные ограничения машинного семиозиса. Положительный ответ на вопрос «может ли машина мыслить?» сопряжен для инженерного подхода с точной оценкой ограничений машины в смысле таких возможностей.

Согласно психолингвистической модели языкового восприятия по Реймолду основными являются три принципа: 1) единичей анализ при восприятии речи является предикативная синтагма; предложение воспринимается (анализируется) синтагма за синтагмой, причем после прохождения границы синтагмы сохраняется только ее семантическое содержание, а синтаксическая структура стирается; 2) при восприятии речи мы имеем одновременный доступ к семантическим и синтаксическим свойствам сигнала, т. е. анализ синтаксиса и семантики предложения идет параллельно; 3) при понимании высказывания используются знания о мире. Учет этих принципов является основной чертой, отличающей данную модель от традиционных трансформационных моделей, которые оперируют с полной синтаксической структурой целого предложения и в которых синтаксический анализ предшествует семантическому.

Традиционное понятие предикативной синтагмы пересматривается в связи с тем, что требуется использовать понятие сверхоперативной памяти, которая может одновременно хранить приблизительно пять сигналов. Синтаксическая структура синтагмы стирается сразу же после завершения ее семантической интерпретации. Синтаксический и семантический анализатор действуют параллельно. Входная синтагма подвергается последовательной обработке, в результате чего устанавливается три вида семантической репрезентации: предварительная, промежуточная и окончательная.

Стратегии установления семантических связей приписывают семантические роли различным частям предложения. Связи устанавливаются как внутри синтагмы, так и между синтагмами. Связи внутри синтагмы устанавливаются с помощью некоторого набора стратегий, связи между синтагмами устанавливаются без обращения к полной синтаксической структуре любой из обработанных предшествующих синтагм. Вводится стратегия временной последовательности, устанавливающая сравнительную хронологию событий, описываемых в терминах последовательности синтагм. Мы не приводим здесь более подробного описания механизма создания семантических представлений, см. (Reimold 1975).

Перевод начинается с понимания составляющих высказывания последовательно на разных уровнях. Однако в переводческом понимании процесс осознания содержания идет параллельно

но с попытками передать это содержание на выходном языке, он происходит методом проб и ошибок, имея принципиально другую ориентировку, а именно — не только понять, но и передать содержание на другом языке. Можно считать, что процесс понимания неотделим от процесса перевода. Доказательством этому служат действия переводчика при переводе, описание которых мы цитируем:

«Возьмем предложение: Oil immersed transformer windings are now usually made with paper covered conductors.

Анализ («пассивная грамматика») предполагает, ...что для каждой словоформы дан полный перечень ее функций, т. е. грамматических значений. Берем первое слово oil, оно может быть подлежащим, переводим слово в именительном падеже: «масло». Второе слово immersed тогда придется рассматривать как сказуемое (данная словоформа вполне может быть глаголом в прошедшем времени). Получаем: «масло погрузило», третье слово переводится тогда как дополнение: «масло погрузило трансформатор». Следующее слово также является существительным, поэтому предшествующее не может переводиться как существительное. Переделываем перевод: «масло погрузило трансформаторную обмотку». Однако следующее слово are показывает, что сделанные раньше предположения о функциях всех слов неверны, т. к. are обязательно является сказуемым или частью сказуемого: между тем, для этого сказуемого нет подлежащего. Таким подлежащим может быть лишь слово windings. Итак, имеем: «трансформаторная обмотка делается». Это заставляет предположить, что immersed не глагол, а причастие (такая функция также имеется в первоначальном перечне значений) и должно переводиться словом «погруженная». Отсюда следует, что слово oil зависит от этого причастия и должно переводиться косвенным падежом. Получаем окончательный перевод этой части: «Погруженная в масло трансформаторная обмотка делается»... Окончательным этот перевод назван потому, что дальше мы не встречаем слов, которые свидетельствовали бы о том, что наш анализ произведен неверно. Так же анализируется и вторая часть предложения. Разберем теперь аналогичный немецкий пример:

Die zusätzliche Produktion kann nicht losgelöst von den Planaufgaben in Angriff genommen werden, wie es in der Vergangenheit vielfach geschah.

Ключом к переводу является правильное выделение подлежащего и сказуемого. Подлежащее в главном предложении выделяется легко (Die Produktion). Сказуемое явно сложное. Первая часть его (kann) стоит, как и полагается, на втором месте. Но какова его вторая часть? Если считать, что это genommen werden, то перевод, как это нетрудно убедиться, получится бессмысленным. В данном случае необходимо знать, что in Angriff nehmen представляет собой единство, имеющее значение «приступить к чему-либо, начинать что-либо». Посмотрим, однако,

нет ли в предложении какой-либо другой формы глагола, которая также может служить частью сказуемого. Рассмотрим с этой точки зрения причастие losgelöst. Оно по формальным признакам могло бы входить в сказуемое (die zusätzliche Produktion kann nicht losgelöst werden). Однако в этом случае, во-первых, указание на объект (von den Planaufgaben) не могло бы стоять после losgelöst, а, во-вторых, часть предложения in Angriff genommen werden повисла бы в воздухе. Таким образом, мы установили, что losgelöst не входит в состав сказуемого. Значит, оно образует самостоятельную группу losgelöst von den Planaufgaben («изолирование от плановых заданий»)» (Ревзин, Розенцвейг 1964, 202).

Рассматривая данную картину действия переводчика с точки зрения заложенных в основу действий логических и предметных посылок и самой последовательности действий, отметим, что перевод здесь базируется на представлении о пассивной грамматике (анализ) и активной (синтез). Переводчик работает методом проб и ошибок. Интуитивно представляется, что в реальном процессе перевода переводчику не понадобится перебирать такое большое число вариантов, среди которых так много ошибочных; однако можно себе представить, что это кажущееся быстрое действие переводчика на самом деле является следствием наличия некоторых стереотипов и незначительным временем поиска среди них. Процесс перевода осуществляется движением слева направо, однако неизбежны рекурсии, связанные с необходимостью понимания связанных словосочетаний (синтагм). Каждая словоформа рассматривается как обладающая некоторым набором возможных синтактико-семантических функций, выраженных в категориях частей речи и членов предложения. Фактически при каждой словоформе возникает некоторая возможность продолжения предложения по разным дорожкам. Эта возможность реализована в известном предсказуемом анализаторе синтаксической структуры; препятствием к практическому применению этого анализатора является его чрезвычайная громоздкость, огромные затраты машинного времени и трудность фильтрации вариантов с целью отбора одного из них для перевода. Когда мы рассматриваем действия переводчика с точки зрения порядка или системы установления переводных соответствий, то обращают на себя внимание следующие обстоятельства:

— анализ синтаксической и семантической структур высказывания происходит одновременно. В переводе это есть анализ по частям речи и членам предложения;

— переводчик не восстанавливает предварительно полной синтаксической структуры предложения;

— анализ и синтез происходят по некоторым отдельным осмысленным кускам («единицам перевода», синтагмам и т. п.);

— при анализе и синтезе происходит постоянная проверка

на осмысленность, в ходе которой анализируемая синтагма сравнивается с предыдущими и последующими.

Сопоставим эти действия с моделью языкового понимания. Можно видеть, что здесь хорошо видны этапы предварительной, промежуточной и окончательной семантической репрезентации. Переход от одной к другой осуществляется методом проб и ошибок: при этом существенно используются правила сочетаний слов в синтагме, которые можно уподобить правилам конкатенации. Хорошо иллюстрированы также правила семантические: так, мы убеждаемся, что масло не может погрузить трансформаторную обмотку. Здесь привлекаются также и экстралингвистические сведения. Понимание предложения в целом из синтагм обеспечивается действием стратегии установления семантических связей. Однако в отличие от языкового восприятия в общем смысле понимание для перевода одновременно есть перевод: переводчик понимает через перевод.

Весьма сложным, однако, остается вопрос о возможности полного и исчерпывающего моделирования действий переводчика на ЭВМ, поскольку каждый раз человек осуществляет проверку осмысленности, чего не может делать ЭВМ. Очевидно, что машина будет «осмысленно» переводить в том случае, если диапазон введенных в нее переводных соответствий достаточно обширен, используется по четким правилам и охватывает достаточно полный круг явлений, чтобы обеспечивать необходимые переводческие трансформации.

Смысл рассматривается, таким образом, как некоторая дискретная категория или свойство высказывания в целом, которое возникает из отдельных элементов высказывания и конституируется в процессе перевода и самим процессом перевода. В этом нет философского противоречия, так же как нет противоречия и в абстрактном определении смысла через перевод.

Таким образом, в моделировании перевода на основе переводных соответствий в качестве аксиомы можно принять, что процессы понимания и перевода осуществляются одновременно.

2.7. Технология и методика построения модели

Модель должна оптимизировать соотношение между лингвистическим описанием и алгоритмом работы с ним. Абсолютная независимость алгоритма от описания и передача составления алгоритма математикам невозможны и нецелесообразны. С другой стороны, полное слияние описания и алгоритма, хотя оно технологически возможно и может дать хорошие инженерные результаты (например, высокая скорость перевода), практически нецелесообразно в связи с тем, что при этом затрудняется осуществление самого главного принципа приближенных вычислений, а именно принципа совершенствования системы по результатам перевода, пополнения словарей и таблиц и т. п. Следовательно, модель должна обеспечивать некоторое опти-

мальное соотношение между независимостью и зависимостью описания и алгоритма (механизма) работы с ним для целей МП.

В отличие от чисто дедуктивного подхода идея приближенных вычислений позволяет строить семантический компонент модели постепенно, идя от текстовых проявлений семантических параметров и от некоторого исходного минимального семантического представления, необходимого для разрешения неоднозначностей принципиального характера (для начального цикла вычислений) ко все более полному набору семантических признаков, вводимых по мере возникновения в них необходимости, обусловленной требованиями разрешения все более сложных случаев неоднозначности и организации трансформационных преобразований.

Можно считать, что методы, с помощью которых при МП можно организовать постепенное обнаружение основных свойств «кода» — элементов языковой системы — взаимообусловлены и создают в совокупности особую методическую теорию, имеющую значение для названного моделирования — теорию контекстной детерминации. Со времен Соссюра и Ельмслева целые школы занимаются описанием знаковых и текстовых структур и рассматривают их в значительной мере как единственные и основополагающие факторы языкового выражения и языковой системы. В каждом языке много явлений, таких, как полисемия, омонимия, и пр., которыми можно доказать, что многие или все «значимые» элементы сами по себе не означают ничего определенного и только в связи с другими, тоже многозначными элементами, приобретают определенное значение, по мере их детерминации контекстом. Однако примеры, показывающие, как с ростом контекста значение языкового элемента постепенно уточняется детерминациями, с логической точки зрения все же не доказывают, что отдельный элемент не обладает никаким значением. Уточнение детерминацией происходит только между элементами, которые сами по себе обладают более общим значением, чем в совокупности с другими элементами того же класса. Понятие тем конкретнее, чем большим числом понятий общего характера оно определяется, детерминируется. Каждое контекстное значение возникает за счет вариации, уточнения — в более общем виде детерминации — со стороны других элементов контекста, которые, в свою очередь, могут детерминироваться данным элементом (Panzer 1965).

В теории контекстной детерминации можно выделить две составляющие. Первая — тип восприятия контекста. Методическая сущность теории детерминации в том, что контекст воспринимается не только как последовательность символов, из которых каждый снабжен некоторой исходной информацией, извлеченной из словаря, но и как иерархически упорядоченная совокупность уровней информации, на каждом из которых существуют свои опорные точки для анализа. Эти уровни не возникают сами собой, а должны быть сконструированы (реконструир-

рованы) по тексту методами реконструкции знаковой системы.

Вторая составляющая теории контекстной детерминации — типология контекстных зависимостей. Зависимость от контекста неоднородна и оценивается дифференцированно для каждой единицы. Это положение удобно иллюстрировать контекстными зависимостями лексики, хотя оно не относится исключительно к последней. Так, есть слова, лексические значения которых полностью зависят от контекста, и есть слова, абсолютно от контекста не зависящие, хотя это видно не сразу и априорно, а лишь в некотором исходном конкордансе. Тип зависимости от контекста играет решающую роль для построения алгоритма анализа языковых явлений, который работает в условиях взаимодействия языковых уровней. Именно теория детерминации определяет порядок и следование уровней формализации, во взаимодействии с которыми языковые уровни должны обеспечить экспликацию языковых явлений, определяющих перевод.

Составляющие теории контекстной детерминации, особенно вторая, имеют смысл только применительно ко всему спектру языковых явлений, подлежащих переводу. Поясним это на примере типологии контекстных зависимостей. Если мы определили, что такое-то многозначное слово имеет такую-то степень зависимости от контекста, которая описывается определенным набором средств, то это еще ничего не дает для возможности инженерного описания и применения этого описания к разрешению лексической многозначности. Однако если в пределах всего корпуса текстов для всех отмеченных в этом корпусе многозначных слов проведена классификация типов контекстных зависимостей, это позволяет строить специальный аппарат разрешения лексической неоднозначности уже с определяемой эффективностью. Комплексное — на всем рассмотренном корпусе слов — решение проблемы делает также возможной и эффективной стыковку решений с алгоритмами, производящими другие виды обработки, так или иначе связанные с лексическими. Точно так же грамматика подязыка может быть задана списком грамматических конструкций.

Дифференцированность зависимости от контекста позволяет принимать статистические решения для определенного числа случаев слабой контекстной зависимости, где от контекстной детерминации возможно перейти к общему значению. Здесь возникает естественный для прикладных методов вопрос о соотношении статистических и детерминистских методов.

В реконструкции языка как знаковой системы статистические методы дают лишь самые общие результаты; к числу задач, решаемых на уровне статистики звуков, букв и графем, а также двух-, трех- и четырехбуквенных сочетаний, относятся, например, разбивка слитного текста на слова, установление типа языка и т. п.¹⁴ Можно согласиться с Р. Г. Пиотровским в том,

что «широко проводимые сейчас работы по статистике и комбинаторике букв, буквосочетаний и слогов дают мало сведений о механизме текстообразования. Гораздо большую ценность для лингвистики текста и инженерной лингвистики представляют статистические исследования в области комбинаторики значащих единиц текста» (Пиотровский 1975, 181). Комбинаторика значащих единиц, которая, в частности, описывает совместную встречаемость двух или более элементов текста, может служить основанием для следующего шага, а именно определения степени и характера детерминации, существующей между совместно встречающимися элементами. Здесь можно квалифицировать детерминацию по содержательному признаку.

Детерминация проявляется, таким образом, как в количественной, так и в качественной стороне. Статистические сведения о комбинаторике значащих единиц текста (например, слов и сочетаний слов) играют определенную роль в оценке целесообразности того или иного вида анализа применительно к данному языковому явлению.

По Б. Панцеру, можно различать три вида детерминации. Первый — автоматическое сочетание двух элементов — состоит в том, что за одним элементом с необходимостью следует другой (так, если говорящий по-латыни употребил слово *sine*, он должен затем употребить аблатив, у него нет выбора). Здесь имеет место абсолютная детерминация. Второй вид детерминации — возможность сочетания со многими элементами без изменения значения. Третий — возможность сочетания со многими элементами с изменением значения; например, латинское *consulere amicum* означает 'просить совета у друга', а *consulere amico* — 'заботиться о друге'. Здесь падежная функция релевантна для понимания значения.

В первом случае мы имеем управление и функциональную релевантность, во втором — свободную иррелевантность и в третьем — функциональную релевантность. Распределение по этим трем рубрикам следует за установлением сочетаемости с учетом идентичности значения сочетания. Таким образом, наблюдение за употреблением в языке, которое позволяет выделить все формальные типы, все же недостаточно. На вопрос о релевантности можно ответить только после определения вида и степени зависимости. Это означает, что мы приходим к типу релевантности только после определения вида и степени взаимозависимости. Последний определяется эмпирически (Rappaport 1965).

Массовость явления служит показателем, с помощью которого можно ориентировочно определить цену затрачиваемых усилий, что важно для приближенных методов, а также величину относительной ошибки. Статистические характеристики элементов, вступающих между собой в отношения детерминации, также позволяют ранжировать алгоритмы анализа, базирующиеся на этих элементах.

¹⁴ Примером работ такого рода может служить исследование о статистической структуре немецких слов (Menzerat — Meyer-Eppler 1950).

МОДЕЛЬ МАШИННОГО ПЕРЕВОДА НА ОСНОВЕ ПЕРЕВОДНЫХ СООТВЕТСТВИЙ

3.1. Общая структура модели

Суммируем требования, которым должна удовлетворять общая структура модели.

Есть основания моделировать переводческую деятельность человека целиком, т. е. от распознавания слов (поиск в словаре) до синтеза выходного предложения и выходного текста. Модель должна, таким образом, иметь последовательность этапов анализа и синтеза, каждый из которых базируется на некоторой исходной информации и дает в качестве выхода информацию следующей ступени. Работа следующего этапа основывается на данных предыдущего плюс некоторая исходная информация, необязательно являющаяся выходными данными предыдущего этапа, например, словарная или табличная. Алгоритм перевода представляет собой лингвистическую операцию преобразования разноуровневых языковых данных, объединенную цельностью замысла и меняющейся от этапа к этапу информацией. Варианты анализа, если таковые возникают, должны оцениваться в пределах тех этапов, где это происходит, и на выход этапа должен выдаваться один вариант. Иерархия соответствий должна позволять строить единую систему преобразований, ведущих к цели, т. е. переводу в рамках данной пары языков.

Будем представлять себе модель перевода на основе переводных соответствий (МПС) в виде некоторого автомата, имеющего два основных состояния: генерации и трансляции. Состояние генерации необходимо по двум причинам. Во-первых, модель должна быть создана изначально на базе исходного набора языковых данных и переводных соответствий для данной пары языков. Во-вторых, при установке на переводные соответствия для каждого подязыка внутри данной языковой пары необходима более точная настройка системы, ее словарей и алгоритмов анализа и синтеза на специфику данного подязыка в соответствии с основными законами типологии научно-технического перевода. Этап генерации характеризуется преобразованием систематизации элементов описания: слов, единиц перевода, основных закономерностей преобразований. Этот этап заключается в создании предметного наполнения системы языковым материалом. Этап трансляции характеризуется собственно работой системы на языковом материале по осуществлению перевода как такового. Здесь также есть элементы пополнения словарей и алгоритмов системы, однако они вторичны

по отношению к основной операции этого состояния, т. е. перевода.

В соответствии с этим будем различать в модели два основных компонента: предметный и динамический. Предметный компонент отвечает на вопрос о том, что должно быть переведено, динамический — как должен быть осуществлен перевод.

В состав предметного компонента входят два основных элемента: собственно предметный и переводной. Составляющие предметного элемента — лексика, грамматика, семантика. Каждая составляющая представляет собой множество элементов,

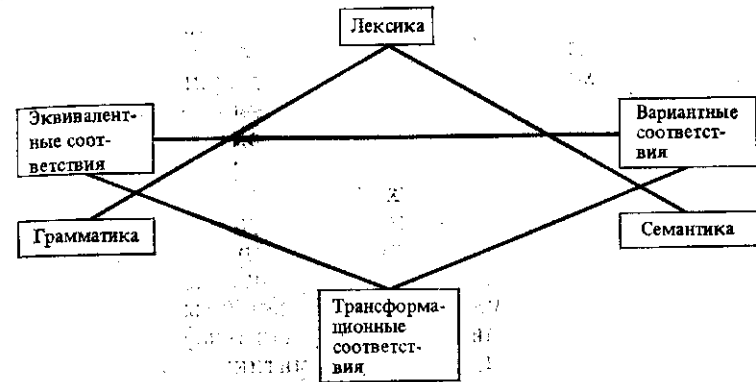


Рис. 4. Структура предметного компонента модели МПС

связанных между собой. Лексика есть собрание (список) слов данного языка (подязыка). Грамматика есть собрание (множество) грамматических признаков данного подязыка, семантика — набор семантических (или семантико-синтаксических) признаков. Способ задания элементов в каждом множестве есть перечисление. Таким образом, составляющие предметного элемента предметного компонента суть образующие текст.

Переводной элемент предметного компонента образуется из типов переводных соответствий. Он содержит межъязыковые соответствия, также заданные в форме перечислений. В соответствии с принятой нами типологией переводных соответствий будем различать соответствия эквивалентные, вариантные и трансформационные. Эквивалентные соответствия устанавливаются всегда простой операцией соотнесения: так, в результате поиска входного слова в списке слов словаря может быть сразу получен переводной эквивалент этого слова, или установлена другого сорта инвариантность и т. п. Вариантные соответствия суть такие, которые устанавливаются с помощью некоторого формализованного выбора. Трансформационные соответствия образуют самый сложный вид соответствий, для их нахождения требуется работа сложных алгоритмов анализа, синтеза и межъязыковых преобразований.

Предметный элемент предметного компонента образует исходную структуру компонента. На эту исходную структуру накладывается сеть, образованная переводным элементом. Одна в целом предметный компонент отражает предмет, статическое описание сущности исходных данных, над которыми производится операция перевода. Структура предметного компонента изображена на рис. 4.

Динамический компонент модели МПС отвечает на вопрос, каким образом должен быть осуществлен перевод. Содержимое этого компонента является процесс установления соответствий, динамика нахождения переводных соответствий в исходном тексте и конструирования некоторого выходного текста, находящегося со входным в отношении коммуникативной эквивалентности. Для осуществления такой динамики нахождения соответствий динамический компонент содержит особого рода инструментарий, который отражен в его структуре. Динамический компонент состоит из двух основных элементов: алгоритмического и программного. Алгоритмический элемент включает собственно алгоритм, трансляционную грамматику, словарь. Алгоритм, который подразделяется на общий алгоритм, организующий систему в целом, и частные алгоритмы (этапов, частей этапов, отдельных решений анализа и синтеза), управляет использованием трансляционной грамматики и словаря на соответствующих стадиях нахождения соответствий и построения эквивалентного текста. Трансляционная грамматика есть специфическая бинарная грамматика, которая оптимальным для перевода образом организует грамматические признаки, нужные для установления соответствий. Сюда входят признаки всех языковых уровней: морфологического, синтаксического, семантического (или семантико-синтаксического) и лексического. Наконец, словарь представляет собой упорядоченное для перевода собрание лексических единиц с необходимой информацией. Это собрание лексических единиц упорядочивается по трем основаниям: по признаку входного/выходного языка, однозначности/многозначности слов, фразеологических единств/отдельных лексем.

Программный элемент представляет собой совокупность программных решений для этапов, назначенных и описанных в алгоритме. Название «программный» для лингвистического обеспечения носит условный характер, но оно необходимо, так как здесь мы выходим на уровень машинной реализации системы, закономерности которой нужны для разработки собственно лингвистического обеспечения. Последнее можно пояснить примером использования системы стандартных операторов: этот язык программирования, предназначенный для лингвистов, определяет возможности и границы применения определенного набора приемов лингвистического анализа. Составляющие программного элемента: списки, схемы лексического анализа и перевода, схемы грамматического анализа. Каждая из этих со-

ставляющих отличается от другой все большим усложнением программно-лингвистического аппарата: так, списки (сюда входят также списки и таблицы для синтеза) реализуют наиболее соответствующие типу словарей перевода однозначных слов и т. п.; схемы лексического анализа — перевод сложных сочетаний и многозначных слов — характеризуются элементарным набором стандартных операторов и моделей перевода; схемы грамматического анализа (куда входят также схемы, организующие перестановки слов в предложении и пр.) используют уже максимально полный набор

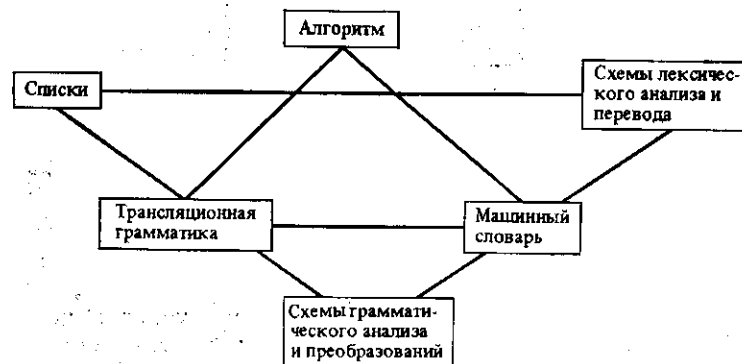


Рис. 5. Структура динамического компонента модели МПС

стандартных операторов. В анализирующей части составляющие программного элемента динамического компонента могут рассматриваться в аналогии с типом переводных соответствий: так, списки реализуют соответствия эквивалентного типа, схемы лексического анализа — варианты, схемы грамматического анализа — трансформационные соответствия.

Структура динамического компонента изображена на рис. 5.

Два компонента модели МПС по-разному взаимодействуют в разных режимах работы системы. Будем различать генерацию системы, т. е. режим, при котором создается предметный компонент для работы на текстах некоторого нового подязыка, и эксплуатацию системы (трансляцию).

В обоих режимах используются оба компонента, однако порядок работы их меняется. В режиме генерации работа происходит согласно рис. 6, т. е. первым вступает в действие предметный компонент, который обеспечивает создание необходимых для модели составляющих и элементов. Физически объем языковой информации, необходимый для создания модели в целом, может быть представлен в виде конкорданса, составленного на материале достаточного количества исходных текстов. После

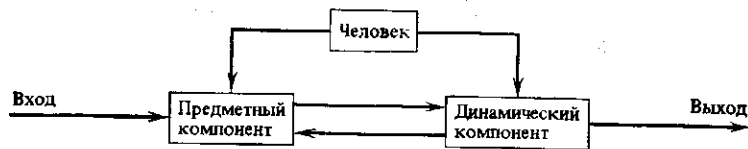


Рис. 6. Генерация системы МТ по модели МПС

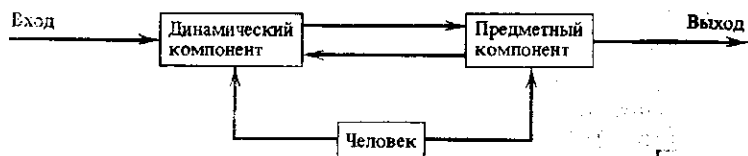


Рис. 7. Работа модели в режиме трансляции

того, как модель с ее двумя компонентами создана, работа системы осуществляется по схеме рис. 7.

Режим трансляции, в соответствии с вышесказанным, характеризуется тем, что здесь осуществляется собственно перевод. В этом процессе в качестве постредактора и/или корректора системы участвует человек. Однако его участие не носит принципиального характера и в некотором идеале система может работать без человека. Режим трансляции изображен на рис. 7.

3.2. Общие принципы алгоритма анализа и синтеза

Общие принципы анализа и синтеза определяют подход к моделированию и основную структуру алгоритма в целом.

Основной единицей перевода, для которой определяются все необходимые для перевода параметры, является предложение, под которым будем понимать либо простое предложение целиком, либо составные части сложного (составного) предложения. Однако известно, что целый ряд контекстных признаков, детерминирующих разрешение неоднозначности, может находиться вне пределов простого предложения. Поэтому анализ производится в общем случае не в пределах предложения, а в границах всего текста, введенного в оперативную память ЭВМ. Таким образом, лингвист (и соответственно алгоритм) имеет в своем распоряжении программные средства, позволяющие выходить за рамки простого предложения и осуществлять поиск детерминирующих признаков вне его пределов, в границах всего текста. Нет необходимости говорить, что этим обстоятельством моделируется существенная и необходимая часть анализирующей деятельности человека, производящего перевод.

Синтактико-семантический анализ производится с помощью привычных для человека синтаксических функций в рамках, не

выходящих за пределы традиционной системы «части речи — члены предложения». Напомним, что анализ переводческой деятельности человека, произведенный нами в предыдущей главе, показал, что это именно те категории, которые необходимы и достаточны для определения основных семантико-синтаксических ценностей при переводе с одного естественного языка на другой. Поэтому в центре исследования ставится не нахождение полной синтаксической структуры предложения, а определение синтаксических функций слов предложения, что представляется наиболее важным и необходимым для установления переводных соответствий человеком и моделирующей его действия машиной.

Единицей перевода может выступать отрезок, меньший, чем предложение, например, синтагма. Если ставить целью точное определение границ синтагм, то эта задача, как показал обширный опыт, недостижима, в рамках имеющихся методов синтаксического анализа. Во-первых, она связана с полным синтаксическим анализом: чтобы сказать, что мы можем членить предложение на синтагмы без какого-либо неопределимого «остатка», мы должны знать полную синтаксическую структуру. Во-вторых, членение на синтагмы предполагает разрешение практически всех видов неоднозначности. Поэтому мы определяем синтагму как некоторое единство, в рамках которого устанавливается переводное соответствие: коль скоро соответствие найдено, установление границ синтагм не производится.

Определим переводное соответствие, поскольку в нашем случае, как видно из вышесказанного, оно играет чрезвычайно большую и решающую роль в анализе.

Переводным соответствием (актуализацией единицы перевода) будем называть пару, образованную отрезком входного текста и отрезком выходного текста, между которыми существует отношение коммуникативной эквивалентности.

Отсюда следует, что понятия синтаксической структуры, синтагмы и пр. являются вторичными по отношению к понятию «переводное соответствие». Анализ направлен не на установление характеристик строя языка безотносительно к переводу, а на установление соответствий для перевода, причем в этом нахождении соответствий используются разнородные сведения. Так, переводное соответствие мы можем установить не только тогда, когда мы нашли полную синтаксическую структуру и границы синтагмы, в которой участвует анализируемая нами, например, лексическая единица. Мы можем говорить о переводном соответствии и тогда, когда мы определили лексический эквивалент данной единицы, нашли в тексте сочетающиеся с ней слова, установили основные параметры согласования сочетающихся слов, получили все необходимые синтаксические и семантические характеристики всех участков данного соответствия.

Из самой идеи установления переводных соответствий ясно, что перевод играет практически основную роль. Если в ситуации, когда мы занимаемся определением полной синтаксической структуры предложения, перевод как таковой может служить завершающим актом или он вообще не важен, поскольку может быть заменен передачей смысла в какой-либо другой форме, в МПС перевод играет роль критерия и поэтому должен быть достаточно рано включен в процесс нахождения соответствий. Это вытекает также и из действий человека-переводчика, который достаточно рано включает в работу метод проб и ошибок — проб на перевод и корректуру результатов перевода. Поэтому в системах, базирующихся на переводных соответствиях, последовательность этапов анализа и синтеза примерно следующая:

- поиск по словарю, идентификация лексических единиц;
- работа словаря оборотов. Исключение из дальнейшего анализа тесно связанных фразеологических единств, грамматический анализ которых не нужен вследствие наличия строго фиксированных переводных эквивалентов;

- грамматический анализ наиболее существенных явлений входного текста. Определение основных синтаксических функций слов предложения. Определение основных грамматических параметров (падежи, время, вид, наклонение, род и пр.) для слов основных лексико-грамматических классов;

- перевод однозначных и многозначных слов. Перевод однозначных слов производится по таблице однозначных соответствий. Перевод многозначных слов может производиться разными способами, в том числе и с помощью контекстного анализа, в ходе которого используется уже имеющаяся к настоящему этапу информация, полученная в ходе работы других этапов;

- грамматический анализ и синтез. Этот этап относится к той части работы, которая может быть сделана только после того, как станут известны переводные эквиваленты, т. е. здесь для полного нахождения переводного соответствия требуется двуязычная ситуация в максимально возможном объеме. Ряд преобразований, направленных на установление, например, трансформационных соответствий, также может быть осуществлен на этом этапе. Пример: перестановка отрицания в переводе с английского на русский язык;

- синтез выходного текста. Здесь происходит окончательное оформление текста в выходном языке, включая морфологический синтез словоформ, грамматический синтез (например, аналитических категорий) и синтаксический синтез предложения и фразы в целом.

Особенностью принятого синтаксического анализа является окончательное разрешение лексико-грамматической омонимии (омографии) на начальной стадии анализа. Этим модель МПС отличается от других способов анализа, в которых омонимия

обуславливает сохранение многих вариантов до окончательного результата (многодорожечный предсказуемый анализатор, метод фильтров и т. п.).

Трансляционная грамматика рассматривается как арсенал средств, аппарат, предназначенный для раскрытия грамматической структуры переводимого текста с целью нахождения соответствий.

Общая структура грамматических средств следующая:

- диагностирующая конструкция;
- детерминанта;
- контекстуальная модель перевода;
- элемент контекстуальной модели перевода.

Диагностирующая конструкция есть некоторый типовой контекст данного языкового элемента, который может быть описан достаточно произвольным образом: последовательность классов слов, окружающих данное слово или словосочетание, совокупность формальных признаков вместе с символами классов слов, и т. д. Диагностирующая конструкция есть наиболее общее понятие, являющееся результатом применения теории контекстной детерминации к задаче постепенного раскрытия грамматических и лексических свойств текста для перевода.

Детерминанта есть некоторое уточнение понятия «диагностирующая конструкция», она выступает как более четкая составляющая последней. Под детерминантой мы понимаем лексико-грамматический класс или подкласс, список слов или отдельное слово, синтаксический, морфологический признак или морфему, которые по отдельности или в некоторой совокупности однозначно определяют одно значение многозначного слова как требующее отдельного лексического эквивалента в выходном языке. Детерминанта является продуктом применения теории контекстной детерминации к наиболее важному для МП уровню — уровню слов и словосочетаний, к лексике.

Контекстуальная модель перевода далее конкретизирует механизм использования контекстных средств в обнаружении соответствий. Контекстуальная модель перевода выступает как полный набор операций, алгоритмических действий, которые производятся после того, как найдена детерминанта и следует осуществить перевод. Модель состоит из элементов, каждый из которых описывает действия с одной детерминантой или их типовой совокупностью.

Моделирование действий человека требует такого количества основных символов, которое бы наиболее оптимальным образом могло быть использовано оперативной памятью (см. выше о психолингвистическом моделировании языкового понимания). Поэтому классификация основных используемых системой языковых элементов должна насчитывать небольшое количество классов и иметь иерархическую структуру. В ранних алгоритмах машинного перевода, носящих экспериментальный характер, иногда применялись обширные классификации, мо-

жет быть, удобные для формального описания, но чрезвычайно неудобные для человека-составителя и пользователя (см., например, Молошная 1960). В нашей классификации поэтому принято небольшое число лексико-грамматических классов слов в качестве основного деления словаря, с достаточным количеством подклассов, определяющих дистрибутивно-статистические свойства лексики, релевантные для поиска переводных соответствий.

Особое значение имеет грамматический анализ. Под грамматическим анализом мы понимаем всю совокупность процедур, направленных на установление грамматического строя текста с точки зрения нахождения этому строю грамматических эквивалентов, необходимых и достаточных для оформления переводного соответствия в рамках грамматических норм языка выходного. Термин «грамматика» означает при этом морфологию, синтаксис и семантику, последнюю в тех пределах, которые неотделимы от синтаксиса (например, главные члены предложения). Семантика как набор специфических признаков, дополняющий грамматику и существующий в описании и алгоритме отдельно от нее, образует особую составляющую предметного компонента. Грамматический анализ является определяющей частью системы, поскольку именно он определяет ее эффективность и разрешающую способность.

Моделирование действий человека-переводчика привело нас к мысли рассмотреть возможности традиционной системы «части речи — члены предложения». При этом важно установить, насколько полно эта система учитывает, или может учесть, формальные признаки.

Рассмотрим кратко общее развитие грамматической традиции в части формальных признаков, а также систему «части речи — члены предложения».

В основе русской грамматической нормы до 20-х годов лежало представление о своеобразии и национальной неповторимости русского языка, основанное на работах фортунатовской школы, наиболее ярко представленной в исследованиях А. М. Пешковского (Пешковский 1959). Школьная грамматика и стилистика Пешковского строились на описании соотношения форм, например, определение и определяемое. Для этой грамматики характерна полнота представления: можно сказать, что в ней были описаны почти все явления русского языка. С развитием русского языка как языка межнационального общения происходит переориентирование русской грамматики на межнациональный норматив, т. е. по сути дела на универсальные семантические категории, присущие не только русскому языку, но и другим языкам. Русская грамматическая структура стала рассматриваться с позиций принадлежности к общезыковой грамматической картине. Эти представления восходят к грамматическому учению Л. В. Щербы, центром которого является система формальных признаков, которая по существу от-

вергает парадигматическую организацию категорий (Щерба 1958). Система формальных признаков отражает равноуровневость признаков, и тем самым она дает возможность построить межнациональный грамматический норматив. Так, по тексту можно вывести систему формальных признаков, например, морфологических, для такого флективного языка, как русский, и затем соотнести эту систему с системой формальных признаков для любого другого языка: становятся возможными межъязыковые корреспонденции соотношения. В свою очередь система формальных признаков Щербы восходит к концепции пяти членов предложения Буслаева, выделявшего среди членов предложения подлежащее, сказуемое, определение, дополнение, обстоятельство. Парадигматические признаки были использованы в структуре частей речи, синтаксические — членов предложения.

Отношение частей речи и членов предложения у Буслаева было уникальным, поскольку в других грамматиках, кроме немецкой (В. Беккер) такого соотношения не было. Практическая приемлемость системы Буслаева объясняется тем, что ей было свойственно понимание генеративного развития системных свойств речи. Эта система вошла в школьный норматив, и русский язык является единственным языком, систематически освещенным с позиции частей речи и членов предложения.

Члены предложения суть универсалии, поскольку в каждом языке универсально представлены пять членов предложения. Когда переводчик производит перевод, сопоставляя текст с текстом, он осуществляет разбор предложения, рассматривая слова с точки зрения грамматической семантики. При этом семантика предложения членится им в терминах членов предложения. Все примеры на перевод даже весьма сложных случаев и ситуаций, которые содержатся в работах по теории перевода убедительно показывают, что переводчик не имеет никакой необходимости заниматься при переводе изучением механизмов словопроизводства, парадигматики или синтаксиса в объеме, превышающем систему членов предложения. Даже перевод с китайского, в котором нет частей речи и соответствующей парадигматики, производится таким образом, что переводчик выделяет в китайском предложении члены предложения и дальнейший анализ производит от них.

Таким образом, есть основания утверждать, что члены предложения являются универсальными семантическими эквивалентами в области грамматики, данными по тексту. Эти универсальные семантические эквиваленты могут служить удовлетворительной основой для организации сопоставительного анализа двух языков с целью нахождения переводных соответствий¹.

¹ «В переводческом сопоставлении языки не выступают в качестве равноправных систем, каждая из которых должна сначала изучаться как самостоятельное целое, а затем и сравниваться с другим языком в качестве единой системы. При сопоставлении, имеющем целью описание переводческих отношений, исходной точкой анализа служит язык оригинала. Задача анализа за-

Таблица 1

	Подлежа- щее	Сказуе- мое	Дополнение	Опреде- ление	Обстоя- тельство
Предложение	+	+	—	—	—
Глагол+связка	—	+	+	—	—
Наречие	—	+	—	—	+
Местоимение и существи- тельное	+	—	+	—	+
Прилагательное	—	—	—	+	—

Часть речи может быть членом предложения или его частью. Часть речи как парадигматический класс входит в разные парадигмы. В некоторой парадигме часть речи может быть редуцированной формой члена предложения. Употребление части речи в данной парадигме означает редуцирование парадигм парадигматического класса. Такое употребление части речи называют основным употреблением части речи.

Идеальный текст представляет редуцированную форму членов предложения и редуцированную форму парадигматических классов, когда некоторая часть речи выступает в основном употреблении, а некоторый член предложения редуцирован до одного слова некоторой части речи. Эта корреляция между уровнями представляет собой использование свойства языка, которое формулируется так: максимальная единица нижележащего уровня равна минимальной единице вышележащего уровня. Системная корреляция может быть осуществлена между всеми единицами данного и другого уровня, с наибольшим удобством между частями речи и членами предложения. Существует следующая схема корреляции между частями речи и членами предложения — см. табл. 1.

Здесь знак «+» обозначает обязательное наличие корреспонденции отдельного слова и члена предложения в соответствующем значении, данном в правом столбце матрицы. Знак «—» означает либо отсутствие такой корреспонденции, либо ее неинверсальный характер, так как корреспонденции могут быть представлены не во всех языках (Рождественский 1969).

Для целей автоматического анализа и синтеза, как неоднократно указывалось ранее в многочисленных работах по машинному переводу и смежным проблемам, традиционная классификация нуждается в коррективах вследствие того, что мно-

ключается в нахождении в текстах перевода отрезков, используемых для передачи значения единиц, которые выделяются в ИЯ, в изучаемых оригиналах. Таким образом, речь идет фактически не о сопоставлении систем двух языков, а об описании системы ИЯ в терминах системы ПЯ» (Комиссаров 1973, 197).

гие релевантные машинному анализу свойства классов слов остаются за ее пределами вследствие ориентированности на человека. Особое значение в машинном анализе приобретают числительные, местоимения, знаки препинания. Мы не будем здесь обосновывать установившееся в МП (в рамках индуктивного подхода и инженерной лингвистики) разделение числительных и местоимений согласно их синтаксическим функциям между существительными и прилагательными; в каждой из этих частей речи числительные и местоимения образуют особые разряды. Знаки препинания, роль которых связана с установлением границ простого предложения, также выступают в машинном анализе в качестве самостоятельных слов. Естественно, что в традиционном представлении о частях речи и членах предложения такая роль знаков препинания не учитывается.

Поэтому в целом система частей речи и членов предложения обычно преобразуется в некоторую сходную систему синтаксических функций и лексико-грамматических классов слов. Так, для англо-русского анализа эмпирическим, индуктивным путем получено распределение, приведенное в табл. 2.

Таблица 2

	Подле- жащее	Сказу- емое	Опреде- ление	Допол- нение	Обстоя- тельство	Пред- ложная	Раздели- тельная
Существительное	+		+	+			
Прилагательное	+	+	+				
Глагол	+	+	+		+		
Наречие					+		
Предлог						+	
Союз							+
Знак препинания							+

Система синтаксических функций (СФ) использует преимущества традиционной системы «части речи — члены предложения» в сочетании с необходимой формализацией существенных для переводного анализа вспомогательных элементов синтаксического строя предложения (приписывание СФ знакам препинания, союзам, знакам разделительным и т. п.)².

² «В основу настоящего алгоритма положена такая гипотеза, которая не противоречит основным нормам «традиционной» грамматики, поскольку в этом случае грамматические сведения, которые будут получены с помощью автоматического лингвистического анализа, можно будет использовать для совершенствования и дальнейшего развития грамматической теории, не дублируя определенные начальные сведения о строении языка» (Бельская 1969, 26).

3.3. Принцип системной организации динамического компонента

Рассмотрим некоторые примеры, которые на первый взгляд свидетельствуют о трудности или даже невозможности машинного перевода. Пример Т. Винограда: *The fish was bought by the cook. — The fish was bought by the river* 'Рыба была куплена поваром' — 'Рыба была куплена у реки' (Winograd 1977). Здесь требуется научить машину различать значения предлога *by* как вводящего деятеля и обстоятельство места. Пример на перевод пассивной конструкции: *The air raids came at dusk and one by one the towns were blacked out as air raid sirens wailed and anti-aircraft guns went into action.* 'Воздушные налеты начались с наступлением темноты, и города один за другим погружались в темноту под вой сирен воздушной тревоги и грохот зенитных орудий' (Швейцер 1973, 108). Пример, которым Бар-Хиллел мотивировал принципиальную невозможность полностью автоматизированного высококачественного МП: машина не может различить значения следующих двух выражений, в каждом из которых употреблено многозначное слово: *The box is in the pen. The pen is in the box.* Если даже принять, что слово *box* однозначно и означает, например, 'коробка' или 'ящик', то при многозначности слова *pen* ('перо' — 'детский манеж') два эти предложения, взятые вместе или отдельно вне контекста, понять невозможно (кстати, не только машине, но и человеку). Тем не менее, для каждого из приведенных примеров нетрудно представить последовательность однозначных действий над некоторыми исходными данными, в результате которой будет получено приемлемое переводческое решение задачи. Так, если начать с последнего примера, можно себе представить словарь, в котором каждое из этих слов будет занимать четкое место в иерархии значений и неоднозначности в их сочетании возникать не будет. Далее, трудно предположить, что в реальном дискурсе не найдется достаточно формальных признаков, чтобы эти значения дифференцировать. В примере на пассивную конструкцию также существуют последовательные действия, позволяющие получить перевод (пример подобных действий приводит А. Д. Швейцер на с. 110). Наконец, в примере Т. Винограда проблема перевода предлога *by* может быть решена проверкой на одушевленность следующего существительного.

Таким образом, можно отдельно изучать вопросы анализа и синтеза для перевода различных ситуаций, от простых до самых сложных, и каждое такое изучение приводит к интересной и самостоятельной методике и к определенной системе лингвистических доказательств. Для каждой ситуации можно найти определенный путь решения. Однако совершенно ясно, что механическое соединение найденных таким образом решений не даст универсальной и всеобъемлющей системы МП потому, что каждый раз для поиска пути решения исходные позиции — ко-

ординаты отсчета — были своими собственными. Противоречие, таким образом, заключается в том, что для каждого трудного и сложного случая существует алгоритмическое решение — однако системы МП не существует, такой, которая могла бы удовлетворительно решать все эти (и другие) трудные случаи.

Очевидно, что алгоритмическое «машиннопереводческое» решение лингвистических проблем перевода возможно и целесообразно только с системных позиций. Под системной организацией будем понимать:

— распространение проанализированных явлений на достаточно широкое языковое поле. Ясно, что «модельное» использование и небольшой объем взятых в рассмотрение явлений не может обеспечить удовлетворительного перевода даже для ограниченной языковой области;

— охват системой всего процесса перевода, от ввода текста до его вывода после работы программ синтеза;

— отказ от индивидуальных окказиональных преобразований, которые хотя и обеспечивают максимальную эквивалентность передачи данного конкретного содержания, в то же время факультативны относительно некоторого более простого варианта, удовлетворительно передающего коммуникативную нагрузку.

Системная организация модели МПС определяется структурой предметного компонента, в которой система соответствий в статике наложена на сеть, образуемую предметным элементом. Системная организация динамического компонента должна обязательно предусматривать возможность пополнения и изменения элементов, входящих в компонент. Система МП в целом должна обладать свойством открытости.

Положение об открытой организации системы МП требует самостоятельного рассмотрения.

Многие системы МП, построенные по индуктивному принципу в начальном периоде, не получили развития вследствие того, что не был удовлетворительно организован процесс пополнения и совершенствования этих систем. При реализации инженерного подхода к задаче МП возникает ряд трудностей, связанных с тем, что накопление информации естественно идет от текстовых ситуаций разнообразного плана. При этом происходит накопление не только общих случаев, но и достаточно индивидуальных текстовых употреблений, затемняющих общую картину. Если учесть при этом разноуровневость релевантных переводу явлений и отсутствие закономерностей в большом количестве случаев, подлежащих анализу, то можно представить, что индуктивная система без продуманного механизма накопления, сортировки, экстраполяции и элементов обучения не может эффективно действовать. Именно такими и было большинство ранних систем МП, принципиально использующих переводные соответствия и индуктивную методику. Они не были принципиально рассчитаны на восприятие изменений.

Принцип открытой организации системы МП должен предусматривать:

— возможность учитывать новые текстовые особенности употребления в их классификации в рамках модели;

— сортировку новых данных в алгоритме такую, чтобы эти данные могли быть учтены при формировании новых алгоритмов или частей алгоритмов;

— постоянный приоритет наиболее массовых случаев над индивидуальными при сохранении и учете индивидуальных особенностей употребления;

— возможность динамической корректировки всех элементов обоих компонентов модели.

Для обеспечения открытой организации компонентов модели МПС с целью создания возможности повышения точности вычислений применен принцип разделения алгоритмов анализа (схем) на зоны.

При индуктивном выявлении основных элементов языковой структуры, релевантных переводу (реконструкции знаковой системы), используется информация морфологическая, лексическая, синтаксическая (порядок слов, следование одних элементов за другими) и семантическая. Естественно принять последовательность выявления элементов в направлении от частного к общему, иначе в противном случае не исключается возможность неправильного определения характеристик. Так, если некоторая неоднозначность разрешается в пересечении синтаксических классов, то до выяснения наличия такого пересечения должны быть выявлены и проанализированы случаи, когда упомянутая неоднозначность определяется другими признаками. Так, омография словоформы *presents* сначала сужается рассмотрением окончания *s*; это означает, что если морфологический признак мы будем рассматривать после анализа окружения в терминах, например, классов слов, то не исключена возможность ошибочного решения. Другими словами, действует принцип, который в других ситуациях именуется принципом *the longest match* (наибольшего возможного совпадения). Введем понятие «диагностирующая конструкция». Будем называть так типовую контекстную ситуацию, определяющую выбор искомой характеристики по тексту. В соответствии с этим принципом порядок выявления диагностирующих конструкций³ целесообразно строить следующим образом:

— комбинированные диагностические конструкции (ДК) с опорой на лексический элемент. Комбинация может состоять из любого числа элементов других уровней, в том числе морфологических;

— комбинированные ДК с опорой на морфологический элемент;

³ Слово «модель», иногда употребляющееся в таком смысле, нежелательно, поскольку в нашем изложении относится к общей концепции.

— морфологические ДК;

— лексические ДК. Последовательность в отношении этих двух типов ДК может быть произвольной; сначала можно использовать лексические, затем морфологические ДК.

Перечисленный набор ДК образует первую зону, отделенную от последующих. Отделение зон необходимо для организации пополнения ДК с тем, чтобы не менять всей последовательности ДК, образующей схему.

Вторая зона состоит из ДК синтаксического характера. На первом месте находятся комбинированные ДК с опорой на синтаксические элементы, однако с учетом морфологических и лексических признаков. Замыкают зону собственно синтаксические ДК.

Третья зона состоит из семантических признаков. По структуре она также начинается с комбинированных ДК и завершается семантическими ДК в собственном смысле. Под ДК здесь понимается сочетание семантических признаков, релевантных в первую очередь различению синтаксических омонимов.

Индуктивное построение набора семантических признаков может быть начато с выявления тех из них, без которых не могут быть осуществлены трансформационные соответствия или разрешение синтаксической омонимии. В той мере, в какой отдельные признаки не могут существовать самостоятельно, они образуют некоторую минимальную семантическую систему, которая, в свою очередь, составляет исходную совокупность параметров.

Рассмотрим в качестве примера вышеизложенного схему определения синтаксической функции (СФ) «определение» системы АМПАР для существительных. Схема имеет вид (блок-схема), изображенный на рис. 8. Можно видеть, что в схеме использованы следующие содержательные признаки (проверки): лексико-грамматические классы слов (существительное, глагол); семантические подклассы (числительное, союзное, местоименное, имя собственное); СФ (однородность, «определение»); отдельные слова и списки слов, знаки препинания.

Применение принципа разделения схемы на зоны в данном случае дает следующую картину.

Первая зона (комбинированные ДК с опорой на лексический или морфологический элемент, морфологические и лексические ДК): блоки 30—31.

Вторая зона (комбинированные ДК синтаксического характера, с лексическими и морфологическими признаками, собственно синтаксические ДК): блоки 24—25—26, 21—22—23, 15—16—17—18—19—20.

Третья зона (комбинированные ДК с использованием синтаксических и семантических признаков, а также лексических): блоки 1—2—3—4—5—6—7—8—9—10—11—12.

Каждая зона отделяется от остальных массивом запасных

ние. Например, в случае с рассмотренной схемой определения СФ существительного должны быть выписаны цепочки существительных, в пределах которых некоторые слова должны получить СФ «определение». Массив исходных примеров должен возможно полным образом характеризовать специфику текстов;

— классификация примеров согласно общему представлению о роли данного признака в системе признаков, релевантных переводу, в соответствии с положениями трансляционной грамматики. В каждой типовой ситуации должно быть дано определение желаемому признаку: например, каждое слово в типовой цепочке существительных должно получить содержательную классификацию с точки зрения релевантных категорий трансляционной грамматики;

— общее определение роли данного признака в системе категорий трансляционной грамматики. Например, для СФ «определение» существительного должна быть произведена оценка возможностей и отношений этой СФ к другим СФ существительного и к СФ остальных частей речи;

— назначение ДК для типовых ситуаций, отраженных в исходном массиве;

— исследование ДК с точки зрения выявления комбинированных случаев пересечения и омонимичности с ДК других родственных или связанных явлений и схем;

— упорядочение ДК по зонам и составление схемы определения данного грамматического явления;

— составление тестов к проверке составленной схемы;

— испытание схемы на тестах и внесение корректив по результатам проверки.

3.4. Лексика. Словарь

Лексика есть составляющая предметного элемента предметного компонента модели. Словарь есть составляющая алгоритмического элемента динамического компонента модели. Основой систематизации лексики и создания машинного словаря в обоих случаях является специальная классификация слов.

Рассмотрим строение входного словаря.

В исходной классификации слов различаются семь основных лексико-грамматических классов и один класс, существующий как словарный только до этапа разрешения омографии включительно — класс омографов. Словарные классы следующие: существительные, прилагательные, предлоги, глаголы, наречия, союзы, знаки раздела (знаки препинания), омографы. Каждой словоформе ставится в соответствие один номер в словаре — цифровой эквивалент. Деление словаря на лексико-грамматические классы и подклассы отражается в распределении цифровых эквивалентов (ЦЭ).

Внутри класса существительных выделяются следующие семантические подклассы, определенные по дистрибутивно-семан-

тическому признаку индуктивным путем, т. е. в зависимости от того, насколько выделение того или иного подкласса определяется потребностями текстового различения или разрешения лингвистического феномена, к какому бы уровню он ни относился.

Существительные: союзные, местоименные, числительные, временные, географические названия, названия организаций, периодические издания, имена собственные, одушевленные, названия документов, названия денежных единиц, прочие.

Дистрибутивно-семантический признак лежит в основе часто применяющегося в МП решения о разделении числительных, местоимений и союзных слов между существительными и прилагательными соответственно их синтаксической роли и грамматическому оформлению.

Особый разряд отведен «новому» существительному. Имена прилагательные делятся на: союзные, местоименные, национальные, географические, прочие относительные, качественные, числительные. К прилагательным относятся также: определенный и неопределенный артикли, обозначения кавычек, тире и отрицание *no*. Особый раздел отведен «новому» прилагательному.

Среди предлогов какой-либо классификации по разрядам (подклассам) нет.

Глаголы разделяются на подклассы следующим образом:

1 форма: модальные, вспомогательные, прочие;

2 форма: модальные, вспомогательные, прочие;

3 форма: вспомогательные, прочие.

Специальный разряд отведен для «нового» глагола.

1 форма — основа настоящего и будущего времени, 2 форма — прошедшее время, 3 форма — причастие прошедшего времени. Синтаксическая роль глагола в предложении определяется на этапе анализа и выражается в его синтаксической функции, семантические характеристики извлекаются из входного словаря.

К лексико-грамматическому классу наречий относятся: частица *to*, отрицание *not*, «новое» наречие.

Внутри союзов деления на подклассы нет.

Знаки препинания разделяются на знаки раздела фраз (точка, точка с запятой, знак абзаца и пр.) и знаки выделительные (запятая, скобка).

Класс омографов разделен на подклассы в соответствии с видами омографии. В качестве исходных для системы приняты следующие подклассы (виды) омографии:

«глагол—существительное», «глагол—прилагательное», «союз—существительное», «союз—предлог», «союз—наречие», «существительное—предлог», «существительное—прилагательное», «существительное—наречие», «предлог—прилагательное», «предлог—наречие», «прилагательное—наречие». Это так называемые двухкомпонентные подклассы омографии, в которых словоформы

могут принадлежать к одному из двух лексико-грамматических классов. Отметим, что среди этих подклассов есть два принципиально открытых: глагол—существительное и существительное—прилагательное. Эти подклассы могут значительно пополняться в текстах новой тематики. Исходный состав этих подклассов определен на начальном корпусе текстов, равно как и основы процедуры разрешения данного вида омографии для словоформ текста.

Далее следуют трехэлементные подклассы: «глагол—союз—предлог», «глагол—союз—прилагательное», «глагол—существительное—прилагательное», «глагол—прилагательное—наречие», «союз—существительное—прилагательное», «союз—предлог—наречие», «существительное—предлог—прилагательное», «существительное—прилагательное—наречие», «предлог—прилагательное—наречие», «существительное—предлог—наречие», «союз—существительное—наречие». Эти подклассы практически закрыты, часто они состоят из одного или нескольких слов. За ними следуют четырехэлементные подклассы: «глагол—существительное—предлог—прилагательное», «глагол—существительное—предлог—наречие», «глагол—существительное—прилагательное—наречие», «глагол—предлог—прилагательное—наречие», «союз—существительное—прилагательное—наречие», «союз—предлог—прилагательное—наречие», «существительное—предлог—прилагательное—наречие», и наконец, один пятиэлементный подкласс: «глагол—существительное—предлог—прилагательное—наречие».

Специальный разряд в рамках ЦЭ омографов отведен для нового слова. Особый подкласс занимает омография цифр и омография букв (например, буква *a* с неопределенным артиклем).

Таким образом, одно из основных исходных делений словаря есть деление слов на лексико-грамматические классы и подклассы слов. Вторым делением является деление на дистрибутивно-семантические и морфологические подклассы внутри классов.

Третье деление ЦЭ отражает распределение слов словаря на однозначные и многозначные слова.

В исходном корпусе текстов можно выделить слова, имеющие один или несколько переводов в русском языке. Каждый перевод слова может быть принят за значение его в данном подязыке, и поэтому можно разделить слова словаря на однозначные и многозначные. Это деление отражается в словаре распределением ЦЭ, который, таким образом, становится носителем информации о многозначности/однозначности слова.

Деление на однозначные и многозначные слова проходит по всему словарю, поскольку внутри каждой части речи (кроме знаков препинания и омографов) могут быть выделены соответственно однозначные и многозначные слова.

Распределение цифровых эквивалентов во входном словаре дано в табл. 3.

Морфологический синтез выходного текста осуществляется на основании данных анализа и словарной информации, записанной в выходном (русском) словаре. Рассмотрим его состав.

Исходными данными для морфологического синтеза являются:

- словарь русского языка;
- промежуточный текст (ПТ) в виде последовательности информационных ячеек с записью данных анализа;
- таблицы классов русских слов;
- цепочки окончаний;
- программы синтеза.

Состав и объем русского словаря определяется набором тех переводных эквивалентов, которые придаются однозначным и многозначным словам английского словаря и английским фразеологическим словосочетаниям на этапе перевода. В словаре выделены следующие лексико-грамматические классы:

- существительные;
- глаголы;
- прилагательные;
- предлоги;
- союзы;
- наречия и вводные слова.

Местонахождения и числительные разделены указанным выше образом между существительными и прилагательными. Причастия и деепричастия в словарь самостоятельно не помещены, а образуются от основ глагола по специальным правилам.

Существительные располагаются в словаре по родам, глаголы — по классам управления. Предлоги также отнесены к классам управления, подобно глаголам. Кроме отдельных слов в словаре имеются также словосочетания, состоящие из двух и более слов. Словосочетания разделяются на изменяемые и неизменяемые. Сочетание в словаре помещается в тот лексико-грамматический класс, в который входит основное его слово.

Машинной основой слова считается самая большая неизменяемая часть слова, общая для всех его грамматических норм.

Основы выделяются внутри одной части речи, а у глагола (см. далее) 4-го видового класса — внутри одного вида. Выделение основ произведено таким образом, что при синтезе словоформ посредством суффиксации, префиксации и флективным путем все только присоединяется и ничего не отнимается. Иногда основа совпадает с корнем слова (*лампа-а, бегущий*), в некоторых случаях она представляет собой законченную грамматическую форму (*палец*). Исторически сложившиеся чередования букв, появление беглых гласных и супплетивность при русском формообразовании находят свое выражение в выделении дополнительных основ в словаре, когда это необходимо. Например, существительное *доска* имеет две основы: *доск-а, досок*, глагол

Таблица 3. Распределение цифровых эквивалентов

Лексико-грамматические классы и подклассы слов	Однозначные		Многозначные	
	с № ...	с № ...	с № ...	с № ...
Существительные	0 0 5	11 30 31	20 0 0	23 31 31
союзные	0 0 5	0 0 10	20 0 0	20 0 9
местоименные	0 0 11	0 1 28	20 0 10	20 1 17
числительные	0 1 29	0 3 4		
временные	0 3 5	0 5 10	20 1 18	20 2 25
географические названия				
названия организаций	0 11 19	2 8 15	20 5 20	20 11 27
периодические издания	2 8 17	5 10 10	20 11 28	20 28 23
имена собственные	5 10 11	6 1 24	20 28 24	21 4 17
одушевленные	6 1 25	6 16 20	21 4 18	21 9 7
название документов	7 16 21	8 8 2	21 9 8	21 27 21
названия денежных единиц	8 11 7	8 15 28	21 29 18	21 31 3
прочие	8 15 29	8 16 30	21 31 3	21 31 31
новое существительное	8 16 29	11 19 8	21 31 4	23 28 27
11 30 31				
Прилагательные	12 0 0	15 31 31	24 0 0	26 31 31
союзные	12 0 3	12 0 11	24 0 0	24 0 9
местоименные	12 0 12	12 0 31	24 0 10	24 0 31
национальные	12 1 0	12 5 25	24 1 0	24 3 24
географические	12 5 26	12 6 29	24 3 25	24 6 27
прочие относительные	12 6 30	13 31 19	24 6 28	25 0 0
качественные	13 31 22	15 30 23	25 0 1	25 31 15
числительные	15 30 24	15 31 30	25 31 16	25 31 30
неопределенный артикль	12 0 1			
определенный артикль <i>the</i>	12 0 2			
отрицание <i>no</i>	25 31 31			
новое прилагательное	15 31 31			
Предлоги	16 0 0	16 1 31	26 0 0	26 1 31
Глаголы	16 2 0	17 31 31	26 2 0	27 31 31
1 форма модальные	16 2 0	16 2 9		
вспомогательные	16 2 10	16 2 19	26 2 0	26 31 31
2 форма модальные	17 24 16	17 24 25		
вспомогательные	17 24 26	17 25 3	27 24 25	27 29
3 форма вспомогательные			27 30 0	27 30 31
прочие	17 28 8	17 31 30	27 30 4	27 30 31
Новый глагол	17 31 31			
Наречия	18 0 0	19 8 31	28 0 0	28 11 31
частица <i>to</i>	18 0 0			
отрицание <i>not</i>	18 0 1			
наречия	18 0 2	19 8 30	28 0 0	28 11 31
новое наречие	19 8 31			
Союзы	19 9 0	19 31 31	28 12 0	28 10 31
Знаки раздела фраз	19 24 16	19 24 31		
Знаки выделительные	19 24 0	19 24 15		
Омографы	28 16 0	31 31 30		

расти имеет основы *раст-и, рос-ла*. Слово помещается в словаре со всеми своими основами: сначала первая, затем следующие. У глаголов четвертого видового класса основы выделяются внутри одного вида. Нумерация основ определяется степенью их употребляемости. Первой основой считается самая употребительная, затем идут менее употребляемые. Данные о частоте употребления основ могут быть получены из частотного словаря.

Для синтеза словоформ произведена классификация всех слов русского словаря по способу образования грамматических категорий — словообразовательная классификация. Для глагола, помимо словообразовательной, сделана также классификация по видам и классификация по управлению.

Словообразовательные классы выделены внутри каждой части речи и объединяют слова, одинаково образующие свои словоформы. Класс указывает на количество основ у слова, а также на то, какая основа и какое окончание участвуют в образовании той или иной грамматической формы. Словообразовательные классы помещены в специальные таблицы, где по горизонтали значатся номера классов, а по вертикали — наименования грамматических форм. Каждое поле таблиц содержит сведения о номере основы и типе окончания для определения грамматической формы определенного класса. Каждое изменяемое слово имеет один словообразовательный класс, за исключением глаголов четвертого видового класса, которые имеют два таких класса — один для форм совершенного, другой — для форм несовершенного вида.

Видовая классификация отражает особенности в образовании видов глагола. Все русские глаголы делятся на одновидовые и двухвидовые. Последние делятся на глаголы, совмещающие в одном слове два вида, и глаголы, у которых разные виды выражены разными словами. Последние образуют противоположный вид либо при помощи суффикса (*кричать—крикнуть*), либо при помощи приставок (*писать—написать*), либо посредством другого слова (*говорить—сказать*).

Глаголы, образующие противоположный вид при помощи суффиксов и посредством другого слова, имеют в словаре две видовые основы. Глаголы, образующие противоположный вид при помощи приставок, имеют в словаре одну видовую основу, а нужная приставка указывается номером видового класса и выбирается из таблиц видовых классов. За основную видовую форму принята форма несовершенного вида.

Видовые классы построены таким образом, что все формы совершенного вида образуются от форм несовершенного вида. Видовых классов 24.

Классы управления глагола содержат следующую информацию:

— класс управления указывает, каким падежом управляет данный глагол или глагольное сочетание в беспредложных конструкциях;

— класс управления является показателем переходности.

Слова и словосочетания в русском словаре записываются в виде словарных статей. Словарная статья представляет собой запись буквенной основы слова для изменяемых слов (или слов в полной форме для неизменяемых слов) в соединении с вспомогательной информацией. Словарные единицы в словаре записываются по ячейкам. Ячейка представляет собой поле, состоящее из 9 байтов. Каждая новая словарная статья записывается с новой ячейки. Порядковый номер первого байта первой ячейки, с которого начинается запись словарной статьи, называется номером перевода (НП). НП является относительным адресом словарной статьи в памяти ЭВМ. В схемах анализа и перевода для удобства используется не само слово, а его номер перевода.

НП является носителем грамматической информации, поскольку указывает диапазон словарных групп и наряду с ЦЭ используется при грамматическом анализе.

Весь русский словарь разбит на грамматические классы. В один класс объединяются лексические единицы с одинаковыми грамматическими характеристиками. Каждый класс занимает в памяти ЭВМ около 32 тыс. байтов. Если словарные статьи занимают более 32 тыс. байтов, то для данного типа лексических единиц с теми же грамматическими характеристиками выделяются дополнительно один или несколько классов. Неизменяемые слова помещаются в словарь в законченном виде, изменяемые — по основам. Рядом со словом записываются грамматические и служебные сведения о нем, — так называемая словарная информация, используемая при синтезе словоформ. Служебные сведения представляют собой данные о количестве байтов, занятых словом, о количестве букв в слове (у неизменяемых слов) или в основах слова (у изменяемых). Грамматические сведения — номер словообразовательного класса слова и номера падежных и видовых классов глаголов. Глаголы четвертого видового класса имеют в словаре основы несовершенного и совершенного вида и номера двух словообразовательных классов, одного для форм несовершенного вида, другого — для форм совершенного. Местоимения — существительные представлены в словаре в следующем виде: в словарь заносится номер класса местоимения и неизменяемая часть его (т. е. частица *то, либо, нибудь, кое* и др.), если она есть. Само местоимение и формы его косвенных падежей содержатся в таблице классов местоимений.

Существует несколько типов словарных статей:

- словарная статья неизменяемого слова или словосочетания;
- словарная статья существительного или прилагательного;
- словарная статья глагола;
- словарная статья изменяемого словосочетания;
- словарная статья местоимения.

Типы статей различаются между собой по структуре.

3.5. Грамматика

Грамматика как составляющая предметного элемента предметного компонента включает в себя морфологию, синтаксис, семантику и выступает как система формальных признаков, необходимая и достаточная для осуществления поиска и обнаружения переводных соответствий с целью достижения эквивалентного перевода.

Грамматика содержит:

— исходную классификацию слов по лексико-грамматическим классам с дистрибутивно-семантическими подклассами, принадлежность к которым определяется семантико-синтаксическим кодом слова;

— морфологические признаки слов (окончания, приставки);

— набор грамматических характеристик для каждого лексико-грамматического класса, с помощью которых описывается словоизменение и согласование данного слова с другими словами в предложении (число, лицо, время, вид, наклонение для глагола, число для существительного и прилагательного и пр.);

— набор исходных семантико-синтаксических категорий — СФ, в терминах которых производится синтаксический анализ входного предложения.

После разрешения омографии словам-омографам приписываются признаки соответствующего лексико-грамматического класса. Это разрешение происходит до начала основного грамматического анализа, поэтому каких-либо специальных грамматических особенностей слов-омографов грамматика не содержит.

Общие требования к грамматике могут быть сформулированы следующим образом:

— грамматика должна обладать свойством системности, которое понимается здесь как обеспечение экономного описания и передачи всех релевантных переводу грамматических черт исходного языка.

Каждая отдельно взятая переводческая задача допускает в большинстве случаев решение в виде последовательности достаточно формализованных шагов. Однако собрание таких частных рецептов разрешения трудностей перевода не образует само по себе ни алгоритмов перевода, ни трансляционной грамматики. Комплексный характер и системность грамматики как составляющей предметного элемента предметного компонента должны обеспечивать возможность и адекватность переводческой задаче трансляционной грамматики как составляющей алгоритмического элемента динамического компонента;

— основные категории грамматики должны быть немногочисленны, удобны для человека-пользователя и соответственно этому близки к традиционным грамматическим описаниям.

Вопрос построения оптимального с указанных точек зрения грамматического описания весьма сложен. В рамках дедуктивного подхода к проблеме МП разработаны достаточно тонкие

Таблица 4

Признак	Код признака	Признак	Код признака
Инфинитив	0	Мужской род	1
Настоящее время глагола	1	Женский »	2
Прошедшее время глагола	2	Средний »	3
Будущее время глагола	3	1 лицо глагола	1
Настоящее время глагола	0	2 лицо »	2
Прошедшее время причастия	1	3 лицо »	3
Без окончания	0	Пассив	0
-s, -th	1	СФ подлежащего	1
-er, -ed	2	СФ сказуемого	2
-ing, -est	3	СФ определения	3
Именительный падеж	1	СФ обстоятельства	4
Родительный »	2	СФ предложения	5
Дательный »	3	СФ дополнения	6
Винительный »	4	СФ разделительная	7
Творительный »	5		
Предложный »	6		
Единственное число	0		
Множественное число	1	Однородный знак преп.	0
Положительная степень прилагательного	0	Неоднородный знак преп.	1
Сравнительная степень прилагательного	2	Изъявительное наклонение	0
Превосходная степень прилагательного	3	Сослагательное наклонение	1
Несовершенный вид глагола	0	Повелительное наклонение	2
Совершенный вид глагола	1		

много обеспечения и для уточнения некоторых представлений и действий в рамках лингвистического обеспечения.

Особое место в системе занимают так называемые правила пропуска. Первоначально идея пропуска несущественных для анализа слов занимала скромное место в алгоритме лингвистического анализа.

В настоящем алгоритме и в системе в целом эти правила играют существенную роль. В отсутствие эксплицитных признаков границ синтагм определение расстояния между существенными для анализа и перевода лексическими единицами выполняет функции установления именно таких границ. Задавая линейное расстояние от исходного лексического элемента и определяя с помощью правила пропуска качество и состав информации к лексическим единицам, объявляемым несущественными для поиска детерминанты или элемента диагностирующей конструкции, мы тем самым делим анализируемую последователь-

приемы грамматического описания, однако практически все из них ориентированы отдельно либо на анализ, либо на синтез. Нет аналогов трансляционной грамматики в разработках и исследованиях по информатике, поскольку грамматика информационно-поисковых языков вследствие своей примитивности чрезвычайно далека от грамматики естественных языков, ориентированной на перевод.

Рассмотрим комплексы признаков, характеризующих части речи (лексико-грамматические классы).

Комплексы признаков, характеризующих несовместимые части речи, являются несовместимыми между собой.

Информационная ячейка (ИЯ) рассматриваемой системы имеет 45 двоичных разрядов, или девять пентад (групп по пять разрядов). Каждой словоформе сопоставляется одна ИЯ, в которой, таким образом, должны быть записаны как словарные, так и несловарные признаки. Структура ИЯ следующая: три старшие пентады — ЦЭ слова; четыре младшие — номер перевода (заполняется на этапах перевода однозначных и многозначных слов), промежуточные пентады — грамматическая информация. Состав грамматических признаков для классов слов:

омограф — грамматической информации нет; **существительное** — одушевленное; нет определения; определение; окончание *s*; число; род; падеж; СФ; **прилагательное** — одушевленное краткое; окончание; степень сравнения; число; род; падеж; СФ; **глагол сказуемое** — наклонение; окончание; род; число; лицо; пассив; вид; время; инфинитив; СФ; **глагол-причастие** (определяется по СФ «определение» для глагола) — одушевленное; вид; время; пассив; число; род; падеж; СФ; **глагол-деепричастие** (определяется по СФ «обстоятельство» для глагола) — СФ; **союз** — неоднородный (разделяющий простые предложения во фразе); СФ; **знак препинания** — то же самое; **предлог, артикль, цифра** — СФ; **новое слово** — СФ.

В табл. 4 приведены коды признаков.

Программным признаком, необходимым для настоящего рассмотрения для того, чтобы была ясна структура ИЯ и методы применения стандартных операторов, является таблица так называемых сдвигов, т. е. расстояний в разрядах от 5 разряда ИЯ до начала записи в ИЯ данного признака. Значения сдвигов приведены в табл. 5.

После того, как описаны все составляющие информационной ячейки для всех частей речи, можно представить структуру ИЯ. На таблице 6 изображена структура ИЯ для частей речи с распределением формальных признаков по ячейке.

После завершения этапов анализа и перевода последовательность символов, введенных в ЭВМ, представляет собой набор заполненных информационных ячеек. Этот набор представляет собой так называемый промежуточный текст (ПТ). Понятие «промежуточный текст» используется для описания програм-

Таблица 5

Признак	Величина сдвига в разрядах		Признак	Величина сдвига в разрядах	
	Не глагол	Глагол		Не глагол	Глагол
Число	20	20	Вид	23	17
Окончание	21	21	Наклонение	—	23
Род	18	21	СФ	12	12
Падеж	15	15	Однородность	15	—
Пассив	22	18	Одушевленность	24	—
Время	21	15	Определение	22	—

* Знак — означает, что данная категория нерелевантна.

Таблица 6

	Разряды			
	45—31	30—21	20—1	
Омограф	Цифровой эквивалент	Грамматическая информация	Синтаксическая функция	Буквы окончания
Существительное				
Прилагательное				
Глагол				
Союз, знак препинания				
Предлог, артикль, цифра				
Новое слово				

ность слов (предложение, или фразу в целом, или его (ее) части) на ряд более мелких синтаксических единиц, в рамках которых происходит установление собственно переводных соответствий. Именно вследствие этого, т. е. своей функциональной эквивалентности установлению границ синтагм, правила пропуска могут быть включены в грамматику.

В табл. 7 приведены правила пропуска. Детализация этих правил сделана в (Марчук, Тихомиров, Щербинин, 1975).

Грамматика выходного языка определяется требованиями синтеза.

На этапе синтеза по данным заполнения ИЯ промежуточного текста воссоздается русская словоформа.

Рассмотрим основные принципы синтеза.

Морфологический синтез предусматривает образование следующих грамматических форм русских слов.

Существительное:

Все падежные формы единственного и множественного числа.

Прилагательное:

Таблица 7. Правила пропуска слов при проверках контекста

№ пп	Содержание групп правил пропуска	Пример (номер конкретного правила и само правило)	Число правил в группе
1	Пропускать все слова, принадлежащие к одному или нескольким лексико-грамматическим классам	1) существительные и прилагательные 9) наречия	15
2	Пропускать слова определенных лексико-грамматических классов с признаками	21) союзы и знаки раздела с признаком «однородный»	3
3	Пропускать слова с определенной синтаксической функцией	28) Слова с СФ «обстоятельство» и «определенные»	4
4	Пропускать все слова в заданных границах	18) все до знака раздела между фразами	1
5	Ничего не пропускать		
Всего:			32

1) положительная степень, полная форма; 2) положительная степень, краткая форма; 3) сравнительная степень; 4) превосходная степень.

Глагол:

1) изъявительное наклонение, актив, инфинитив; 2) изъявительное наклонение, актив, формы настоящего времени; 3) изъявительное наклонение, актив, формы будущего простого времени; 4) изъявительное наклонение, актив, формы будущего сложного времени; 5) изъявительное наклонение, актив, формы прошедшего времени, несовершенный вид; 6) изъявительное наклонение, актив, формы прошедшего времени, совершенный вид; 7) изъявительное наклонение, пассив, инфинитив, несовершенный вид; 8) изъявительное наклонение, пассив, инфинитив, совершенный вид; 9) изъявительное наклонение, пассив, формы настоящего времени несовершенного вида; 10) изъявительное наклонение, пассив, формы настоящего времени совершенного вида; 11) изъявительное наклонение, пассив, формы будущего времени, несовершенного вида; 12) изъявительное наклонение, пассив, формы будущего времени совершенного вида; 13) изъявительное наклонение, пассив, формы прошедшего времени несовершенного вида; 14) изъявительное наклонение, пассив, формы прошедшего времени совершенного вида; 15) повелительное наклонение, несовершенный вид; 16) повелительное наклонение, совершенный вид; 17) сослагательное наклонение, актив, инфинитив, несовершенный вид; 18) сослагательное наклонение, актив, инфинитив, совершенный вид; 19) сослагательное наклонение, актив, формы прошедшего времени, несовершенный вид; 20) сослагательное наклонение, актив, формы прошедшего времени, совершенный вид; 21) причастие действительное настоя-

шего времени; 22) причастие действительное прошедшего времени; 23) причастие страдательное настоящего времени полное; 24) причастие страдательное прошедшего времени полное; 25) деепричастие несовершенного вида; 26) деепричастие совершенного вида.

Грамматические формы слов русского языка можно разделить на простые и сложные (состоящие из двух и более отдельных слов). К последним относятся:

1. Сложная сравнительная степень прилагательного.
2. Сложная превосходная степень прилагательного.
3. Изъявительное наклонение, актив, будущее сложное время глагола.
4. Изъявительное наклонение, пассив, несовершенный вид, будущее время глагола.
5. Изъявительное наклонение, пассив, совершенный вид, прошедшее время глагола.
6. Изъявительное наклонение, пассив, совершенный вид, будущее время глагола.
7. Изъявительное наклонение, пассив, инфинитив.
8. Сослагательное наклонение, актив, прошедшее время глагола.
9. Сослагательное наклонение, актив, инфинитив.
10. Сослагательное наклонение, пассив, инфинитив, несовершенный вид глагола.
11. Сослагательное наклонение, пассив, инфинитив, совершенный вид глагола.
12. Сослагательное наклонение, пассив, прошедшее время, несовершенный вид глагола.
13. Сослагательное наклонение, пассив, прошедшее время, совершенный вид глагола.

3.6. Семантика

Основа семантической составляющей предметного элемента предметного компонента модели задается системой лексико-грамматических классов и синтаксических функций слов в предложении. Семантико-синтаксические (или синтактико-семантические) характеристики этой исходной системы определяются близостью к традиционно сложившейся универсальной системе семантических категорий текста — системе «части речи — члены предложения». Дальнейшее развитие этой системы происходит в семантических подклассах, выделяемых дистрибутивным путем: в составе существительных и прилагательных. Наконец, релевантные переводу синтаксические и сочетаемостные свойства лексических единиц находят реализацию в виде списков слов, которые по объему меньше, чем словарные группы, и поэтому не находят отражения в словаре. Эти списки слов используются непосредственно в схемах грамматического анализа и допускают обобщение к группам (подклассам) слов входного словаря.

Выходной словарь, как было указано выше, имеет более сложную структуру, в рамках которой некоторые существенные для синтеза и анализа после перевода синтактико-семантические свойства (например, глагольное управление) входят в словарную информацию и отражены в номере класса слова.

Модель МПС позволяет иметь семантическую составляющую практически открытой для пополнения семантическими признаками, которые могут вводиться не только как словарная информация, но и в виде отдельных кодов, не обязательно фиксированных в словаре, а способных возникать и действовать только в пределах алгоритма. К числу таких кодов можно отнести всякого рода пометы о тематике текста или временные признаки, приписываемые анализируемым единицам для правильного синтеза в выходном тексте.

Пополнение семантической составляющей осуществляется двумя основными путями: включением новых семантических параметров (различительных признаков) и обобщением от отдельных слов к спискам и от списков к семантическим группам, подклассам и классам слов. Рассмотрим эти два пути.

Возьмем следующий текст из двух предложений: *Впервые Советский Союз участвовал в здеишей ярмарке ровно пятьдесят лет назад — в 1928 году. Сохранились фотографии и другие документы, показывающие то, что мы могли тогда предложить — отдельные виды сельскохозяйственной продукции, кустарные изделия, редкие промышленные товары* («Правда»). Слово *редкий* имеет по крайней мере два значения: 1) 'особый, выдающийся, уникальный по свойствам' (*редкая красота*); 2) 'мало частотный (в противопоставление частому), нераспространенный, не массовый'. В данном тексте слово *редкий* использовано во втором значении. Отграничить эти два значения, пользуясь только контекстом, не представляется возможным. Вводим дифференциальный семантический признак «массовость/немассовость (уникальность)». С помощью этого признака можно было бы описать различие между двумя значениями, если бы эксплицитно они требовали двух разных переводных эквивалентов. Однако в условиях системной организации семантический признак, прежде чем он может быть включен в список используемых, должен пройти проверку на неоднозначность использования, т. е. следует определить, в каком круге неоднозначных явлений можно будет его использовать, при этом следует убедиться, что такое использование не будет однократным. Введение новых семантических подклассов и классов будет обосновано в том случае, если оно требуется из соображений удобства описания дистрибуции.

Не требует особых доказательств положение о том, что распространение от отдельных слов, описывающих индивидуальную лексическую сочетаемость, к спискам и подклассам слов не всегда возможно даже для свободных словосочетаний. Пример: многозначное слово *interception* в сочетании с существительным

telephone переводится как 'подслушивание', в остальных случаях — 'перехват'. Создание некоторого списка вместо слова *telephone* представляется маловероятным. Рассмотрим существительное *interest*. Контекстологический словарь (Марчук, 1976) дает на это слово четкую словарную статью, в которой приводятся обусловленные контекстом три разных перевода этого слова. Один из них 'доход', детерминируется, во-первых, сочетанием с прилагательным *annual*, во-вторых, сочетанием с названием организации. Можно себе представить, что слово *annual* может на самом деле быть составляющей некоторого списка слов со сходным значением, и, кроме того, что вместо 'названия организации' или в дополнение к этому подклассу могут быть введены и другие семантические подклассы. Однако изучение словарной статьи, например Большого англо-русского словаря (БАРС 1972), не дает ответа и не иллюстрирует возможностей расширения, поскольку в нем делается упор на разъяснение семантических значений и оттенков слова посредством приведения как можно большего количества различных примеров. В некотором смысле целевые установки контекстологического словаря и БАРСа вообще различны: БАРС демонстрирует тонкие значения и их оттенки, широко опираясь на индивидуальные свойства слов, контекстологический словарь выносит на передний план наиболее массовые и индивидуальные свойства слова.

Таким образом, распространение и само создание составляющей «семантика» обусловлено релевантностью вводимых семантических признаков для перевода, различения значения и разрешения многозначности.

3.7. Переводной элемент предметного компонента

Переводной элемент предметного компонента образуется из типов соответствий.

Типы соответствий естественно рассматривать с точки зрения уровней языковой структуры, релевантных для перевода. Для моделирования перевода можно взять в качестве первого уровня уровень слов, поскольку уровни фонем и графем, хотя и могут быть самостоятельно использованы в МП, какой-либо особой проблематики не порождают⁴. Далее следуют уровни словосочетаний, предложений, текста, каждый из которых так или иначе связан с типами соответствий.

Рассмотрим эквивалентные соответствия.

Реализация эквивалентных соответствий на уровне слов может быть достаточно полно представлена однозначными словами. Так, английский глагол *inform* в исходном конкордансе всегда переводится как 'информировать'. Исходное слово может переводиться словосочетанием и обратно, словосочетание может

переводиться отдельным словом. Соотношение двух первых уровней при эквивалентных соответствиях может быть представлено так:

- слово — слово;
- слово — словосочетание;
- словосочетание — слово;
- словосочетание — словосочетание.

Примеры: *inform* 'информировать'; *still* 'все еще'; *get up* 'подняться'; *find favour* 'находить поддержку'.

Осуществление перевода предложения с помощью только эквивалентных соответствий представляется маловероятным. Если соответствия первых двух уровней и могут оказаться только эквивалентными, то соответствия в грамматике и семантике трудно себе таковыми представить для всего предложения в целом, если только оно не представляет собой исключительно простую конструкцию. Тем более это относится к эквивалентным соответствиям на уровне текста.

Вариантные соответствия на уровне слов и словосочетаний описываются теми же соотношениями, что и для эквивалентных соответствий, однако для первых возникает проблема выбора по тексту одного нужного перевода из нескольких возможных. Примером вариантных соответствий могут служить алгоритмы перевода многозначных слов. Соответствия на уровне предложения почти всегда варианты (если не относятся к трансформационным) вследствие высокой грамматической омонимичности аналитических форм английского языка. Рассмотрим пример из текста, переведенного машиной⁵.

A very important step has been taken in the direction of efficient and systematic organization of information transfer 'Самый важный шаг был взят в направлении эффективную и систематическую организацию информации перевозит'. Ошибки в согласовании прилагательных «эффективный» и «систематический» произошли вследствие неправильного определения падежа существительного после предлога *of*, затем этот падеж от управляющего существительного схемами анализа был передан прилагательным. Получила неправильное разрешение лексико-грамматическая омонимия слова *transfer*, вследствие чего слово *information* не было переставлено, что произошло бы при правильной квалификации *transfer*. Пример демонстрирует множественность выбора грамматических форм и омонимичность грамматической информации, вследствие чего соответствия должны быть отнесены к вариантным.

Трансформационные соответствия на уровнях слов и словосочетаний фиксируются в схемах перевода многозначных слов и сложных словосочетаний (словаря оборотов). Здесь возможны и реализованы достаточно сложные преобразования. Так, в ал-

⁴ Кроме графематического анализа небуквенных включений и морфологической омонимии, а также словесных комплексов, связанных знаком «←».

⁵ Здесь и далее примеры взяты либо из переведенных на ЭВМ текстов, либо из контекстологического словаря (Марчук 1976).

горитме перевода многозначного глагола *fail* учтены следующие изменения структуры *failing this (that)* — 'если это не будет сделано'. При этом широко используются приемы приписывания некоторым словам словосочетания нулевых переводов, вставки новых слов и т. п. Так, в обороте *fall in with* предлогу *in* приписывается нулевой перевод, а предлог *with* остается для последующего перевода предложными схемами. Трансформационные соответствия на уровне слов и словосочетаний могут осуществляться с использованием грамматической информации, например, «пассив» для глаголов, «множественное число» для существительных и т. п. Так, глагол *involve* в пассиве переводится 'быть втянутым' (с последующей детерминантой *struggle*). Такого рода детерминирование производится лишь в тех случаях, когда при этом меняется лексическое значение (общий перевод глагола *involve* 'включить').

На уровне предложений, с использованием грамматической и семантической информации трансформационные преобразования осуществляются алгоритмами:

— анализа и перевода предложений с отрицаниями и отрицательными наречиями (*no, not, never, nowhere, nothing, nobody* и т. п.);

— перемены порядка слов для цепочки существительных, из которых первое является определением к последующим (*information transfer* 'передача информации');

— анализа и перевода глагольных конструкций со вторичной предикацией (*I saw him run*);

— анализа и перевода конструкций, использующих некоторые глаголы (*I like him* 'он мне нравится'; *the speaker was reported to have said* 'как сообщили, оратор заявил, что...' и т. п.);

— анализа и перевода некоторых специальных конструкций (*there is (are, was, were)*).

Семантика глаголов, при которых могут быть локализованы те или иные трансформации структуры предложения, может быть задана в явном описательном виде (глаголы слушания, говорения, физического восприятия и пр. для конструкций со вторичной предикацией); однако более экономным и четким является задание этих глаголов списком.

Описанный арсенал средств осуществления трансформационных соответствий, как можно себе представить, не охватывает все возможные случаи требуемых преобразований, для достижения полной коммуникативной эквивалентности. Однако если учесть, что существует отмечаемая теоретиками перевода близость синтаксических структур основных европейских языков и что данные трансформации охватывают наиболее частые языковые явления, они вполне могут быть положены в основу трансформационной составляющей переводного элемента с достаточной долей вероятности того, что основные необходимые трансформации окажутся предусмотренными. Случаи комплексных трансформаций могут быть переданы человеку-редактору.

Трансформационные соответствия на уровне текста, как представляется, во многих отношениях вывели бы нас на задачу алгоритмизации и эксплицитного представления смысла текста. В модели МПС представляется целесообразным рассматривать трансформационные соответствия на уровне текста как сумму трансформационных (и прочих) соответствий на уровнях предложений, составляющих текст.

3.8. Динамический компонент модели

Важнейшим элементом динамического компонента модели является алгоритмический элемент, который содержит в качестве составляющих алгоритм, трансляционную грамматику и словарь.

Рассмотрим принципы построения общего алгоритма, управляющего частными в анализе и синтезе текста.

В приведенных ранее примерах перевода английского и немецкого предложений анализ основан на последовательном движении по тексту слева направо при целом ряде рекурсий и переборе возможностей, связанных с грамматическими характеристиками встречающихся словоформ и их потенциальной сочетаемостью с элементами предложения. Машинная модель такого способа анализа — предсказуемый анализатор (Куно—Эттингер 1971) связана с большим перебором, громоздкими таблицами даже для ограниченных фрагментов языка и поэтому распространения не получила.

Движение по тексту в какую-либо одну сторону не является единственным способом анализа. Для выявления синтаксической структуры предложения может быть использована стратегия опорных точек (Гарвин, 1971). Практический неуспех метода, на наш взгляд, объясняется тем, что в качестве целевой установки при этом было выбрано получение полной синтаксической структуры предложения.

Если не ставить задачи нахождения полной синтаксической структуры предложения и ограничиться определением СФ, можно применить экономный метод, который мы называем циклическим. Цикличность метода заключается в том, что на тексте, введенном в память ЭВМ, последовательно работают программы основных этапов алгоритма, каждый из которых просматривает текст слева направо. Внутри же каждого этапа и отдельного частного алгоритма (схемы) могут быть реализованы проверки введенного текста (который не ограничивается предложением или фразой) в любую сторону от исходного элемента.

Блок-схема общего алгоритма показана на рис. 9.

Идентификация десигната охватывает довольно широкий круг операций: поиск входной цепочки символов в словаре, приписывание ей словарной информации; если слово не обнаружено, производится его специальный анализ как нового слова; распознавание оборотов (фразеологических единиц), которые в

дальнейшем анализе выступают как одна лексическая единица с фиксированной грамматической информацией; морфологический анализ словоформ.

Монограмматический анализ назван так потому, что в нем используются данные только входного текста. Здесь производятся следующие действия: разрешение лексико-грамматической омонимии; определение СФ слов предложения; определение ряда грамматических категорий, для которых достаточно данных выходного текста — число существительного, время глагола и т. п.

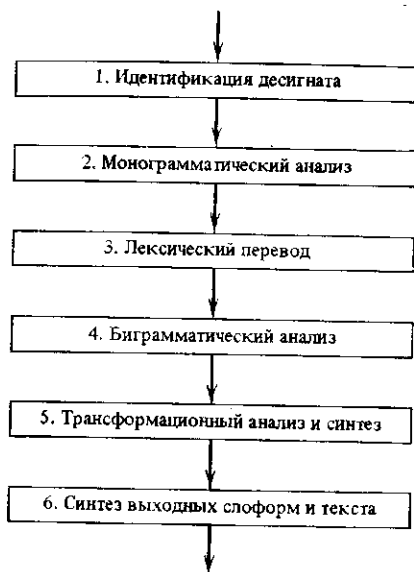


Рис. 9. Блок-схема общего алгоритма модели МПС

Блок лексического перевода описывает: перевод однозначных слов; перевод многозначных слов.

Биграмматический анализ использует грамматические данные как входного, так и выходного языка. Здесь производятся: грамматическое и морфологическое оформление управления и согласования, определение грамматических категорий с использованием данных выходного текста в виде номеров переводов, полученных на этапе лексического перевода.

Трансформационный анализ и синтез включает в себя группу алгоритмов, осуществляющих трансформационные преобразования типа описанных выше. Часть преобразований этапа может быть рассредоточена по другим блокам общего алгоритма.

Синтез выходных словоформ и текста завершает машинную обработку. Кроме синтеза словоформ здесь синтезируются сложные грамматические формы и конструкции, например, пассивные конструкции для перевода глагольных и причастных форм.

Для перевода с английского языка на русский алгоритм модели МПС сводится к следующей последовательности этапов (см. рис. 10).

Содержание этапов следующее.

1. Первичная обработка состоит в поиске словоформ входного текста в словаре основ алгоритма, сопутствующем морфологическом анализе (при ненахождении слова от входной словоформы отбрасываются окончания), замене буквенно-цифровых кодов слов цифровыми эквивалентами и выделении новых слов (им дается специальный цифровой эквивалент и приписывается

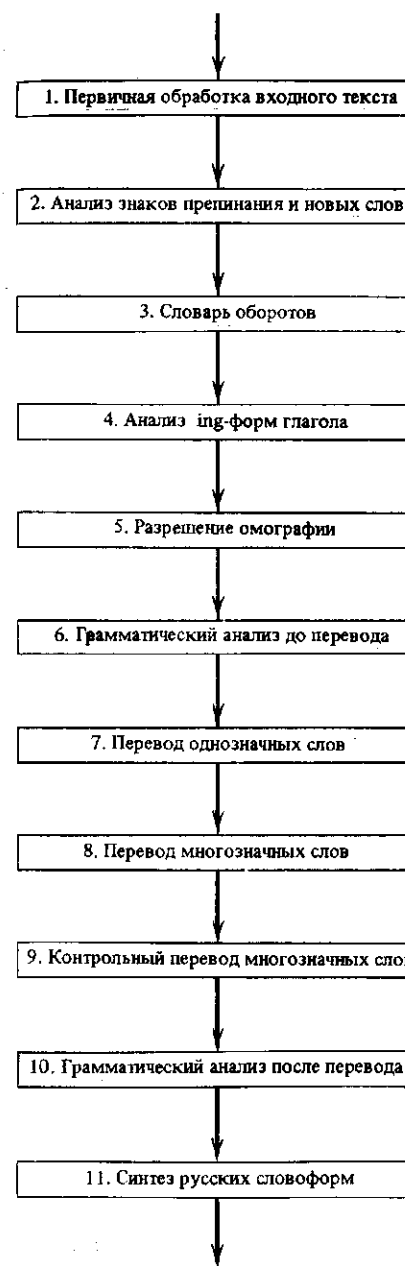


Рис. 10. Алгоритм англо-русского перевода

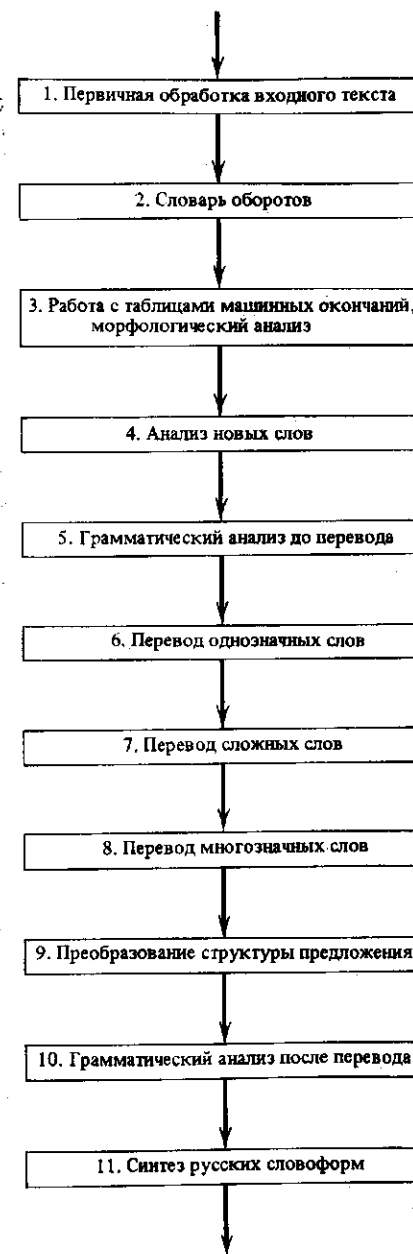


Рис. 11. Алгоритм немецко-русского перевода

морфологическая информация). В результате работы этапа входной текст преобразуется в последовательность информационных ячеек. Каждому слову сопоставляется одна информационная ячейка.

2. На этапе анализа знаков раздела получают разделительные признаки знаки препинания, выполняющие всегда одну синтаксическую функцию: точка с запятой, запятая перед союзами и пр. Для новых слов с помощью анализа контекста и по морфологическим данным определяется лексико-грамматический класс.

3. На этом этапе работает словарь оборотов. В тексте выбираются идиоматические неразложимые словосочетания, каждому из которых ставится в соответствие один цифровой эквивалент. Каждому обороту приписывается определенная синтаксическая функция.

4. На этапе анализа *ing*-форм глагола анализируются некоторые формы, синтаксическая функция которых может быть определена сравнительно просто: отглагольное существительное, герундий в неомонимичных контекстах и пр. Формы, не проанализированные на этом этапе, поступают на вход схемы сказуемого.

5. Этап разрешения лексико-грамматической омонимии (омографии) использует контекстные признаки, главным образом в виде порядка слов и морфологических данных, для разрешения омографии всех словоформ-омографов.

6. Грамматический анализ до этапа перевода состоит в нахождении синтаксических функций слов предложения, а также определении некоторых грамматических категорий английского текста применительно к русскому синтезу. Так, здесь определяются категории «вид» с помощью следующих правил: английские совершенные формы требуют русского совершенного вида при переводе, английские продолженные — русского несовершенного, английские глаголы некоторого списка (в соответствии с приписанными им переводами) — несовершенного, все остальные случаи — совершенного вида (статистически преобладающего).

7. Однозначные слова, т. е. имеющие в данном корпусе текстов один переводной эквивалент, переводятся по таблице соответствий. Содержание этапа и составляет данная таблица соответствий с правилами перевода существительных, получивших СФ «определение», для которых из другой таблицы извлекается соответствующее однокорневое прилагательное или другой какой-нибудь переводной эквивалент для такой ситуации.

8. Перевод многозначных слов осуществляется специальным алгоритмом для каждого многозначного слова.

9. Контрольный перевод многозначных слов нужен для ликвидации последствий сбоя в работе этапа перевода многозначных слов. Этап работает по таблице соответствий, наподобие перевода однозначных слов. Многозначному слову приписывается его наиболее общий и частый перевод.

10. На этапе грамматического анализа после перевода определяются для синтеза грамматические категории по русскому тексту, т. е. фактически по номерам переводов и уже имеющейся в информационной ячейке постоянной и переменной информации. На этом этапе работают схемы определения падежа, числа, рода, схемы согласования глаголов и прилагательных с существительными и схемы определения antecedентов местоимений и местоименных слов. В результате работы этого этапа полностью заполняются информационные ячейки слов.

11. На этапе синтеза по данным заполненной информационной ячейки воссоздается русская словоформа.

Необходимые синтаксические и грамматические преобразования для полного соответствия выходного предложения входному производятся на предыдущих этапах грамматического анализа. Существенная часть преобразований производится при переводе, что будет описано ниже при рассмотрении действия контекстологического словаря.

Для немецко-русского перевода последовательность этапов обработки текста имеет вид (рис. 11).

Содержание этапов следующее:

1. Поиск словоформы текста в словаре и приписывание каждой словоформе информационной ячейки. В результате работы этапа текст заменяется последовательностью информационных ячеек. Обращается информация о заглавных буквах, производится поиск основ сложных словоформ в словаре основ.

2. В тексте выбираются неразложимые фразеологические словосочетания, приписывается грамматическая информация (СФ). Информационные ячейки слов, входящих в оборот, заменяются информационной ячейкой оборота.

3. В информационную ячейку (ИЯ) слов заносится грамматическая информация из словаря. Разрешаются общие случаи лексико-грамматической омонимии: глагол-существительное и пр., в том случае, если эта омонимия может быть разрешена анализом окончаний.

4. Определяется лексико-грамматический класс нового слова (по наличию заглавной буквы и составу окончаний). Производится транслитерация и частичный перевод слов латинского и греческого происхождения и пр.

5. Предварительная сегментация текста, анализ знаков препинания, установление типа порядка слов (для поиска главных членов предложения), приписывание словам синтаксических функций. Последнее начинается с установления СФ для кратких форм прилагательных, причастий, адвербиальных предложных групп. Завершается устранение лексико-грамматической омонимии. Все СФ для слов предложения определяются на этом этапе. Находятся antecedенты местоимений и местоименных слов.

6. Перевод однозначных слов осуществляется приписыванием номеров перевода соответствующим цифровым эквивалентам.

7. Производится перевод основ сложных слов: сначала для основ в виде прилагательных (в составе существительных, прилагательных, наречий), затем в виде существительных в родительном падеже с инверсией именной группы относительно перевода конечной основы, далее перевод остальных случаев сложных основ.

8. Осуществляется вызов алгоритма перевода многозначного слова в зависимости от контекста и работа вызванных алгоритмов.

9. Операции по изменению позиций глагола, преобразования в группе сказуемого, преобразования в определительных и причастных оборотах, преобразования отрицательных конструкций, изменения в порядке слов в других случаях.

10. Установление грамматических значений для русских глаголов, затем для существительных и относящихся к ним определений, установление степеней сравнений для прилагательных и пр.

11. Синтез выходной словоформы по данным заполненной информационной ячейки.

Приведенная выше блок-схема анализа для немецкого текста отражает существенные особенности немецкого синтаксиса и морфологии, которые должны быть отражены в автоматическом анализе и при переводе. Мы можем здесь опять вернуться к последовательности действий переводчика, когда он переводил немецкое предложение в рассмотренном выше примере. Из этой последовательности действий видно, что основную роль в раскрытии структуры предложения в терминах его главных членов играет порядок следования словсформ, особенно тех, которые потенциально могут образовать сказуемое.

В анализ и перевод включены специальные этапы анализа сложных немецких слов с помощью словаря основ, который участвует в анализе тогда, когда искомая словоформа не обнаружена в словаре словоформ. Используется специальная таблица соединительных элементов. На этапах перевода производится перевод отдельных основ сложного слова. Интерредактор имеет возможность корректировать неправильный перевод или набор информации к сложному слову, и в этом случае последствия ошибки на этом этапе для последующего анализа и перевода исключаются.

Тот факт, что в рассмотренной блок-схеме отдельно существует этап преобразования порядка слов, не говорит о существенном отличии от англо-русской блок-схемы. Этот этап содержит практически те же разновидности преобразований, которые в англо-русской схеме осуществляются на этапах грамматического анализа до и после перевода: перестановка определяемого и определяющего слов, изменения порядка слов в отрицательных конструкциях и т. п., т. е. он представляет собой модификацию этапа «Трансформационный анализ и синтез» общего алгоритма.

Общим для англо-русского и немецко-русского алгоритмов является алгоритм синтеза русских выходных словоформ. Опишем основные принципы синтезирования, использованные в последнем блоке алгоритма.

Синтез простых форм осуществляется путем присоединения к основе, выбранной из словаря или выбранной и затем преобразованной, нужного окончания. Синтез сложных форм состоит в образовании форм составных компонентов в отдельности, раздвижке текста и помещении в него в определенном порядке дополнительных слов.

Все многообразие способов образования русских грамматических форм, простых и сложных, можно свести к ряду правил, отображающих некоторую последовательность действий.

Правило 1. Из словаря выбирается нужная основа и к ней присоединяется нужное окончание или формообразовательный суффикс (например, при образовании деепричастий, сравнительной степени прилагательного и т. п.) из таблицы словообразовательных элементов.

По этому основному правилу образуется большинство простых грамматических форм, как то: все формы существительных, полная и краткая форма прилагательных, простая сравнительная степень прилагательных, инфинитив, все личные и временные формы изъявительного наклонения действительного залога (кроме будущего сложного), повелительное наклонение глаголов несовершенного вида и четвертого видового класса глаголов совершенного вида, деепричастие несовершенного вида.

Правило 2. От основы слова, выбранной из словаря, путем суффиксации формируется производная основа и к ней присоединяется окончание. Так образуются причастные формы глаголов несовершенного вида и совершенного вида четвертого видового класса и простая превосходная степень прилагательного.

Правило 3. От основы слова, выбранной из словаря, путем префиксации формируется производная основа и от нее по правилам 1 и 2 образуется нужная форма.

Так образуются все формы глаголов совершенного вида всех видовых классов, кроме четвертого и третьего.

Правило 4. В соответствии с данными, полученными при анализе, из таблиц выбирается полная грамматическая форма слова.

Так образуются личные местоимения (*я, он, мы* и т. п.).

Правило 5. По правилу 1, 2 или 3 образуется нужная грамматическая форма слова и к ней присоединяется справа частица *ся* в соответствии с таблицей *ся*.

Так образуются:

- все формы глаголов с *ся*;
- пассив переходных глаголов несовершенного вида;
- все формы прилагательных с *ся*.

Правило 6. Образуется сложная форма, состоящая из двух изменяемых слов, основного и дополнительного, каждое из которых образуется по правилу 1. Дополнительное слово стоит перед основным и согласуется с ним в роде, числе и падеже.

Так образуется сложная превосходная степень (*самый ближний*).

Правило 7. Образуется сложная форма, состоящая из двух слов, основного и дополнительного, каждое из которых образуется по правилу 1, причем основное слово всегда образует одну и ту же форму (инфинитив), но используя при этом лишь информацию о времени и виде, дополнительное же слово образует разные формы, используя информацию о лице и числе, и стоит оно перед основным.

Так образуется будущее сложное время глаголов несовершенного вида (*будет сообщать*).

Правило 8. Образуется сложная форма из двух слов, основного и дополнительного; основное слово образуется по правилу 3 (если это не глагол четвертого или третьего видового класса), либо по правилу 2, дополнительное слово образуется по правилу 1 или находится перед основным. Основное слово использует информацию о виде и залоге, дополнительное — о времени, числе и лице.

Так образуется страдательный залог (пассив) глаголов совершенного вида.

Правило 9. Образуется сложная форма из двух слов, основного и дополнительного, основное слово изменяемо и образуется по правилу 1, дополнительное — неизменяемо, стоит перед основным.

Так образуется сложная сравнительная степень прилагательного.

Правило 10. Образуется сложная форма, состоящая из двух или трех слов, основного и дополнительного (дополнительных); основное слово образует по правилам 1, 2, 3, 5, 8 определенную форму (например, инфинитив или прошедшее время), дополнительное слово представляет собой частицу и стоит после основного слова.

Так образуется сослагательное наклонение глагола.

Следует выделить некоторые особенности образования глагольных форм, связанные с существенными различиями в видовременной системе глагола. Эти различия имеют большое значение в передаче соответствий; они связаны с оформлением будущего времени и страдательного залога.

Будущее простое время состоит из одного слова: *сообщит, решит*. Будущее сложное — из двух слов: *будет сообщать, будет решать*. Страдательный залог может выражаться одним словом — различными временными формами глагола с *ся* — *дом строится*. Он может также выражаться двумя словами — формами вспомогательного глагола *быть* и кратким страдательным причастием (*дом был построен*).

Поэтому при переводе форм будущего времени действуют следующие правила:

— при наличии у переводимого глагола признаков «будущее время, совершенный вид» образуется будущее простое по правилам 1, 3;

— при наличии у переводимого глагола признаков «будущее время, несовершенный вид» — образуется будущее сложное по правилу 7.

В обоих случаях английский вспомогательный глагол не переводится, а информация, которую он несет, заносится в ИЯ смыслового глагола и используется при образовании русской грамматической формы.

При переводе пассивных конструкций предусмотрено образование двух структур:

— при наличии у смыслового глагола признаков «пассив, несовершенный вид» — образуется форма глагола с *ся* по правилу 5;

— при наличии у глагола признаков «пассив, совершенный вид» — образуется структура из вспомогательного глагола *быть* и краткого страдательного причастия смыслового глагола по правилу 8. При переводе английского сослагательного наклонения происходит замена на русскую структуру по правилу 10. Английский вспомогательный глагол здесь также не переводится, а его информация заносится в ячейку смыслового глагола. Модальные глаголы во всех этих случаях переводятся, так как сами они не являются носителями только грамматических признаков.

В алгоритме синтеза имеются некоторые особенности в образовании причастий и деепричастий: они образуются от основ глагола путем прибавления к ним суффиксов причастия и деепричастия (из соответствующих таблиц), и последующего присоединения окончаний соответствующих классов прилагательных (для причастия).

Трансляционная грамматика есть грамматика, ориентированная на перевод. Включение ее в алгоритмический элемент динамического компонента не означает ее полного слияния с алгоритмом. Она выступает в качестве отдельной самостоятельной составляющей. На перевод ориентированы ее единицы грамматического описания. В то же время они структурно и физически независимы от собственно частных алгоритмов (схем), допускают самостоятельную корректировку и пополнение. Состав и структура грамматических единиц трансляционной грамматики, их порядок следования отвечают требованиям схемы: в первую очередь определяются наиболее массовые случаи, которые должны решаться с оптимальной эффективностью.

Структуру трансляционной грамматики рассмотрим в следующей последовательности:

— упорядочение формальных признаков английского предложения в их связи со структурой алгоритма перевода;

- диагностирующие конструкции,
- детерминанты.

Английский язык в сопоставлении с русским обладает рядом алгоритмизируемых особенностей. Наибольшим числом формальных признаков обладает глагол, наименьшим — имена. Значительная информация может быть извлечена из порядка следования слов в предложении. Поэтому анализ целесообразно строить, исходя из глагольных форм, их морфологии и порядка следования. Сходство между английским, французским и русским языками в виде соответствия синтаксических конструкций облегчает задачу перевода⁶.

Синтаксический анализ может быть произведен лишь тогда, когда разрешена лексико-грамматическая омонимия (омография, пример: *second* — существительное, глагол, прилагательное). В интегральных (непрерывных) методах синтаксического анализа типа предсказуемого или метода фильтров омография передается на последующие этапы и разрешается одновременно с установлением синтаксической структуры предложения. В модели МПС омография составляет самостоятельный этап, работа которого предшествует дальнейшим этапам алгоритма, производящим анализ и синтез. Изучение всех омографов в заданном корпусе текстов позволяет выделить классы их и в дальнейшем рассматривать способы разрешения омографии для классов омографов. Так, слово *back* входит в класс омографии «глагол—существительное—прилагательное—наречие», слово *second* входит в класс «существительное—глагол—прилагательное» и т. п. Ясно, что классы омографии не могут быть закрытыми; установлено, что закрытость классов есть некоторая функция от числа значений в каждом классе: чем больше значений, тем меньше класс и тем меньше вероятность пополнения этого класса новыми членами.

Этап разрешения омографии использует самые разнообразные признаки — порядок следования словоформ, дистрибутивные характеристики отдельных элементов, морфологические признаки. Если есть возможность сначала сузить омографию, то она используется: так, окончание *s* у английской словоформы *presents* позволяет исключить значение «прилагательное». Важны два момента — опора на индивидуальные лексические и грамматические свойства слов внутри заданного класса омографии и учет статистически наиболее значимых ситуаций, когда контекст не дает возможности принять решение. Определенные дополнительные возможности создает последовательность

⁶ «Сходство между синтаксическими конструкциями английского и французского языков, с одной стороны, и русского, с другой, нередко дает возможность дословного перевода: без изменения структуры предложения и без существенного изменения порядка слов» (Рецкер 1974, 77).

омографов, в которой могут быть разрешены неоднозначности отдельных членов (Марчук 1973а).

Необходимо, чтобы ошибки в разрешении омографии были сведены к минимуму, поскольку они наиболее существенным образом сказываются на последующем анализе и переводе. Исправление ошибок этапа омографии на последующих этапах не может быть произведено.

Особое значение имеет анализ словоформ с окончанием *-ing*. Предварительный анализ словоформ с *-ing* заключается в том, что в составе таких словоформ выделяется глагольная основа и суффикс (при поиске по словарю и сопутствующем морфологическом анализе). Затем среди этих словоформ определяются выполняющие синтаксическую функцию существительного: отглагольное существительное и герундий. (Различие между ними для автоматического синтаксического анализа нерелевантно.) Словоформы с *-ing*, получившие признак существительного, в дальнейшем анализе участвуют как существительные.

Особую роль играет признак определения в применении к глаголу. Глагол в СФ определенно есть причастие, действительное или страдательное, настоящего или прошедшего времени: под эти категории попадают английские глагольные словоформы с окончаниями *-ing* и *-ed*. Признак обстоятельства получают глаголы, имеющие окончание *-ing* и выполняющие функции обстоятельства.

Таким образом, при анализе английского глагола используются следующие формальные признаки:

- окончания глагольных словоформ (*-ing*, *-ed*);
- некоторые формы исследуемых глаголов, такие как формы первого, второго и третьего лица модальных и вспомогательных глаголов;
- позиция исследуемого глагола относительно модального или вспомогательного глагола, а также относительно существительных в предложении.

В анализе по СФ мы отходим от принципа выделения синтаксических групп (грамматических групп), соответствующих членам предложения, подобно тому, как это было сделано в алгоритме Бельской. С точки зрения синтеза и определения формообразующих категорий для слов здесь целесообразна редукция, но не такая, как в традиционной картине, рассмотренной выше, при которой одно слово представляет собой один член предложения, а некоторая рассредоточенная редукция, при которой признак члена предложения приписывается каждому слову группы.

В анализе используются также и имплицитные формальные признаки, к числу которых можно отнести, например, принадлежность словоформ глагола к вторым и третьим формам спряжения неправильных глаголов, супплетивность, чередование корневых морфем и т. п.

Некоторые категории определяются из соображений алгоритмического удобства. Так, признак сказуемого приписывается инфинитиву для противопоставления его глагольным формам с признаками СФ «определение» и «обстоятельство». Последующий анализ приписывает инфинитиву информацию «инфинитив». Определение СФ «сказуемое» основывается на следующих главных признаках:

— наличие перед анализируемым глаголом модального или вспомогательного. Сами модальные и вспомогательные глаголы;

— глаголы с окончанием *s* третьего лица единственного числа;

— вторая форма сильных глаголов или супплетивных. Глаголы с омонимичными формами анализируются специальными схемами.

Кроме этих признаков, существуют также и другие (порядок слов, особенности глагольной дистрибуции и пр.).

Английское существительное имеет гораздо меньше формальных признаков, чем глагол. Некоторые существительные получают признак СФ в начале анализа, когда выделяются сложные предлоги и наречия (*in case of, with a view to*). В последовательности существительных все существительные принимаются как определения к последнему существительному в ней. Это решение оставляет возможность для дальнейшего алгоритмического уточнения информации.

Некоторой опорой анализа являются аббревиатуры типа *dr.*, которые помогают выделить собственные имена.

Схема определения подлежащего основана на отрицательных признаках английского подлежащего в тексте: отсеиваются случаи, когда существительное не является подлежащим, и всем, не прошедшим этот фильтр, присваивается соответствующий признак.

Признак «подлежащее» не получают существительные: а) перед которыми стоит предлог, непосредственно или отделенный прилагательным, причастием и пр.; б) стоящие в притяжательном падеже; в) стоящие в предложении после глагола с признаком «сказуемое» (за исключением вопросительных предложений); г) местоимения в косвенном падеже (*him, them*).

Возможны случаи неправильного определения подлежащего; так, признак этот по схеме получают некоторые беспредложные дополнения. Для ликвидации случаев неправильного определения подлежащего вводится в работу вторая часть схемы, которая выбирает в предложении первое по порядку подлежащее и осуществляет анализ по следующим основным правилам:

1) подлежащее в предложении должно быть одно (кроме случая однородных членов); 2) личные местоимения *I, he, she* остаются подлежащими в любых ситуациях; 3) если имеется группа подлежащих в виде однородных членов, то все они остаются подлежащими; 4) название должности, ученое звание и пр., стоящее перед собственным именем, определенным как подле-

жащее, тоже остается подлежащим. Вторая часть схемы зачеркивает в предложении слева направо все подлежащие, оставляя первое. При такой организации схемы: а) упрощается первая часть, где при другой методике пришлось бы вводить сложные проверки для отсеивания беспредложных дополнений, особенно с соединительными союзами; б) способ зачеркивания лишних подлежащих легко осуществим в стандартных операторах и схема получается достаточно компактной.

Схемы определения сказуемого и подлежащего, построенные с учетом описанных выше закономерностей, решают наиболее частые и относительно простые случаи. Принятые решения не противоречат усложнению схем по мере выявления и алгоритмизации дополнительных проверок контекста.

Прилагательное может быть либо сказуемым (именная часть составного сказуемого), либо определением. Оба случая решаются довольно простым контекстным анализом. Предлог представляет трудности для определения СФ только в случае омонимии с союзами или наречием, однако эта омонимия разрешается ранее. Заметим, что при этом разрешении приходится использовать в ряде случаев почти ту же информацию, которая нужна для последующего определения СФ слов; в таком случае разница заключается в исходном кодировании этой информации.

Наречия получают признак «обстоятельство», союз — признак «разделительный», в том случае, если он знаменует границу предложения, и не получает такого признака, если соединяет однородные члены. В английском языке все союзы разделительные, кроме трех: *and, but, or*, анализ которых на указанный предмет производится специальной схемой. Знаки препинания получают тот же признак в тех же ситуациях: особая роль в этом отношении принадлежит запятой, которая выполняет те же функции, что и названные выше три союза.

Таким образом, в результате такого анализа все словоформы текста получают признак СФ и готовы участвовать в дальнейшем анализе и определении бинарных категорий, необходимых для перевода, последующего грамматического анализа и синтеза.

Если перейти от содержательного описания грамматических средств к более формальным правилам их использования в модели, то следует привлечь понятия «диагностирующая конструкция» в случае, когда мы говорим об установлении соответствий на уровне предложения и о структуре схем грамматического анализа, и «детерминанта», в тех случаях, когда соответствия устанавливаются на уровнях слов и словосочетаний. Эти понятия будут раскрыты в описании динамического компонента. Понятия контекстуальной модели перевода и элемента контекстуальной модели будут рассмотрены в главе о контекстологическом словаре.

Словарь как составляющая алгоритмического элемента ди-

намического компонента делится дихотомически по трем разным основаниям:

- словарь входного/словарь выходного языков;
- однозначные/многозначные слова;
- фразеологические единства (обороты)/отдельные лексемы.

Словарь оборотов включает фразеологические единства типа *for many years* 'в течение ряда лет', *in accordance with* 'в соответствии с' и т. п. В связи с тем, что точные критерии фразеологической связанности отсутствуют, принят принцип, согласно которому в словарь оборотов включаются единства, имеющие устойчивые переводные эквиваленты и выполняющие в предложении одну синтаксическую функцию.

3.9. Программный элемент динамического компонента

Несмотря на то, что термин «программный», как указано ранее, применяется здесь довольно условно, поскольку работа в целом посвящена лингвистическому обеспечению систем МП, этот элемент должен излагаться хотя бы в самой общей связи с основными принципами математического и собственно программного обеспечения, из которого наиболее важно представление о стандартных операторах.

Общие принципы математического обеспечения заключаются в следующем: учет специфики создания и эксплуатации в виде большого разнообразия используемых информационных массивов (неоднородность информации), необходимость постоянной коррекции и пополнения данных системы, потребность в учете различных факторов при этой коррекции.

Основные особенности математического и программного обеспечения системы АМПАР следующие:

- использование специализированного языка программирования для записи частных алгоритмов МП (схем);
- наличие двух баз обрабатываемых данных (отладочного и рабочего массивов);
- разбиение обрабатываемых массивов на тематические подмассивы;
- использование языковых средств (языка управления процессом) для описания инструкций ввода—вывода обрабатываемых массивов и режимов их обработки;
- возможность реорганизации структуры системы (создание различных версий для выбора наиболее эффективного варианта системы);
- возможность получения результатов работы системы с любого этапа в удобном для анализа виде;
- разбиение процесса перевода на несколько стадий.

Процесс перевода в системе АМПАР разбит по функциональным признакам на несколько крупных стадий:

- предварительной обработки;
- интерредактирования;

- автоматического перевода;
- постредактирования;
- выдачи результатов.

В свою очередь, стадия может быть разбита на несколько этапов, служащих для реализации наиболее крупных алгоритмов системы. Этапы, которые в процессе эксплуатации будут часто подвергаться модификации, разбиты на этапные схемы, каждая из которых реализует частный алгоритм анализа или перевода. Схема состоит из схемных операторов (стандартных операторов), реализующих некоторую лингвистическую операцию. Эти операторы составляют специальный язык, ориентированный на использование лингвистами. На других этапах используются этапные программы, составленные на языке «ассемблер» с целью минимизации времени работы этих программ.

Использование модульного принципа построения программной части системы, при котором алгоритм разбит на достаточно мелкие частные алгоритмы и каждый частный алгоритм программно реализуется некоторым модулем, обеспечивает существенное упрощение процесса программирования и позволяет достаточно просто генерировать различные версии системы при ее отладке и коррекции в процессе эксплуатации.

Опишем систему стандартных операторов, так как она предназначена непосредственно для использования лингвистов и в значительной мере определяет собственно лингвистические возможности алгоритмов и системы в целом, поскольку очерчивает круг операций над текстом.

Можно определить главные требования, которым должен удовлетворять язык программирования для промышленной системы МП. Эти требования следующие:

- язык должен быть достаточно удобным для человека, не знакомого с программированием как таковым. Языком должен пользоваться лингвист, хорошо знающий собственно лингвистические особенности и возможности системы для постоянного внесения корректив в словарь и схемы анализа и синтеза;
- язык должен быть достаточно близок к машинному языку, с тем, чтобы не допускать существенного увеличения времени работы алгоритма и его частей, поскольку для МП время работы практически является одним из важных показателей эффективности системы.

Эти требования в значительной мере противоречат друг другу, поскольку известно, что удобные для человека языки программирования высокого уровня обычно требуют много времени работы транслятора, что сказывается на скорости перевода, а языки, близкие к машинным кодам, мало удобны для человека, хотя и дают хорошую скорость работы программ.

Удобный язык программирования фактически определяет возможности системы МП. Так, описание Джорджтаунской системы начинается с описания языка SLC, предназначенного для пользователя: тем самым подчеркивается, что пользование си-

стемой начинается с овладения языком. Для потребителя, использующего систему пассивно, достаточно поверхностного знания языка, для желающего активно работать над совершенствованием системы в применении к новому материалу требуется довольно глубокое знакомство с языком программирования (Jordan, Brown, Hutton 1977).

Если рассматривать язык программирования МП с этой точки зрения, то система стандартных операторов, которая будет ниже описана, превосходит язык SLC в том отношении, что знакомство с операторами позволяет осуществлять сразу две цели, а именно использование системы и ее совершенствование и пополнение для применения на новом материале. Главное отличие системы стандартных операторов АМПАР от языка SLC в том, что первый не охватывает таких малозначительных для лингвиста операций как поиск по словарю, морфологический анализ, сопутствующий этому поиску, и пр. За счет сосредоточения внимания на нескольких главных лингвистических действиях язык стандартных операторов более понятен лингвисту и более эффективен. Для овладения SLC надо быть знакомым с программированием, язык стандартных операторов этого не требует.

Выделение стандартных операций, производимых лингвистом над текстом, стало основой для разработки программ операторов, которые представляют собой стандартные подпрограммы, применяемые всякий раз, когда требуется осуществить данное лингвистическое действие над текстом. Язык стандартных операторов для записи алгоритма МП был разработан Ю. А. Моторным и опубликован в 1970 г. (Моторня, Марчук 1970).

Позже язык стандартных операторов был модифицирован Б. Д. Тихомировым (Драмбян, Истомина, Семенова и др. 1977), который ввел новые операторы, дав тем самым лингвисту дополнительные возможности работы над текстом. Однако для целей настоящего исследования достаточен первый вариант языка, который мы вкратце и опишем.

Язык стандартных операторов состоит из восьми операторов, каждый из которых выполняет одно или несколько точно определенных лингвистических действий. Операторы и выполняемые ими действия следующие:

Оператор А служит для нахождения в тексте одного или нескольких конкретных английских или русских слов или для проверки слова, найденного предыдущим оператором схемы, на конкретный ЦЭ или список ЦЭ.

Оператор В предназначается либо для проверки у слова, найденного ранее, заданного признака, либо для поиска в пределах окружающего контекста слова с заданными признаками.

Оператор С служит для проверки или поиска слов, принадлежащих к определенным классам или подклассам.

Оператор D предназначается для занесения номера перевода в ИЯ слов текста.

Таблица 8. Структура операторов для записи схем алгоритма МП

Оператор	Разряды, №№				ЦЭ	Сдвиг
	45...41	40...36	35...26	25...21		
A	Код оператора				ЦЭ	Сдвиг
B	Вид работы					
C	Адрес отхода		Адрес отхода	Адрес проверяемого слова	ЦЭ/МП	—
D	Адрес отхода		Адрес проверяемого слова	Адрес проверяемого слова	МП	—
E	Адрес отхода		Адрес слова	Адрес слова	МП	Сдвиг
F	Адрес отхода		Адрес отхода	Вид отбора	ЦЭ/МП	Сдвиг
G	Адрес отхода		Адрес отхода	№ подпрограммы пропусков	МП	—
H	Адрес отхода		Адрес отхода	№ ЦЭ в таблице	ЦЭ	—

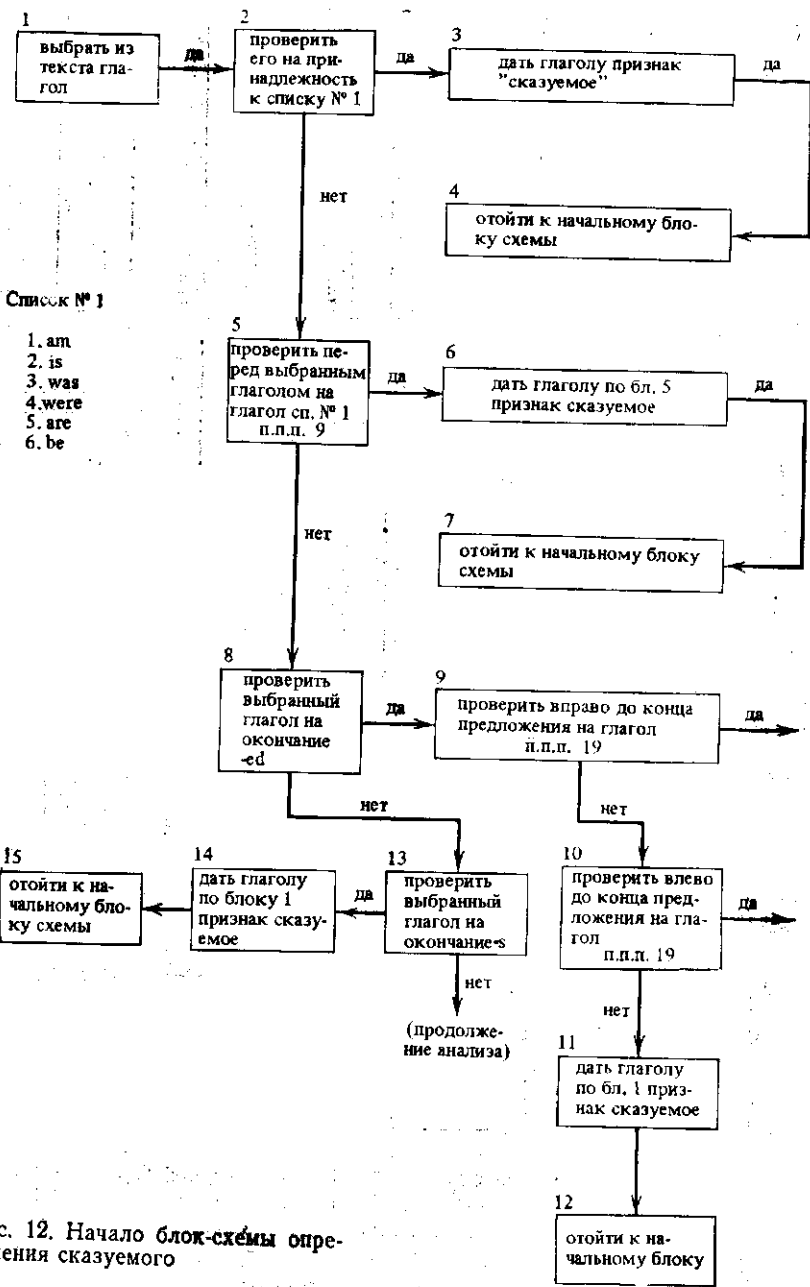


Рис. 12. Начало блок-схемы определения сказуемого

Список № 1

1. am
2. is
3. was
4. were
5. are
6. be

Таблица 9. Пентады

№№ операторов		9	8	7	6	5	4	3	2	1	
1	F	1		отход к след. блоку	0	0	1	0	0	0	Выбор из текста глагола
2	A	1	0	5	0	0	0	0	1	6	Проверка на сп. № 1
3	E	0	0	1	0	0	3	2	12	12	Дать признак сказуемого
4	H	0	0	1	0	0	0	0	0	0	Отойти к блоку 1
5	A	1	0	8	9	0	0	3	3	6	Сп. № 1
6	E	0	0	1	0	0	3	2	12	12	
7	H	0	0	1	0	0	0	0	0	0	Окончание -ed
8	B	0	0	13	0	1	2	2	21	21	Глагол
9	C	1	0	10	0	0	0	19	0	0	Глагол
10	C	1	0	11	0	0	0	19	0	0	
11	E	0	0	1	0	0	3	2	12	12	
12	H	0	0	1	0	0	0	0	0	0	Окончание -s
13	B	0		отход к след. блоку	0	1	2	1	21	21	
14	E	0	0	1	0	0	3	2	12	12	
15	H	0	0	1	0	0	0	0	0	0	

и т. д. схема продолжается.

Оператор E заносит признаки, выработанные в результате работы схем, в ИЯ слов текста.

Оператор F — оператор-диспетчер, который выбирает слова с заданными признаками.

Оператор H — является служебным оператором и используется при переадресации отходов при работе схем. В таблице 8 отражена структура операторов по пентадам.

Для иллюстрации использования стандартных операторов в записи схем анализа приведем начальную часть схемы определения СФ сказуемого для глаголов.

В целях наглядности представим процедуру в виде графической блок-схемы (см. рис. 12).

В этой блок-схеме содержится пример использования операторов для решения вопросов лингвистической логики: так, проверка, является ли данный глагол единственным в предложении, осуществляется блоками 9 и 10, с отсеиванием случаев, подходящих под категорию «нет».

В операторной записи эта часть схемы определения сказуемого выглядит следующим образом (см. табл. 9).

Теперь можно описать составляющие программного элемента динамического компонента.

Составляющая, которая названа «списки», реализует простейшие лингвистические операции типа поиска в словаре (срав-

нение двух списков на совпадение), перевода однозначных слов, работу с единицами выходного словаря в простых случаях синтеза. В программировании этого этапа лингвист не использует стандартных операторов, практически этап реализуется программистами с помощью специальных программ.

Составляющая «схемы лексического анализа и перевода» включает схемы перевода многозначных слов и схемы сложных словосочетаний, составляющие словарь оборотов. Перевод многозначных слов осуществляется специальным контекстологическим словарем. Здесь используются стандартные операторы, однако не в полном их составе, а только четыре из них, что достаточно для реализации теории детерминант. Использование операторов открывает следующие возможности:

— обработку (в том числе приписывание нулевого перевода, перестановку слов и т. п.) четырех слов помимо основного, посредством применения оператора D с дополнительной ячейкой;

— неограниченное пополнение списков слов и введение в схемы новых списков, посредством применения оператора A;

— включение в схемы новых операторов, реализующих новые проверки.

Эта составляющая позволяет устанавливать широкий диапазон переводных соответствий на уровнях слов и словосочетаний. Кроме того, состав лингвистических операций, содержащихся в ней, удовлетворяет требованиям более сложных случаев синтеза, чем описываемые составляющей «списки».

Составляющая «схемы грамматического анализа и преобразований» охватывает сложный комплекс грамматических схем, реализующих варианты и трансформационные соответствия на разных этапах алгоритма. В этих схемах все возможности всех стандартных операторов используются в полном объеме, в результате чего становится возможным использовать:

— отдельные слова, списки, классы и подклассы слов для выявления диагностических конструкций в полном объеме транслационной грамматики;

— осуществлять преобразования типа перестановок слов, вставки новых слов, поиска antecedentов за пределами предложения;

— пополнять списки схем, включать новые классы и подклассы слов, включать в систему признаков новые признаки, которые могут быть заданы предметным компонентом;

— включать в схемы анализа и перевода новые операторы;

— включать в этапы новые схемы, исключать старые схемы, менять порядок следования схем в этапах и т. д.

Для иллюстрации того, каким образом работают схемы и как осуществляется важнейшая часть анализа — синтаксический анализ, приведем перечень схем 6-го этапа системы АМПАР с краткими указаниями на исходные данные и совершаемые схемами действия. Для наглядности и удобства понимания диагно-

стические конструкции будут даваться нами в содержательном описании, без какой-либо специальной символики.

К началу работы этапа омография уже разрешена.

Первые схемы этапа определяют синтаксическую информацию для случаев, когда она получается сравнительно легко, например, проверкой окружающих слов на лексико-грамматический класс. Сначала следует схема разбивки фраз на простые предложения. Эта схема выбирает союзные существительные (*who*) и союзные прилагательные (*what, whatever*), и, если перед нами нет знака раздела в виде запятой, раздела фраз, союзных существительных и прилагательных, проставляет такой знак в виде запятой. Эта искусственно восстановленная запятая в дальнейшем служит важным знаком раздела предложения. Следующие две схемы приписывают СФ «обстоятельство» наречиям и СФ «предложная» — предлогам. Схема определения функции числительного анализирует числительное, которое, в зависимости от диагностической конструкции, получает СФ «дополнение», «определение» или никакой (для последующего решения).

Схема СФ «определение» для существительного выбирает существительное с отсутствующей СФ, осуществляет ряд проверок контекста на классы и подклассы слов и в результате приписывает существительному СФ «определение», если это необходимо. Пример: в сочетании *coordination center* слово *coordination* получает СФ «определение». В дальнейшем обработка такого сочетания происходит так (на этапе синтеза): если от слова *coordination* есть прилагательное—определение 'координационный', то оно выдается в качестве переводного соответствия, если такого прилагательного—определения нет, то производится перестановка и словосочетание получает перевод 'центр координации'.

Схема затирания СФ у отглагольных существительных (анализируются предшествующими этапами) корректирует схему СФ «определение» для случаев, когда после отглагольного существительного стоит другое существительное. Схема приписывания признака неоднородности союзам и восстановленной запятой заносит этот признак союзам, которые им обладают.

Схема СФ «сказуемое» для глаголов (кроме тех, которые имеют окончание *-ing*) состоит из шести частей. Она производит анализ последовательно каждого встретившегося во фразе глагола и приписывает ему одну из СФ: «сказуемое», «определение», «обстоятельство». Глагол, получивший СФ «сказуемое», переводится личной формой либо инфинитивом, «определение» — причастием, «обстоятельство» — деепричастием. Первая часть схемы после анализа наиболее простых случаев обрабатывает глаголы, 2-я и 3-я формы которых совпадают (*brought, brought*). Влево от исследуемого глагола делается проверка на парадигму глаголов *be* и *have*, и при положительном ответе глагол получает признак «сказуемое». Здесь обрабатывается также во-

просительная форма для таких глаголов. Наличие вопросительного знака в конце предложения определяет смену направления проверки. Работа второй части схемы заключается в том, что признак «сказуемое» дается вспомогательному и модальному глаголам. Здесь предусматривается также случай инверсии при вопросительной конструкции предложения: *Are the goods transported by air?* При проверке влево от основного слова на список глаголов *seem, become, appear* основной глагол получает признак СФ «определение» и переводится причастием.

Пример: *He seemed outraged by the results of the investigation.* «Он казался рассерженным результатами расследования».

В третьей части схемы обрабатываются неправильные глаголы в 3-й форме типа *written* и правильные глаголы с окончанием *-ed*. Предварительно осуществляется проверка на наличие у данного глагола СФ (которая может быть дана предыдущими частями схемы). В этой части решаются случаи, когда глагол отделен от вспомогательных глаголов целым вводным предложением. Анализируется ситуация типа: *The government was, according to information from various sources, forced to take drastic measures.* Особо рассматриваются случаи, когда перед единственным глаголом в предложении стоит частица *to* или слово из списка *how, when, while, what, which*. Такие глаголы получают признак «сказуемое».

В четвертой части схемы рассматриваются ситуации, не подвергавшиеся анализу в предыдущих частях: глаголы в форме настоящего времени, глаголы с *-ed* без модальных и вспомогательных перед ними, однородные глаголы-сказуемые и ряд других ситуаций. Глаголы в форме настоящего времени Present Indefinite получают признак «сказуемое», в форме прошедшего времени Past Indefinite тот же, но по «общему выходу» схемы. Случаи с однородными глаголами-сказуемыми решаются с помощью проверок на наличие союзов и запятых, которые в момент работы этой части схемы выступают как однородные.

Пятая и шестая части схемы решают менее распространенные случаи, в частности такие, когда перед глаголом с окончанием *-ed* стоит однородная точка или предложный оборот, перед которым знак раздела фразы.

Затем работает схема определения неоднородности запятой, которая состоит из 2-х частей.

Эта схема нужна для членения фразы на простые предложения. Ситуации, когда получение запятой признака неоднородности значения не имеет, не рассматриваются в схеме и запятая получается однородной. Например: *Thirdly, the call to revolt was...* Запятая после *thirdly* может быть или не быть однородной. То же с обращениями, выделяемыми с обеих сторон — эти запятые также остаются однородными. Признак неоднородности дается запятой в следующих основных случаях, когда:

а) непосредственно за ней следует предлог, за которым идет союз с признаком неоднородности;

б) непосредственно за ней следует союзное существительное или союзное прилагательное;

в) непосредственно за ней следуют кавычки, за которыми идут личные местоимения и глагол-сказуемое;

г) нераспространенное обращение вводится словом *dear*;

д) за запятой следует распространенный оборот с *there* и модальным или вспомогательным глаголом. Пример: *..., there was confidence that the task would be achieved step by step*;

е) перед запятой имеется глагол-сказуемое, а после запятой идет существительное, после которого стоит глагол-сказуемое: случай, когда запятая расчленяет сложное распространенное предложение на простые;

ж) влево от запятой идет существительное, а вправо союзы *and, but* и *or* с последующими существительными и глаголами. При наличии однородных членов предложения схема дает запятой признак «однородная». Общий выход — признак однородности — дается во второй части схемы.

После работы этой схемы повторяется работа 3-ей, 4-ой и 5-ой частей схемы определения СФ глаголов с окончанием *-ing*. Это нужно для определения указанной СФ для тех глаголов, которые не получили СФ на 4-м этапе работы алгоритма вследствие неразрешения там омографии. Далее работает ряд схем служебного характера по замене ЦЭ в соответствии с переводами определений.

Затем следует схема определения СФ для глагола *have* с окончанием *-ing*. Признак «сказуемое» дается *having*, если:

а) перед ним имеется парадигма *be*, например: *The copies were having been distributed among...*;

б) перед ним имеется парадигма *be*, а перед ней стоит знак раздела фразы, т. е. случай, когда *having* употреблено в предложении вопросительного типа;

в) перед *having* стоит однородный союз *and, but, or*, а перед ним сказуемое с окончанием *-ing*.

СФ «обстоятельство» дается *having*, если:

а) вправо от него стоят слова *be, been*, а влево существительное, причем после *be, been* следует глагол, а после этого глагола — глагол-сказуемое;

б) последний глагол не является сказуемым, а имеет после себя еще один глагол, уже получивший признак «сказуемое»;

в) слева от *having* имеются слова *by, without* или *after*;

г) влево от *having* находится существительное, а вправо — глагол-сказуемое.

СФ «определение» дается *having* в том случае, если:

а) вправо от него находятся глаголы, не имеющие признака «сказуемое», а перед ним имеется существительное;

б) влево имеется однородный союз *and, but, or*, а перед ним стоит глагол с СФ «определение» с окончанием *-ing*;

в) перед ним стоит наречие *as*. Пример: *This is conceived as having three aspects.*

Признак отглагольного существительного *having* получает, если:

а) после *having* стоят слова *be, been, being* с последующим глаголом, а перед ним предлог, кроме предлогов *by, without, after*;

б) в той же ситуации, но между *having* и предшествующим предлогом имеется отрицание *not*;

в) перед *having* стоит однородный союз *and, but, or*, а перед ним глагол с окончанием *-ing*, уже получивший признак отглагольного существительного;

г) непосредственно перед *having* стоят глаголы *like, mind*, а также в некоторых других случаях.

Затем действует схема определения СФ для глагола *be* с окончанием *-ing*. Эта схема, как и предыдущая, анализирует диагностические конструкции для этой частой словоформы, играющей значительную роль в анализе.

После работы ряда вспомогательных схем, уточняющих СФ для отдельных случаев, работает схема определения СФ прилагательного, которая выбирает и анализирует прилагательное в зависимости от позиции существительного. Прилагательное, получившее СФ «сказуемое», на этапе синтеза обращается в крайнее прилагательное — именную часть сказуемого.

Далее работает схема особого рода — схема восстановления союза *that*. Восстановление союза необходимо для указания границ предложения. Схема основывается на том положении, что для ряда существительных, глаголов, прилагательных, при соблюдении ряда условий, выражающихся в следовании классов слов с СФ, после этих слов должен следовать союз *that*, часто опускаемый как избыточный. В русском тексте этот союз должен быть также восстановлен, поскольку ему следует приписать перевод «что», без которого русское предложение будет неполным.

Восстановление союза *that* является частичным решением общей проблемы «нормализации» текста, т. е. восстановления опущенных в нем элементов формальной структуры, которые необходимы при переводе. Это специальная проблема, которой мы касаться не будем. Заметим только, что степень и характер «нормализации» должны определяться потребностями выявления переводных соответствий.

Следующей работает схема определения неоднородности союзов *and, but, or*, которая определяет СФ «разделительная» для этих союзов. Неоднородный союз разделяет, однородный — соединяет предложения в составе фразы и однородные члены. Такое членение необходимо для анализа. Основной принцип схемы: если справа и слева от союза имеются существительные, а соответственно левее и правее их — глаголы с СФ «сказуемое», то такие союзы получают признак «неоднородный». Отсюда должны быть исключены случаи, когда СФ «сказуемое» приписана глаголу-инфинитиву.

Затем после ряда вспомогательных схем работает схема определения подлежащего, состоящая из трех частей. Основные положения схемы описаны ранее. Дальше особая схема определяет подлежащее в оборотах *there is (are...)*. Схема подлежащего, таким образом, завершает определение СФ слов предложения. Существительные, не получившие СФ, классифицируются как «дополнения».

Остальные схемы этого этапа: схема определения неоднородности скобки, первая часть схемы определения числа существительного, схема определения одушевленности существительного, схема определения степени прилагательного, наклонения глагола-сказуемого, вида, времени, лица, числа, пассивной формы, одушевленности, инфинитива для глагола-сказуемого и глагола-причастия (определения).

Далее работают схемы, осуществляющие перестановку и необходимый перевод отрицательных конструкций.

3.10. Эффективность работы системы и качество перевода без редактирования

Рассмотрим следующие вопросы:

— является ли система такого рода алгоритмически реализуемой, т. е. возможно ли ее осуществление на ЭВМ с требуемыми параметрами. Вопрос этот в отношении систем МП является актуальным, поскольку многие проекты систем оказались неосуществленными и не осуществимыми в ближайшем будущем даже на всех мыслимых ЭВМ, так как предусматривали для реального подязыка огромный перебор вариантов;

— каково качество получаемого машинного продукта;

— какие переводные соответствия и языковые явления, существенные для перевода, вызывают затруднения (неправильно переводятся) или остаются вообще вне рамок модели, т. е. какова разрешающая сила модели в общем случае.

По первому вопросу показательным является список машинных параметров, полученных при реализации системы на ЭВМ второго поколения. Перевод текстов газетного характера был осуществлен со скоростью примерно 7 знаков/сек, качество перевода приемлемое. Этот широкий эксперимент с распечаткой информационных ячеек промежуточного текста описан нами подробно в (Марчук 1973) и здесь воспроизводиться не будет. Приведем таблицу, показывающую затраты времени на реализацию машинных этапов системы от входа до выхода.

Из данных табл. 10 видно, что основное машинное время затрачивается на этапы грамматического анализа до и после перевода, а также на этап синтеза. В отношении последнего можно сказать, что его реализация в стандартных операторах, которая в данном эксперименте не была осуществлена, даст возможность, как показали последующие эксперименты и подсчеты, значительно сократить машинное время на этот этап. Отно-

Таблица 10

Название и функции этапа	Время работы (сек)
1. Первичная обработка	34,0
2. Предварительный этап	20,3
3. Словарь оборотов	2,0
4. Анализ глагольных конструкций	39,5
5. Разрешение омографии	81,0
6. Грамматический анализ до перевода	130,0
7. Перевод однозначных слов	6,5
8. Перевод многозначных слов	9,6
9. Контрольный перевод многозначных слов	3,0
10. Грамматический анализ после перевода	114,0
11. Синтез русского текста	162,0
	601,9

сительно грамматического анализа следует иметь в виду, что данная организация этапов наиболее эффективно работает по сравнению с другими способами анализа.

Таким образом, можно считать, что ответ на первый вопрос положителен. Временные параметры системы (и соответственно стоимость) таковы, что делают возможным ее промышленную эксплуатацию уже на существующих ЭВМ. Тем более ясен ответ на вопрос о том, реализуема ли данная система в принципе.

Для ответа на второй вопрос требуется рассмотреть качество машинного перевода.

Первый вариант системы АМПАР, задачей которого было упорядочить общеупотребительную лексику, строился для перевода текстов широкой тематики, поэтому исходные словари и схемы анализа были составлены на материалах английской публицистики. Ниже дается образец перевода газетного текста с английского языка на русский с распечатками информационных ячеек для контроля перевода и приписывания словам грамматической информации. Все необходимые сведения о характере и способах кодирования содержатся в предыдущих главах.

Английский текст

Athens.

The year one nine six six was the broadest and greatest resistance of the working people in Greece to the anti-working class and anti-democratic policies of the reactionary forces who were behind the coup of July, one nine six five.

Throughout one nine six six, the Greek working class maintained with honor its vanguard role in developing the struggles of the people. This is seen clearly in the extent of the strikes waged by workers and employees.

Some two, five eight five, zero zero zero workers participated in three nine five strikes, large or small, or stoppages of work in one nine six six.

Never in the post-war years was the strike movement so broad and extensive. In one nine six six, five zero various branches of workers participated in strikes.

The main demand in almost all the strikes, was the improvement of the economic position of the workers. This demand, which arises from the low level of remuneration and the high cost of living, was advanced in various ways. In some instances, through a demand for a bonus for unhealthy work, family allowances, long-service bonus, cost of living bonus, etc.

The main characteristics of the strikes were:

The mass participation of workers and the realization of unity in almost all the strikes;

The frequency of the strikes (frequent stoppages of work or two four-hour strikes in the same industry) or their continuation until victory;

The increase in the proportion of the strikes of long duration in comparison with previous years;

The effectiveness of the strikes — practically all of them ending in partial or total victory.

The strike movement was directed also against the anti-working class and anti-democratic policies of the July, one nine six five coup against democracy and thus constituted an important contribution of the workers to the struggle of the whole Greek people against the coup and for the restoration of democracy.

Certainly, however, the struggles of the working people could have developed on an even wider scale and have combined with greater stability and clarity, the economic and political demands of the working class. The demand of the whole democratic people, for instance, for the withdrawal of the fascist bill on «unions and persons» put forward in their strike meeting, by the one five zero, zero zero building workers, as well as other demands (repeal of law five zero nine against «communism», abolition of loyalty certificate, equality for the parties, including the Greek communist party etc.), ought and could have been advanced with more firmness and consistency by the workers' organizations and linked with economic claims.

Also, the struggle for trade union rights and the independence of the trade union organizations could and ought to have struggled for with greater determination.

The Greek working class enters one nine six seven armed with greater experience of organization and struggle and the effectiveness of its struggles. For one nine six seven to see an end to the anti-democratic situation in Greece and a turn towards democratic change, the working class — the most progressive and democratic force — must stand united as the worthy advance guard of the people's democratic struggles.

Русский перевод

АФИНЫ.

ГОД 1966 ЯВИЛСЯ ШИРОЧАЙШЕГО И САМОГО БОЛЬШОГО СОПРОТИВЛЕНИЯ РАБОТАЮЩЕГО НАРОДА В ГРЕЦИИ К АНТИ-РАБОЧЕМУ КЛАССУ И АНТИ-ДЕМОКРАТИЧЕСКИМ ПОЛИТИКАМ РЕАК-

ЦИОННЫХ СИЛ КТО ОТСТАЛИ ОТ ГОСУДАРСТВЕННОГО ПЕРЕВОРОТА
ИЮЛЯ, 1965.

ВСЕ ВРЕМЯ 1966, ГРЕЧЕСКОГО РАБОЧЕГО КЛАССА ПОДДЕРЖИ-
ВАЛАСЬ С ЧЕСТЬЮ ЕГО АВАНГАРДНАЯ РОЛЬ В РАЗВИТИИ БОРЬБЕ
НАРОДА. ЭТО ВИДИТСЯ ЯСНО В СТЕПЕНИ ЗАБАСТОВОК ВЕДЕННЫХ
РАБОЧИМИ И СЛУЖАЩИМИ.

ПРИБЛИЗИТЕЛЬНО 2 585,000 РАБОЧИХ УЧАСТВОВАЛИ В 395 ЗА-
БАСТОВКАХ, БОЛЬШИХ ИЛИ НЕБОЛЬШИХ, ИЛИ ПРЕКРАЩЕНИЯХ РА-
БОТЫ В 1966.

НИКОГДА НЕ В ТЕЧЕНИЕ ПОСЛЕВОЕННЫХ НЕСКОЛЬКО ЛЕТ
БЫЛ СТАЧЕЧНЫМ ДВИЖЕНИЕМ НАСТОЛЬКО ШИРОКИМ И ОБШИР-
НЫМ. В 1966, 5 С РАЗЛИЧНЫХ ОТРАСЛЯХ РАБОЧИХ УЧАСТВОВАЛ В
ЗАБАСТОВКАХ.

ГЛАВНОЕ ТРЕБОВАНИЕ В ПОЧТИ ВСЕХ ЗАБАСТОВКАХ, ЗАКЛЮ-
ЧАЛОСЬ В УЛУЧШЕНИЕ ЭКОНОМИЧЕСКОГО ПОЛОЖЕНИЯ РАБОЧИХ.
ЭТО ТРЕБОВАНИЕ, КОТОРОЕ ПРОИСТЕКАЕТ ИЗ НИЗКОГО УРОВНЯ
ВОЗНАГРАЖДЕНИЯ И ВЫСОКАЯ СТОИМОСТЬ ЖИЗНИ, БЫЛА ВЫ-
ДВИНУТА НА РАЗЛИЧНЫХ СПОСОБАХ. В НЕКОТОРЫХ СЛУЧАЕВ,
ЧЕРЕЗ ПОСРЕДСТВО ТРЕБОВАНИЯ (001) ДЛЯ НЕЗДОРОВОЙ РАБОТЫ,
СЕМЕЙНЫХ АССИГНОВАНИЙ, ДОЛГО — ОБСЛУЖИВАЕТ (002), СТОИ-
МОСТЬ ЖИЗНИ (003), И ТАК ДАЛЕЕ.

ГЛАВНЫЕ КАЧЕСТВА ЗАБАСТОВОК ЗАКЛЮЧИЛИСЬ В СЛЕДУЮ-
ЩЕМ:

МАССОВОЕ УЧАСТИЕ РАБОЧИХ И РЕАЛИЗАЦИИ ЕДИНСТВА В
ПОЧТИ ВСЕХ ЗАБАСТОВКАХ;

ЧАСТОТА ЗАБАСТОВОК (ЧАСТЫХ ПРЕКРАЩЕНИИ РАБОТЫ ИЛИ
24 — ЧАСА УДАРЕТ В ТОЙ ЖЕ ПРОМЫШЛЕННОСТИ) ИЛИ ИХ ПРО-
ДОЛЖЕНИЯ ДО ТОГО, КАК ПОБЕДЫ;

УВЕЛИЧЕНИЕ ПРОПОРЦИИ ЗАБАСТОВОК ДЛИННОЙ ПРОДОЛ-
ЖИТЕЛЬНОСТИ ПО СРАВНЕНИЮ С ПРЕДЫДУЩИМИ ГОДАМИ;

БОЕСПОСОБНОСТЬ ЗАБАСТОВОК — ПРАКТИЧЕСКИ ВСЕ ИЗ НИХ
ЗАКАНЧИВАЮЩИХ ЧАСТИЧНУЮ ИЛИ ОБЩУЮ ПОБЕДУ.

СТАЧЕЧНОЕ ДВИЖЕНИЕ БЫЛО НАПРАВЛЕНО ТАКЖЕ ПРОТИВ
АНТИ-РАБОЧЕГО КЛАССА И АНТИ-ДЕМОКРАТИЧЕСКИХ ПОЛИТИК
ПОСЛЕ, 1965 ГОСУДАРСТВЕННОГО ПЕРЕВОРОТА ПРОТИВ ДЕМОКРА-
ТИИ И ТАКИМ ОБРАЗОМ ЯВИЛОСЬ ВАЖНЫМ ВКЛАДОМ РАБОЧИХ
К БОРЬБЕ ВСЕГО ГРЕЧЕСКОГО НАРОДА ПРОТИВ ГОСУДАРСТВЕН-
НОГО ПЕРЕВОРОТА И ВОССТАНОВЛЕНИЯ ДЕМОКРАТИИ.

КОНЕЧНО, ОДНАКО БОРЬБА РАБОТАЮЩЕГО НАРОДА МОГЛА
БЫ РАЗВИЛАСЬ В ДАЖЕ БОЛЕЕ БОЛЬШОМ РАЗМЕРЕ И ОБЪЕДИНИ-
ЛА С БОЛЕЕ БОЛЬШОЙ СТАБИЛЬНОСТЬЮ И ЯСНОСТЬЮ, ЭКОНОМИ-
ЧЕСКИХ И ПОЛИТИЧЕСКИХ ТРЕБОВАНИИ РАБОЧЕГО КЛАССА, ТРЕ-
БОВАНИЕ ВСЕГО ДЕМОКРАТИЧЕСКОГО НАРОДА, НАПРИМЕР, ИБО
УХОД ФАШИСТСКОГО ДОКУМЕНТА НА «СОЮЗАХ И ЛИЦАХ» ПРЕД-
СТАВЛЯЕМЫХ НА ИХ СТАЧЕЧНОЙ ВСТРЕЧЕ, 150,000 СТРОЯЩИЕ РА-
БОЧНИЕ, ТАКЖЕ КАК ДРУГИХ ТРЕБОВАНИИ (АННУЛИРОВАНИЕ ЗА-
КОНА 509 ПРОТИВ «КОММУНИЗМА, ОТМЕНА ЛОЯЛЬНОГО СЕРТИФИ-
КАТА, РАВНОПРАВИА ДЛЯ ПАРТИИ, ВКЛЮЧИВ ГРЕЧЕСКУЮ КОМ-
МУНИСТИЧЕСКУЮ ГРУППУ И ТАК ДАЛЕЕ, СКОБКА НАЧАЛЬНАЯ.

ДОЛЖН И МОГ БЫ БЫТЬ ВЫДВИНУТ ОТ БОЛЬШЕ НАСТОИЧИВОСТИ
И ПОСЛЕДОВАТЕЛЬНОСТИ ОРГАНИЗАЦИЯМИ РАБОЧИХ И СВЯЗАН-
НЫХ С ЭКОНОМИЧЕСКИМИ ШАНСАМИ.

ТАКЖЕ, БОРЬБА ЗА ПРОФЕССИОНАЛЬНЫИ СОЮЗ ПРАВА И НЕ-
ЗАВИСИМОСТЬ ПРОФЕССИОНАЛЬНЫИ СОЮЗ ОРГАНИЗАЦИИ МОГЛА
И БЫЛА БЫ ДОЛЖНА БОРЬТЬСЯ ЗА С БОЛЕЕ БОЛЬШОИ РЕШИ-
МОСТЬЮ.

ГРЕЧЕСКИЙ РАБОЧИИ КЛАСС ВНОВЬ ПРИСОЕДИНЯЕТ 1967 ВО-
ОРУЖЕННЫЕ С БОЛЕЕ БОЛЬШИМ ОПЫТОМ ОРГАНИЗАЦИИ И
БОРЬБЫ И С СУЩЕСТВУЮЩИМ ДОВЕРИЕМ К ЕГО ОРГАНИЗАЦИОН-
НОЙ СИЛЕ И ЭФФЕКТИВНОСТИ ЕГО БОРЬБЫ. ДЛЯ 1967 УВИДЕТЬ
КОНЕЦ К АНТИ-ДЕМОКРАТИЧЕСКОЙ ОБСТАНОВКЕ В ГРЕЦИИ И ПО-
ВОРОТЕ К ДЕМОКРАТИЧЕСКОМУ ИЗМЕНЕНИЮ, РАБОЧЕМУ КЛАС-
СУ — НАИБОЛЕЕ ПРОГРЕССИВНОЙ И ДЕМОКРАТИЧЕСКОЙ СИЛЫ —
ДОЛЖН НАХОДИТЬСЯ ОБЪЕДИНЕННАЯ КАК ДОСТОИНАЯ СООТВЕТ-
СТВУЮЩАЯ ОХРАНА ДЕМОКРАТИЧЕСКОЙ БОРЬБЫ НАРОДА.

Информационные ячейки к тексту

9 24 16 00000000010000 00 03 26		
02 13 18 00001010010012 11 18 02	athens	Афины
19 24 16 00000000010000 00 03 26		
19 24 17 00000000010001 17 30 01		
12 00 02 00000000000110 00 00 01	the	
20 02 08 00000010010011 14 19 00	year	год
20 00 21 00000010010110 00 01 09	one	1
00 02 12 00000010010111 14 21 23	nine	9
00 02 20 00000010010111 17 11 18	six	6
00 02 20 00000010011101 17 11 18	six	6
27 24 30 00010111100100 19 25 08	was	явился
12 00 02 00000000001100 00 00 01	the	
25 01 12 00110110100110 14 03 24	broadest	широчайшего
19 09 02 00000000000000 00 04 30	and	и
25 08 00 00110110100110 04 25 12	greatest	самого большого
10 13 02 00000110101102 06 31 04	resistance	сопротивления
26 00 00 00000000001010 00 00 01	of	
12 00 02 00000000000110 00 00 01	the	
24 15 05 00000010100110 07 28 29	working	работающего
22 24 11 00000010101101 15 21 08	people	народа
26 01 14 00000000001011 11 20 00	in	в
02 10 15 00000101101101 25 13 27	Greece	Греции
26 01 19 00000000001010 29 06 06	to	к
12 00 02 00000000000110 00 00 01	the	
24 07 08 00000010110113 07 01 20	anti	анти
12 00 00 00000000000110 00 03 25	—	—
24 15 05 00000010110110 07 28 31	working	рабочему
22 04 21 00000010111101 15 04 30	class	классу
19 09 02 00000000000000 00 04 30	and	и

24 07 08 00001100110113 07 01 20	anti	анти
12 00 00 00000000000110 00 03 25	—	—
14 04 19 00001100110110 05 10 13	democratic	демократическими
10 02 18 00001100111101 26 23 29	policies	политикам
26 00 00 00000000001010 00 00 01	of	
12 00 02 00000000000110 00 00 01	the	
14 18 28 00001100100110 13 06 08	reactionary	реакционных
22 13 08 00001100101101 27 05 19	forces	сил
19 24 14 00000000010000 00 00 01		
19 10 16 00000000010011 14 08 03	who	кто
27 24 31 00011111100100 28 21 08	were	отстали от
26 00 18 00000000001010 00 00 01	behind	
12 00 02 00000000000110 00 00 01	the	
09 02 27 00000010101101 23 10 18	coup	государственного переворота
26 00 00 00000000001010 00 00 01	of	
00 03 19 00000010101101 15 01 01	july	июля
19 24 00 00000000100000 00 03 06	,	,
20 00 21 00000010010110 00 01 09	1	1
00 02 12 00000010010111 14 21 23	9	9
00 02 20 00000010010111 17 11 18	6	6
00 02 03 00000010011101 16 16 07	5	5
19 24 16 00000000010000 00 03 26	.	.
19 24 17 00000000010001 17 30 01	.	.
26 01 17 00000000001010 02 23 23	Throughout	все время
20 00 21 00000010010110 00 01 09	one	1
00 02 12 00000010010111 14 21 23	nine	9
00 02 20 00000010010111 17 11 18	six	6
00 02 20 00009010011101 17 11 18	six	6
19 24 00 00000000000000 00 03 06	,	,
12 00 02 00000000000110 00 00 01	the	
12 02 05 00000010100110 05 08 01	greek	греческого
24 15 05 00000010100110 07 28 31	working	рабочего
22 04 21 00000010101101 15 04 30	class	класса
26 26 26 01010010100111 01 13 24	maintain	поддерживает
26 01 26 00000000001011 11 29 00	with	с
22 15 29 00000101011101 27 18 24	honor	честью
12 00 22 00000100010113 07 15 27	its	его
00 09 04 000001000010110 04 14 00	vanguard	авангардная
10 14 30 00000100010011 27 03 01	role	роль
26 01 14 00000000001011 11 20 00	in	в
11 30 30 00000111101102 06 21 26	developing	развитии
12 00 02 00000000000110 00 00 01	the	
10 20 15 0000010110 10 25 06 16	struggles	борьбе
26 00 00 00000000001010 00 00 01	of	
12 00 02 00000000000110 00 00 01	the	
22 24 11 00000010101101 15 21 08	people	народа
19 24 16 00000000010000 00 03 26	.	.
20 00 26 10000110010012 04 09 23	This	Это

26 02 15 00110110010100 00 00 01	is	
27 31 03 00110000010100 31 14 03	seen	видится
28 00 26 00000000001000 02 08 12	clearly	ясно

Рассматривая текст в целом, можно отметить как его общую понятность, так и понятность практически каждого предложения, а также «готовность» текста к редактированию (неоднократно отмечавшаяся профессиональными редакторами). Однако рассмотрение каждого предложения показывает довольно большое количество разнообразных ошибок. Так, первое предложение следовало бы перевести: '1966 год явился годом широчайшего и самого большого сопротивления рабочего класса Греции антирабочей и антидемократической политике реакционных сил, которые стояли за переворотом в июле 1965 года'. Сразу же отметим, что мы предпочитаем вариант, наиболее близкий к тексту. Нетрудно представить себе дальнейшее улучшение текста и вкусовую правку, например: 'сил, которые вдохновляли (поддерживали) государственный переворот' и пр. Полученный машинной текст перевода этого предложения содержит следующие ошибки: неправильный перевод *working people*, ненужный перевод предлога *to* 'к', отсутствие согласования определения *anti-working class* с существительным *policies*, неправильный перевод многозначного глагола *be* в сочетании с наречием *behind*, отсутствие идиоматичного оформления даты ('в июле 1965 года'). По нормам русского языка требовалось бы также повторить слово 'год' в начале предложения.

По поводу этих ошибок следует сказать, что они в большинстве случаев устранимы в рамках алгоритма. Трудность представляет повторение слова 'год', обеспечение согласования определения с существительным в условиях разрывных составляющих. Ошибки же в переводе слов, однозначных и многозначных, в данном случае трудностей не представляют и могут быть легко учтены в контекстологическом словаре, как будет показано далее. Однако общая сложность текста, наличие разрывных составляющих и в ряде случаев необходимость существенных преобразований порядка слов затрудняют достижение полностью автоматизированного высококачественного МП таких текстов.

Второй образец содержит перевод научно-технического текста по вычислительной технике. Для осуществления этого перевода потребовалось существенно пополнить англо-русский словарь и усовершенствовать многие алгоритмы анализа. Анализ показывает, что два первых предложения вообще можно признать переведенными правильно. В третьем предложении имеется одна смысловая ошибка: 'внутренний текущий источник' вместо 'внутренний источник тока' — и две стилистических — 'также показанный есть' и 'в половине дуплексном режиме'. Здесь неправильно сработал алгоритм анализа омографа *current*. В четвертом предложении следовало бы *four wire twisted pairs* перевести 'пары из четырех скрученных проводов'. Вооб-

ще говоря, относительно последнего случая следует заметить, что здесь имеется смысловая неоднозначность, разрешить которую можно только после детального анализа текста специалистом, поэтому вряд ли в данном случае алгоритм сможет решить этот вопрос без постредактора.

Английский текст

Full duplex and 20 milliamperere operation are selected by wiring. Figure 3—2 illustrates the wiring for the various modes of operation. Also illustrated is how an internal current source can be used in half duplex mode. Four wire twisted pairs are recommended for full duplex.

Русский перевод

Полный дуплекс и 20 миллиамперная операция выбираются посредством монтажа электропроводки. Рисунок 3—2 показывает монтаж электропроводки для различных режимов операции. Также показанный есть как внутренний текущий источник может быть использован в половине дуплексном режиме.

Четыре проводные скрученные пары рекомендуются для полного дуплекса.

Нетрудно заметить, что алгоритмическое решение, из-за которого происходят ошибки, может создать неоднозначность других типов, поэтому, видимо, следует определить цену, которую придется платить за «автоматизацию» разрешения обнаруженных неоднозначных ситуаций. Может быть, усложнение алгоритмов окажется менее выгодным, чем оставление ошибок такого рода на устранение редактора.

Ответить на третий вопрос, а именно, каковы возможности системы и какие типы переводных соответствий находятся вне рамок модели, можно следующим образом. Во-первых ответ на этот вопрос дается априорно конструкцией модели и самим подходом, в русле которого она создана. Это означает, что система может не найти удовлетворительного переводного соответствия (или находить чисто случайно) тогда, когда существует семантическая неоднозначность, разрешение которой требует полной системы семантических отношений, еще не созданной в семантической составляющей. До завершения этой составляющей не учтенные в ней случаи, естественно, получать разрешения не будут. Модель в целом, однако, будет осуществлять некоторый «перевод» этих ситуаций, верный лишь приблизительно или вообще неверный, требующий активного вмешательства постредактора. По мере построения семантической составляющей предметного компонента и создания соответствующих механизмов в динамическом компоненте таких случаев будет становиться все

меньше, но окончательно (как и для человека-переводчика), они не исчезнут.

Во-вторых, опыт использования и анализ ошибок показали, что система с трудом переводит некоторые случаи, границы которых, вообще говоря, предварительно самой структурой модели не очерчиваются. Это следующие ситуации:

— наличие большого количества однородных членов, особенно в ситуациях с разрывными составляющими. В этих случаях алгоритмы анализа испытывают затруднения в связи с использованием правил пропуска;

— широкое употребление числительных в комбинациях с существительными и прилагательными. Особенности сочетания числительных и сложности словоизменения и согласования при этом существенно влияют на определение синтаксических функций слов предложения, и здесь возможны многочисленные ошибки, для исправления которых необходим более тонкий учет особенностей словоизменения и синтаксической роли числительных, чем тот, который имеется в алгоритме. Следует заметить, что этот учет требует специального исследования в направлении, которое не очень разработано в традиционной грамматике;

— большое количество собственных имен (названий фирм и пр.) в разнообразных комбинациях с существительными и прилагательными.

Система может производить целый ряд преобразований, обеспечивающих нахождение трансформационных соответствий. Так, специальный этап разрешает вопросы порядка слов в отрицательных конструкциях, предусмотрена перемена порядка слов в определительных конструкциях с причастиями, схемы перевода многозначных слов могут действовать в широких пределах по нахождению соответствий на лексическом уровне и пр. Однако естественным представляется, что система не может производить преобразований в масштабе, например, коренного изменения порядка следования слов в предложении, таких ситуаций, которые требуют экспликации различий между темой и ремой и т. п. Указанные трудности могут в значительной мере устраняться при точной ориентировке системы на тексты ограниченной специфики.

Нет необходимости особо останавливаться на трудностях, связанных с большим количеством новых слов. Словарь системы должен быть достаточно полным, в противном случае трудности возрастают пропорционально числу слов, не найденных в словаре.

ПРОБЛЕМЫ СЛОВАРЯ:

ВЫЧИСЛИТЕЛЬНАЯ ЛЕКСИКОГРАФИЯ

4.1. Основные отличия машинных словарей от обычных

Словарь в системе машинного перевода играет ведущую роль. Если грамматический анализ для МП уникален, то в части словаря накоплен большой опыт автоматической обработки текстов, который позволяет сформулировать основные черты вычислительной лексикографии.

Чтобы утверждать, что вычислительная лексикография представляет собой особую науку, а не является отдельной областью уже определившихся наук, рассмотрим сначала объект исследования вычислительной лексикографии — машинные словари. Мы выдвигаем предположение о том, что особенности структуры, состава и построения машинных словарей таковы, что они представляют собой объект, значительно отличающийся от объектов исследования лексикографии традиционной. При этом наиболее важным представляется использование машинных словарей.

Вычислительная лингвистика занимается преимущественно методами автоматического анализа и синтеза, морфологией, синтаксисом, семантикой, и лишь в незначительной степени словарями, обычно в той мере, в какой тот или иной словарь необходим для работы алгоритма. Вопросы вычислительной лексикографии как таковой редко ставились в научной литературе. Между тем многие ученые неоднократно обращали внимание на тот факт, что машинный словарь в системе автоматической обработки текстов моделирует многие важные функции человеческого интеллекта и поэтому изучение словарей имеет самостоятельное значение.

Цели и задачи, которым служили и служат словари в течение веков, чрезвычайно разнообразны. Главные функции словарей:

- регистрация в словесной форме объективных данных в свойственной данному языку и эпохе форме восприятия мира (энциклопедические и толковые словари);
- упорядочение в понятийно-словесной форме плана содержания данного языка (идеографические словари);
- нормализация словоупотребления с целью способствовать языковой коммуникации (нормативные и терминологические словари);

- систематизация лексического материала для обучения языку (учебные словари);
- перевод с одного языка на другой (переводные словари);
- прочие вспомогательные операции языкового общения (словари-справочники, номенклатуры, специальные словари).

Названные функции словарей не являются взаимоисключающими, но тесно связанными. Так, энциклопедический и толковый словарь могут быть использованы в переводе, переводной — при обучении языку и т. п. В широкой применимости словарей отражается основное свойство лексики как уровня языка, наиболее тесно связанного, с одной стороны, со сферой содержания, с другой — с коммуникативным актом.

Машинные словари используются в системах обработки языковой информации для следующих целей:

- хранение информации в ЭВМ для того, чтобы по запросу выдать код, отражающий смысловое содержание единицы хранения информации — документа, либо сам документ. В этом употреблении машинные словари участвуют в переводе с естественного языка на информационный и обратно;
- перевод с одного естественного языка на другой;
- производство других операций, предусмотренных системой.

В системе МП могут работать несколько словарей, имеющих самостоятельное значение. Строение системы зависит от принципа, положенного в ее основу, однако существуют этапы, общие для всех систем или незначительно отличающиеся для разных систем. В каждом из этапов обработки может участвовать машинный словарь.

Суммируя, можно сказать, что машинные словари выполняют одну из трех функций — хранение информации, манипуляции с ней, трансляция. Эти функции строго дифференцированы. Обычные словари также выполняют эти три функции, однако все они скрыты в норме. Норма определяет и хранение, налагая ограничения на отбираемый в словарь словарный материал, и трансляцию, ограничивая набор переводных эквивалентов, и манипуляцию, т. е. режимы пользования словарем. При этом каждый или во всяком случае один и тот же словарь может выполнять несколько или все из названных функций, т. е. здесь нет такой строгой дифференциации, которая имеет место в вычислительной лексикографии.

Словарь, составленный для человека, работает в системе человеческого мышления и коммуникации, для которого характерна сопоставительная с машинной скоростью протяженность во времени. Так, писатель, научный работник и пр. могут неоднократно и с разными целями обращаться к словарю. При этом эффективность и полезность словаря оцениваются в основном субъективно. Ближайшая объективная оценка — количество изданий. Машинный словарь работает в машинном времени для решения задач, поставленных алгоритмом обработки.

Пользователь гораздо быстрее получает то, чего он хотел от словаря; качество работы при этом проверяется сразу же и получает объективную оценку. Машинный словарь действует в системе с оперативной обратной связью, которая не есть свойство только физической организации словаря. Машинный словарь может быть включен в работу при практически нулевом словнике, и тогда пользование словарем совмещается с его построением. Такое использование словаря неизвестно традиционной лексикографии.

Будем называть составом словаря характеристику его с точки зрения составляющих его лексических единиц. Состав объединяет все то, что может быть сказано о каждой его составляющей. В рассмотрении с точки зрения состава словарь выступает как набор единиц. Под структурой словаря будем понимать способ упорядочения лексических единиц в словаре, способ организации словаря как некоторого массива данных. Структура описывает словарь как некоторое единство, входящее в систему автоматической обработки текстов и действующее в ней.

Номинальной единицей, составляющей словарь, в большинстве случаев является слово; к отдельному слову удобно «привязывать», если это необходимо, лексические единицы любой длины.

Рассмотрим словарную информацию в некоторых словарях. Большой англо-русский словарь под редакцией И. Р. Гальперина к английскому слову *amid* дает следующую информацию: «нахождение среди каких-либо предметов: в окружении, посреди, среди, между — the trees и др.». Информация к слову содержит общее в конкретных реализациях данного предлога и не перечисляет все отдельные употребления, поскольку пользователь-человек почерпнет необходимые ему сведения уже из этого описания. В словаре М. Николсон словарная статья имеет такой вид: *A — bomb, atomic bomb, atom bomb* — «используются все три формы; первичной является вторая, она предпочтительнее в письменной речи». Словарь исходит из презумпции, что основное значение слова или словосочетания пользователю известно (Nicholson, 1958). В словарях издательства Гарнье—Фламариона при отсутствии какой-либо грамматической информации, кроме указания на часть речи, и каких-либо данных о лексической сочетаемости и употреблении, сам подбор слов и их переводов в соответствующих двуязычных словарях (англо-французском — франко-английском и пр.) не механический, а подводящий читателя к правильному пониманию по возможности точным подбором эквивалентов (Vincent, 1964; Blancourt—Herselin 1964). В этих и других словарях, предназначенных для человека, существенным образом используется информация, которой человек уже располагает, и именно это обстоятельство определяет состав и характер информации к слову в словаре.

Словарная информация, таким образом, комплементарна к той, которая уже есть у человека-пользователя.

Слово для машинного словаря — совокупность знаков между двумя пробелами. Машинный словарь также имеет в качестве основной единицы чаще всего слово; словари оборотов и т. п., как правило, организованы с привязкой лексических единиц к отдельным словам. В условиях особой знаковой ситуации, в которой работает машинный словарь, он не может давать информацию, дополнительную к той, которой располагает пользователь, а должен давать всю необходимую для данного вида обработки. Для ЭВМ, у которой нет «фоновых» знаний, существует только та информация, которая в ней записана в явном виде, и эту информацию может содержать только машинный словарь (или некоторое устройство, принципиально от него не отличающееся). В этом смысле машинный словарь автономен.

Поясним примером. Машинные словари информационно-поисковых систем для обработки технических текстов должны содержать не только набор терминов, но и достаточно большое количество общеупотребительных слов. Обычные терминологические словари содержат только термины.

С точки зрения структуры обычные словари организуются по двум основным принципам: алфавитному и предметному. Организация машинных словарей подчиняется другим, более сложным критериям. Способы организации машинного словаря оцениваются в соответствии:

- с объемом памяти, занимаемым под словарь;
- с объемом памяти, занимаемой под алгоритмы определения адреса слова;
- со средним временем поиска по словарю;
- со сложностью переделки программ и таблиц словаря при его пополнении.

Суммировать отличия в составе и структуре обычных и машинных словарей можно следующим образом. С точки зрения состава словари ориентируются на одну и ту же исходную единицу — слово. Однако традиционная лексикография эксплицитно или имплицитно понимает слово как сложную лингвистическую единицу, вызывающую в сознании человека-пользователя обширную сферу коннотаций. Основная задача словаря — дать информацию, дополнительную к той, которая уже имеется у пользователя. Эта дополнительная информация может быть незначительной или вообще отсутствовать для известного слова (если словарь не сообщает потребителю ничего нового) и может быть достаточно обширной, если потребитель не знал ранее этого слова и того, что с ним связано. Разумеется, разные варианты этого положения по-разному реализуются словарями разных типов. Однако в общем можно утверждать, что в таком смысле обычный словарь комплементарен. Машинный словарь, как уже отмечено — автономен.

Таким образом, в том, что касается состава словарей, основное отличие машинных словарей от обычных заключается в качестве исходной информации. Отличия в отношении структуры связаны с переменностью состава машинных словарей. Чуждая логической и формальной структуре обычных словарей переменность состава предусматривается организацией машинного словаря — способом записи информации и поиска по словарю, программами изменения и пополнения словарного состава.

Мы не будем утверждать, что отмеченные нами закономерности имеют абсолютное теоретическое значение для лексикографии в целом. Представляется, что полная теория вычислительной лексикографии еще должна быть создана. Мы можем лишь утверждать, как и в отношении других отличий машинных словарей от обычных, что названные выше параметры существуют и они значимы с точки зрения инженерно-лингвистического подхода к проблеме.

Построение словарей включает в себя следующие моменты:

— отбор слов (исходных лексических единиц), построение словаря;

— определение состава словаря: назначение основной словарной единицы, определение объема и фактическое заполнение грамматической и другой информации к словам в зависимости от цели словаря;

— определение структуры словаря: выбор способа организации и расположения словарных единиц и фактическое построение словаря.

Рассмотрим отбор лексики в традиционной лексикографии. Исходный список слов словаря (словник) определяется типом словаря и в значительной мере лексикографическим чутьем, интуицией, вкусами и компетентностью (субъективными качествами) лексикографа, что дает основание многим исследователям определять лексикографию не столько (и не только) как науку, но и как искусство¹.

Источниками, по которым составляются словари и из которых отбирается словник, являются:

— уже существующие словари такого же типа или сходных типов;

— письменные тексты данной области знания;

— собственные наблюдения авторов над речью и собственный языковой опыт.

На отбор слов для словарей существенно влияют:

а) норма;

б) объем словаря;

в) целевое назначение словаря и другие факторы.

Определенные рекомендации для состава словаря могут быть получены из рассмотрения взаимоотношений между лексикографией и другими науками лингвистического цикла. Эти

науки могут способствовать решению проблемы «включать или не включать» данное слово в словарь, а также определения значения данного слова и состава информации к нему. Способ определения значений в словаре также в какой-то мере определяет состав словаря.

Основными признаками нормы считаются:

— употребление данных языковых фактов писателями прошлого и длительность существования этих фактов;

— соответствие языковых фактов закономерностям и тенденциям языка;

— распространенность.

Принцип нормативности и стремление отразить наиболее употребительное в языке противоречат друг другу, поскольку высокая употребительность может обладать слова и выражения, выходящие за пределы литературной речи.

Однако в строгом смысле эти критерии применены быть не могут. Употребление писателями отдельных слов может, как показывает В. П. Берков, содержать прямые ошибки. Вторым критерий — соответствие языковых фактов внутренним закономерностям языка — трудно применим вследствие возможности различного формального выражения аналогичных образований, отсутствия точного определения продуктивных моделей и трудности выявления внутренних законов применительно к лексике. Распространенность явления также не может всегда служить показателем его нормативности. В результате нормой в языке чаще всего оказывается система правил словоупотребления, «характерная для лиц, создающих нормативные пособия» (Берков 1973, 52).

Отбор лексики для машинных словарей рассмотрим в тех же аспектах, а именно:

— регистрация существующего положения с данным словом;

— «нормативность» употребления;

— постоянство состава;

— цель словаря.

Регистрация значений слов в вычислительной лексикографии имеет свои существенные особенности.

Исходным материалом для набора словарей машинных словарей служат тексты и уже имеющиеся словари данной области. Однако в отличие от традиционной лексикографии здесь приобретают большое значение формальные методы отбора лексики. Известны следующие методы построения информационных словарей, основной частью которых является словарь:

— логико-интуитивный метод, предполагающий неформальный анализ употреблений слов;

— дистрибутивно-статистический метод, основанный на формальных характеристиках текста и изучении статистического распределения элементов текста;

— метод компонентного анализа смыслов слов, частично формализуемый;

¹ Анализ разных определений лексикографии дан в работе (Берков 1973, 4).

— метод ассоциативного эксперимента, основанный на психолингвистическом подходе к определению значений слов и связей этих значений.

Таким образом, регистрация значений слов в машинных словарях отличается следующим:

— отбираемые слова и значения должны строго соответствовать конкретным информационным потребностям и выполнять нормирующую функцию;

— отбор лексики производится с помощью точных методов по формальным критериям. Логико-интуитивный метод является на нынешнем этапе развития информационных систем решающим, однако он не единственный и не исключительный.

Понятие нормы в триаде «индивидуальная речь—норма—структура» как раз противопоставляется первому члену триады по принципу «то, что понятно не только данному носителю языка, но и всем другим членам данной языковой общности», и в этом смысле отражает потребности акта коммуникации, имея аналогию с названным выше нормированием для целей информационного поиска. Однако потребности нормирования при поиске гораздо более сильны, и поэтому необходимо отмечать разницу между этими двумя понятиями нормирования.

На стадии построения машинного словаря особенно проявляется его роль как инструмента информатики. Машинный словарь характеризуется особым типом комплектования словника. В принципе машинный словарь может обеспечить максимально полный комментарий — толкование — каждому слову, но объем и характер этого комментария в первую очередь определяется задачей словаря. Дифференцированный комментарий при возможности максимально полного — вот основное отличие машинного словаря от обычного при построении и отборе лексики. Информационный словарь обязательно составляется по документам, какое-либо другое составление практически запрещено. Система должна быть строго ориентирована на запросы потребителя. Неправильный отбор исходного материала, неучет требований потребителя приводят к тому, что созданный словарь не выполняет своих функций в системе, и это сразу обнаруживается по результатам работы.

Таким образом, есть существенные отличия в построении словарей с точки зрения отбора лексики. Регистрация лексических единиц и их значений в вычислительной лексикографии происходит в рамках определенных подязыков, которые задаются совокупностями текстов. Она имеет целью обеспечить наилучшее выполнение информационным языком, частью которого является словарь, коммуникативной функции за счет отказа от других функций, присущих человеческому языку. Решающую роль в выборе лексических единиц и их значений играет целевое назначение данного машинного словаря. На первый план выходит нормирующий аспект употребления слова.

Отбор лексики в машинные словари гораздо более «норма-

тивен», чем отбор лексики в обычные словари. Критерии норм в традиционной и вычислительной лексикографии различны. В первой — норма носит характер неформализованных общих рекомендаций, следование которым затруднительно. Во второй — нормированное отличается от ненормированного по результатам выдачи, и поэтому можно утверждать, что критерий нормы реализуется в вычислительной лексикографии с помощью обратной связи.

Обратная связь касается не только отбора лексических единиц в словарь, но и отбора значений этих единиц, поскольку в информационной системе выдача зависит не только от того, какие дескрипторы использованы, но и какое значение придано этим дескрипторам в машинном словаре информационного языка. В машинном переводе показателем эффективности словаря является выдача в виде перевода текста на выходном языке.

Важной особенностью словаря является постоянство словарного состава. В традиционной лексикографии ценность словаря определяется устойчивостью употребления включенных в него слов. Отбор лексики здесь есть единая операция, предшествующая составлению словаря и его дальнейшему использованию. Новые слова образуют отдельный массив, изолированный от основного тела словаря. В отличие от этого, в машинных словарях отбор слов не прекращается перед составлением словаря, а продолжается в течение всего времени функционирования словаря. Постоянно происходит исключение устаревших или оказавшихся ненужными слов. В отличие от обычных словарей переменность состава, необходимость которой диктуется требованиями повышения эффективности или новыми условиями работы, должна быть обеспечена структурой словаря и его машинной организацией. Можно утверждать, что переменность состава и способность воспринимать изменения в ходе функционирования в системе являются принципиальными характеристиками, отличающими машинные словари от обычных в том, что касается построения словарей.

Цель и назначение машинного словаря гораздо более четко сформулированы, чем назначение словаря в традиционной лексикографии. Обычные словари обладают несомненно более широкой сферой применения в силу того, что они рассчитаны на человеческий, а не ограниченный машинный интеллект. Однако это не исключает полифункциональности машинного словаря в системе.

4.2. Предмет, методы и теория вычислительной лексикографии

Предметом вычислительной лексикографии является изучение способов построения и использования машинных словарей естественных языков.

Вычислительную лексикографию мы подразделяем на механизированную и машинную. В механизированной машинный словарь представляет собой словарь обычный, только в иной физической форме: к механизированной лексикографии можно отнести все виды использования механических и автоматических средств для помощи в обычной лексикографической работе. В машинной лексикографии машинный словарь используется как орудие автоматической обработки текстов с определенными целями, он есть часть системы такой обработки.

Механизированная лексикография в использовании словарей не имеет принципиальных отличий от лексикографии обычной. Машинная лексикография действует в пределах науки информатики и подчиняется закономерностям последней.

Главной смысловой особенностью информатического текста является его справочный характер. Информатический текст не должен содержать ошибок в калькуляции сведений или в точности извлечения сведений из исходного документа. Но информатический текст не может рассматривать содержание исходного документа с точки зрения его отношения к действительности, оценивать содержание исходного документа. Поэтому в информатический текст не входят все фигуры стиля документа, его риторическая установка.

Построением информатических текстов в указанном смысле занимаются информационно-поисковые и информационно-логические системы. В той мере, в какой естественный текст, подвергнутый машинному переводу, должен удовлетворять требованиям информационного, соответствующие задачи должны решаться и алгоритмом МП. Это делает МП задачей гораздо более сложной, чем прочие лингвистические задачи прикладного характера.

Создание адекватной для МП грамматики и словаря безусловно равноценно созданию цельной теории языка или его модели, в случае, однако, если это создание охватывает не фрагмент языка, а достаточно широкую языковую область и позволяет осуществить обработку большого массива текстов. «В каком-то смысле машинная грамматика и машинный словарь являются методико-лингвистическими описаниями языка, т. е. описаниями, совмещающими методический и лингвистический аспекты. Составление алгоритмов машинного перевода не менее теоретично, чем составление академической грамматики или словаря, в неявном виде учитывающих интуицию носителей языка, их способность читать такие грамматики и словари «между строк», «домысливать» недосказанное или высказанное туманно» (Денисов 1974, 169).

В вычислительной лексикографии можно отметить общие методы, объединяющие ее с лексикологией и лексикографией, а также с другими науками лингвистического цикла, и частные методы, обусловленные спецификой лингвистических вычисле-

ний². Из общих методов наибольшее значение имеет использование языка-эталона для описания значений слов. Практическое применение этого метода имеет место в теории детерминант. Среди частных методов важно использование билингв. Кроме того, последнее время больше внимания стало уделяться вопросам математического моделирования словаря.

Так, в работе А. Эттингера, посвященной машинному словарю для МП и написанной с математической и технической точек зрения, нет теории словаря (Oettinger 1960). Наиболее полная теория словаря, по-видимому, дается в работе Д. Санкова (Sankoff 1971), в которой развивается идеализированное представление о словаре Д. Болинджера (Bolinger 1965). По Д. Санкову, существуют два эмпирически мотивированных подхода к формализации понятия словаря. Первый исходит из признания того, что существуют некоторые простые неразложимые элементы, называемые маркерами, булева комбинация которых описывает любое значение слова. Другой подход основан на иерархическом упорядочении общих и специфических терминов в таксономиях. Чтобы сравнить эти два подхода, нужно формализовать их так, чтобы они имели дело с едиными исходными данными.

Метод билингв имеет широкое применение в виде использования параллельных текстов для составления словарей и словников. В качестве примера можно рассмотреть составление семантических частотных словарей. Состав словника при этом во многом зависит от выбранной методики выделения единиц перевода в исходном и переводящем тексте. При составлении англо-русского семантического частотного словаря по радиоэлектронике единица перевода была определена как минимальный отрезок исходного английского сегмента (слово или словосочетание), для которого в соответствующем русском сегменте нет лексических единиц, передающих лексическое значение составных частей выделенной единицы перевода, если она является словосочетанием. Это рабочее определение исходит из понятия сегмента (Марчук, Тихонова, Убин 1975).

Математическая модель словаря исходит из идеализированных представлений о словаре как некоторой структуре, задающей отношения порядка на множестве слов. Сравнительно с другими объектами математической и вычислительной лингвистики словарь лишь недавно стал объектом математического моделирования. Природа множеств, на которых словарь задает

² «Специфика языка состоит в том, что любое его описание относительно и верно лишь постольку, поскольку оно удовлетворяет той цели, ради которой было предпринято... Любое описание — это тоже язык, который надо понять, т. е. как бы снова описать на семантическом языке. Следовательно, если существуют два описания одного и того же аспекта языка, но одно из них непонятно данному потребителю, а другое — понятно, то он сможет воспользоваться (и тем самым убедиться в истинности) только понятного для него описания» (Денисов 1974, 168).

отношения порядка, может быть различной. Словарь может рассматриваться и как структура, упорядочивающая буквенные комбинации, и как структура, строящая отношения между элементами смысла, выраженными словами и словосочетаниями.

Оценивая технологические возможности математической теории словаря, следует признать, что здесь еще не получено таких результатов, пользуясь которыми можно было бы оценивать проектируемые или имеющиеся словари по существенным для них содержательным лингвистическим параметрам.

В математической теории тезауруса можно выделить два направления. В первом из них формализуется одна сторона функционирования тезауруса, а именно классификационная. Тезаурус рассматривается как система, реализующая иерархическую классификацию фактов (Dagobert 1972). Исходной позицией при этом служит требование о минимальной сложности индексирования. Для каждого элемента смысла тезаурус дает ряд последовательно уточняемых координат (Овчинников 1977). Другое направление отражает смысловую структуру тезауруса. Всякое отношение типа сходства может быть задано набором канонических признаков. Эти признаки естественно отождествлять с «элементарными смыслами». Таким образом, описание группы родственных по смыслу слов (семьи) равносильно заданию множества пар слов и множества элементарных смыслов. Вопрос о том, между какими парами имеет место сходство, выходит за рамки математической модели и решается информантом. Важно то, что при таком подходе система признаков и тем самым «элементарных смысловых единиц» определяется выбором подмножества допустимых пар и установлением сходства на парах. Затем уже эти элементы смысла могут получать содержательные наименования (Шрейдер 1971).

Этот подход противостоит другому, при котором заранее, априорно, выделяются элементарные смыслы и уже в терминах этих элементарных смыслов описываются свойства слов. Априорное выделение таких семантических параметров неудобно для описания имеющегося словарного материала. П. Н. Денисов показывает, что для толкования 2996 слов с позиций априорного моделирования требуется 7262 других слова (Денисов 1974). С самых общих позиций метод предварительного жесткого деления семантического поля на составляющие, а в этом и заключается идея комбинаторного моделирования, противоречит принципу скользящего и переменного описания семантики, меняющейся диалектически в подъязыках и аспектах общения (см. также Котелова 1975).

Теория вычислительной лексикографии строится на тех же основных положениях, что и теория лексикографии в целом как языковедческой науки: абстрактные теоретические объекты лексикографии суть понятия инварианта, образца или типа словаря, максимальной или минимальной лексической системы, универсальной структуры словарной статьи, дефиниции, дескрипции

слова и т. п. Однако эта теория содержит также положения, специфические для лексикографии вычислительной как дисциплины, связанной с кибернетикой и информатикой.

Специфические положения, лежащие в основе теории вычислительной лексикографии, суть следующие:

— соотношение словаря и алгоритма автоматической обработки текстов;

— типология машинных словарей в машинной и механизированной лексикографии с противопоставлением общих для них и специальных для каждой типов словарей;

— представление о языке как многоуровневой иерархической системе, основным назначением которой является коммуникация. Эта система допускает постепенную расшифровку основных черт, релевантных коммуникации. Такое представление лежит в основе использования приближенных методов вычислений применительно к проблемам лексикографии и составления машинных словарей.

Словарь является статическим собранием языковых фактов; с этими фактами оперирует механизм — алгоритм решения конкретной задачи. Разделение описания и алгоритма влияет также на структуру машинного словаря. Машинный словарь может иметь разный объем информации к лексическим единицам. П. Гарвин, автор системы МП «Фулькрум», делит системы МП в зависимости от соотношения словаря, алгоритма и грамматического описания на двучленные и трехчленные³. Грамматическая информация может располагаться либо в специальных таблицах, либо в словарных статьях. Примером первого решения является система МП с русского языка на французский Гренобльского центра (модель синтаксического анализа M—2 (Vauquois 1968); примером приписывания информации словам может служить система «Мультистор» (Глазерсфельд, 1971), а также системы МП, разрабатываемые с установкой на множественный синтез (Шаляпина 1974 и др.). Объем словарной статьи, в которую вносятся сведения о синтаксических особенностях, модели управления, лексической сочетаемости и пр., чрезвычайно возрастает.

Типологию словарей механизированной лексикографии целесообразно строить по видам лексикографической работы, в результате которой получают те или иные словари на ЭВМ. Обработка текстов с помощью ЭВМ происходит в тесном взаимодействии человека с машиной.

Машинные словари можно классифицировать по различным признакам. Самой общей классификацией является классификация по двум основаниям: — по характеру лексических единиц, включенных в словарь, и по принципу упорядочения в нем лек-

³ Двучленная система состоит из машинного словаря и алгоритма, трехчленная — из словаря, алгоритма и отдельной таблицы грамматических правил (Гарвин 1971, 26).

сических единиц, т. е. по способу организации словаря. По характеру лексических единиц мы делим словари на: 1) словари основ; 2) словари словоформ; 3) словари оборотов. По способу организации словаря машинные словари подразделяются на: 1) частотные, 2) алфавитные (прямые и обратные), 3) словари-тезаурусы, 4) словари-конкордансы, 5) специальные словари, к которым можно отнести, например, автоматический контекстологический словарь для перевода многозначных слов. Машинная и механизированная лексикография отличаются лишь некоторыми разновидностями из перечисленных выше словарей. Так, в рамках той и другой используются частотные словари, алфавитные, конкордансы и пр. Разница, однако, в принципиальной конечной цели, которой служит словарь. В механизированной лексикографии словарь служит конечной целью исследования. С помощью ЭВМ получают некоторые данные о лексическом составе текстов или лексических характеристиках языка. В машинной лексикографии машинный словарь используется как орудие автоматической обработки текстов, с его помощью добываются некоторые новые данные.

4.3. К проблеме лексической неоднозначности

Не имея возможности подробно рассматривать специфику вычислительной лексикографии и ограничившись определением ее общих границ и тезисом о самом ее существовании, перейдем к ее аспектам, наиболее интересным и необходимым для поставленной задачи — моделирования перевода. Наиболее важной проблемой здесь является разрешение лексической неоднозначности (многозначности).

Теория контекстной детерминации требует прежде всего уточнения понятия лексической неоднозначности. Она (неоднозначность) отделяется от грамматической и лексической неоднозначности (омографии). Если слово в заданном корпусе текстов имеет один перевод, оно однозначно, если более одного — многозначно. При этом значение приравнивается переводу, что можно считать допустимым в прикладных целях.

В данной постановке задачи о многозначности к последней приравниваются также случаи лексической омонимии. Границы между многозначностью и лексической омонимией нерелевантны переводу, причем не только машинному; переводчик вряд ли задумывается над тем, имеет ли он дело с лексическим омонимом или многозначным словом.

Разрешить лексическую многозначность по тексту — это значит найти и выбрать один перевод многозначного слова на естественный язык (при МП) и на информационный (при информационном поиске).

Проблема лексической неоднозначности возникла в МП одновременно с самой проблемой МП. Именно лексической неоднозначностью мотивировал И. Бар-Хиллел концепцию о невоз-

можности высококачественного и полностью автоматизированного МП (Бар—Хиллел 1962). Первые исследования многозначности в связи с МП в основном посвящены описанию фактов лексической многозначности и ее классификации, связи с семантикой, в формализации которой уже тогда усматривались серьезные трудности.

С самого возникновения проблемы ни у кого не возникало сомнений, что разрешение многозначности должно осуществляться с помощью контекста. Однако в понимании контекста, который должен служить этой цели, не было единообразия. Мы можем выделить два основных подхода к алгоритмическому разрешению многозначности по контексту.

Начало подходу, который можно назвать синтаксическим, было положено В. Ингве (Ингве 1957), который рекомендовал искать разрешения многозначности в пересечении синтаксических классов совместно встречающихся слов. Такое предложение было выдвинуто потому, что чисто дистрибутивные методы в их классической форме, например (Harris 1970), были непригодны из-за присутствия в них в неявном виде обращения к смыслу. Родоначальники синтаксического метода осознавали также, что в разрешении лексической многозначности еще нельзя будет опираться на сведения о полной синтаксической структуре предложения, поэтому методика предусматривала использование как бы отдельных частей (кусков) синтаксической структуры. Кроме того, использовались сведения о семантических характеристиках отдельных элементов предложения. Приведем в качестве примера две методики такого рода.

В лаборатории информационного поиска компании Локхид (США) разрабатывается система автоматического индексирования, составной частью которой является специальный словарь для разрешения одноязычной многозначности английских слов. Например, слово *elevation* означает либо 'высоту' — 'дом высотой в 100 футов', либо акт продвижения вверх — 'газеты объявили о его повышении'. *The building has an elevation of 100 feet. The newspapers announced his elevation to the peerage.* Каждое многозначное слово как единица словаря (primary) имеет набор так называемых типов контекста (patterns), каждому из которых соответствует одно определенное значение многозначного слова. В табл. 11 изображен фрагмент словаря для двух многозначных слов. Тип контекста для слова *elevation* изображен в табл. 12.

Методика такого рода основывается на предположении о том, что каждому семантическому значению соответствует одна синтаксическая структура⁴.

В алгоритме МП университета штата Вашингтон применена несколько иная, но сходная методика для русских слов.

⁴ По состоянию на 1970 год работа была сделана на 70%, словарь включал примерно 8000 многозначных слов и 40 000 типов контекста. Не сообщается, каким образом обеспечивается необходимая исходная информация и какова эффективность на новом материале (Robinson 1970).

следовательно проверяется вхождение данного омонима в типовые конструкции. Если омоним входит в конструкцию, омонимия разрешена, если нет, то поиски разрешающего правила продолжают. Здесь используются списки слов и отдельные слова. Эффективность алгоритма — 86% при 8% ошибок и 6% неразличений (Чернышев 1970). Примерно такой же метод предложен в (Трибис 1973).

Определенные возможности для разрешения многозначности представляет сужение тематической области. Так, в (Madhu, Lytle 1965) использовалась вероятность, с которой данное слово встречается в различных областях науки и техники. Каждому значению в каждой области соответствует свой коэффициент. Исследование проведено на материале в 21 русское предложение. Недостатком метода является неточное определение границ областей. Постепенное сужение поля функционирования и выбор значения из все более сужающегося круга осуществлены в алгоритме (Зубов 1971). Рассматривается макроконтекст, в котором число возможных значений даже для очень многозначных слов (глагол *faire*) сокращается в два и более раза. Затем используются вероятности различных значений слова, при этом устойчивые сочетания задаются списком.

В информационном поиске различие лексической многозначности в основном происходит с помощью словаря при индексировании. Однако текстовое различение многозначности дает большое количество шума и, по-видимому, проблема алгоритмического разрешения скоро возникнет и здесь (Певзнер 1973).

Подход, который можно было бы назвать лексическим, характеризуется следующим порядком этапов:

- определение границ применимости метода и границ исследуемой подсистемы;
- установление объема контекста;
- выработка инвентаря различительных признаков контекста;
- формализация лексической сочетаемости;
- учет дистрибутивных и статистических признаков языковых элементов;
- выработка собственных синтаксических и семантических параметров и приемов, дополняющих лексический уровень;
- построение алгоритма и проверка его на новом материале.

Формальное определение лексической многозначности в вычислительной лексикографии позволяет поставить актуальный вопрос о ее распространенности в словаре и текстах. Статистические характеристики многозначности позволяют оценить возможность и необходимость применения тех или иных методов разрешения многозначности и «цену» за их применение.

Рассмотрим распространенность многозначности в словаре. Приравнивая многозначность «многопереводности» и считая слово многозначным, если оно встретилось в разных рубриках,

Таблица 15

Части речи	Однозначных слов	Процент	Многозначных слов	Процент	Всего	Коэффициент многозначности
Существительные	6812	71,3	1561	40,5	8373	0,18
Глаголы	1015	10,5	1254	32,3	2269	0,55
Прилагательные	1101	11,5	860	22,3	1961	0,44
Наречия	603	6,7	132	3,4	735	0,18
Союзы	0	0	8	0,2	8	1,0
Предлоги	0	0	53	1,3	53	1,0
Итого	9531	100	3868	100	13 399	

можно определить наличие многозначности в словаре типа тезауруса; так, для тезауруса П. Роже примерно половина всех слов многозначна (Вишнякова 1973). Для разных частей речи соотношение многозначных и однозначных слов различно. В табл. 15 приведено распределение многозначных и однозначных слов в словаре системы АМΠΑР⁵.

Из этой таблицы видно, что однозначные слова преобладают в классе существительных и наречий, многозначные — в классе глаголов. Прилагательных примерно поровну. Общее соотношение многозначных и однозначных слов в словаре АМПАР примерно такое же, как в словаре Роже. Распределение слов лексико-грамматических классов и омографов в словаре АМПАР⁶, показанное на рис. 13, характерно тем, что глаголы, наречия и предлоги, как строевые компоненты текста, присутствуют примерно в равных долях в каждой сотне слов частотного словаря, упорядоченного по убывающей частоте. Для существительных и прилагательных эта доля существенно меняется. Это и означает, что именно за счет этих классов происходит прирост словаря, что соответствует данным в (Алексеев 1964).

Распределение многозначных слов в частотном словаре всегда привлекало внимание лингвистов. Общепринятым считается, что самые частые слова наиболее многозначны и обратно — самые многозначные слова наиболее часты. Однако линейной зависимости, как это вытекает из закона Ципфа, здесь нет, и в этом можно убедиться на примере графика зависимости числа значений слова от его ранга в частотном словаре, построенного для словаря АМПАР. Из графика (рис. 14) следует, что в пределах первых ста рангов число значений слов крайне невелико. Здесь нужно учесть, что в словарь АМПАР, как и в многие другие машинные словари, знаки препинания входят как самостоя-

⁵ Только по словарю общеупотребительной лексики.

⁶ Частотный словарь АМПАР описан в (Марчук 1973б).

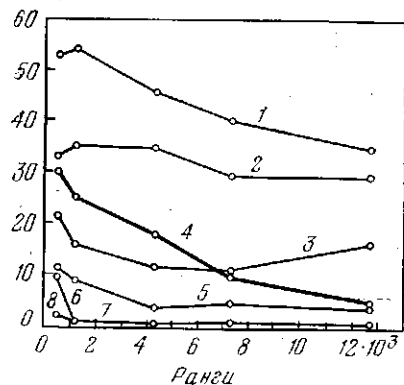


Рис. 13. Распределение лексико-грамматических классов и омографов в частотном словаре АМΠΑР
 1 — существительные; 2 — глаголы; 3 — прилагательные; 4 — омографы; 5 — наречия; 6 — предлоги; 7 — союзы; 8 — знаки препинания.
 Примечание: расчет произведен по сотням слов частотного словаря.

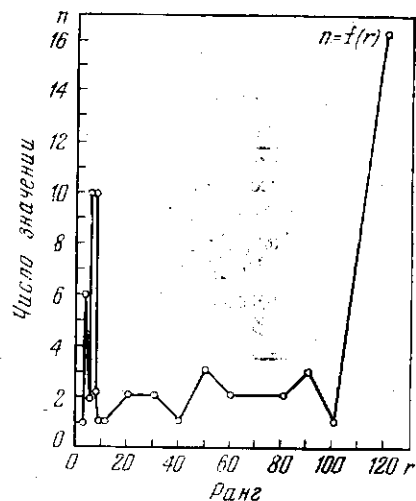


Рис. 14. Зависимость числа лексических значений слова от его места в частотном словаре

тельные слова. В частотном словаре ранги от 3 до 7 занимают предлоги, у каждого из которых есть самостоятельное лексическое значение, но число таких значений невелико. В диапазоне рангов с 20 по 100 находятся модальные и вспомогательные глаголы (здесь кривая построена по точкам, кратным 10), каждый из которых имеет небольшое число лексических значений. Некоторые прилагательные в этом диапазоне однозначны. В целом многозначные слова в начале частотного списка чередуются с однозначными и зависимость выражается не прямой, а кривой линией, имеющей пересечения с осью однозначности.

Оценка распределения лексической многозначности в тексте возможна только при наличии словаря, в котором многозначные слова отделены от однозначных, и большого количества текстов, в которых были бы отмечены словоупотребления многозначных и однозначных слов. Последнее затруднительно при ручной обработке и удобно с использованием ЭВМ.

Запрограммированный и введенный в машину словарь АМΠΑР позволяет получить исчерпывающие сведения о распределении лексической многозначности по тексту. Оно определяется по промежуточному тексту. Есть некоторые особенности алгоритма, которые нужно учесть для правильного распределения слов по категориям «однозначность/многозначность», а именно:

Таблица 16

Номер текста	1	2	3	4	5	6
Объем текста в словоупотреблениях	626	453	296	634	230	354
Число употреблений многозначных слов	287	196	118	255	101	161
Процент употребления многозначных слов среди всех словоупотреблений	46	43	50	40	44	46

- знаки препинания выступают как однозначные слова;
- перед союзными словами восстанавливается служебная запятая;
- многозначные слова учитываются всеми своими отдельными словоформами: так, *may have received* — три многозначных слова.

Было обследовано шесть переведенных машиной текстов. Результаты сведены в табл. 16. Из нее следует, что многозначные слова составляют в текстах примерно 40—50% всех словоупотреблений.

Если принять, что между числом многозначных слов в словаре и числом словоупотреблений многозначных слов в текстах существует прямая пропорциональная зависимость, то это означает, что некоторый абстрактный, состоящий только из употреблений многозначных слов «полностью многозначный текст» может образоваться при объеме многозначного словаря примерно в половину от общего количества слов словника.

5.1. Принципы построения

Рассмотрим множество многозначных слов в исходном языке. Полагаем, что выбор эквивалента из списка возможных переводов каждого слова осуществляется не случайно, а детерминируется употреблением переводимого слова в той или иной конкретной ситуации, т. е. особенностями того отрезка текста, в котором он встречается. Некоторые особенности могут быть связаны с самим переводимым словом, — например, если на перевод слова влияет то, в какой грамматической форме оно стоит в тексте. Другие особенности могут быть связаны с контекстом, с окружающими словами.

Иначе, каждому из значений, фиксированных посредством отдельного слова или словосочетания эталонного языка, в исходном языке должна соответствовать некоторая форма, диагностирующая или детерминирующая его; опираясь на форму, детерминирующую в каждом конкретном случае то или иное значение многозначного слова, переводчик и выбирает каждый раз определенное значение в выходном языке. Таким образом, переводы помогают нам эксплицировать то различие в значениях, которое фиксирует английский язык, пользуясь в каждой конкретной ситуации строго определенным набором языковых единиц.

Очевидно, что можно было бы поручить задачу снятия многозначности ЭВМ, если бы в построенном нами словаре фиксировался набор форм, разводящих значения данного многозначного слова, и набор переводов, однозначно соотносенных с этими формами. Главная трудность на этом пути заключается в том, что такой словарь нельзя построить априори.

В качестве рабочего метода для построения такого словаря примем метод постепенного улучшения исходного конкорданса. Суть метода состоит в следующем.

За исходный примем словарь, составленный на основании биллингов таким образом, что в нем представлена дистрибуция слов, частота их употребления и все имеющиеся в данном объеме текстов значения каждого из слов, образующие словарные статьи. Назовем такой словарь конкордансом. На основании этого конкорданса построим алгоритмы разведения значений каждого многозначного слова. Осуществим машинную проверку этих алгоритмов на текстах, не вошедших в исходные биллинги. Исследуя полученные результаты, будем совершенствовать конкорданс в двух отношениях: 1) будем увеличивать количе-

ство переводов многозначных слов (если полисемия данного многозначного слова в конкордансе отражена недостаточно) и при этом будем фиксировать соответствующие им детерминирующие формы; 2) будем дополнять уже имеющиеся в словаре детерминирующие формы, если их недостаточно для распознавания тех значений, которые зафиксированы в словаре.

Очевидно, что в результате этой процедуры мы получим новый конкорданс. Этот конкорданс будет лучше, чем предыдущий — с его помощью, несомненно, будет вскрываться больше таких значений многозначных слов, которые релевантны для переводчика. Однако ряд значений останется «за бортом». Некоторые из них всплывут в ходе исследования результатов новой машинной реализации, и мы сможем включить их в следующий конкорданс. Но некоторые снова попадут в неформализованный остаток и т. д., пока не будет составлен идеальный конкорданс, дающий возможность решать задачу о распознавании всех омонимов и значений, которые может распознать человек с помощью ЭВМ.

Этот путь достаточно трудоемкий, но, по-видимому, более легкого пути к решению такой задачи нет.

Ясно, что таким путем мы не сможем охватить все без исключения слова исходного конкорданса: так, без разрешения останутся случаи авторских ошибок, неологизмы; однако процент неразрешенных ситуаций может оказаться достаточно малым, чтобы им пренебречь.

На первый взгляд кажется, что в качестве исходного конкорданса можно взять обычный переводной двуязычный словарь. Однако такой словарь не может служить основой построения специального машинного словаря по следующим причинам:

а) переводной словарь дает не переводы, а ряды синонимов, поясняя значение; лишь в частных случаях значение может совпасть с переводом. Машинный словарь должен во всех случаях давать переводы;

б) отдельные значения слов даются в переводном словаре вне сочетаемости; указания на дистрибуцию и сочетаемость случайны, по ним не могут быть построены дистрибутивные модели. Восстановление дистрибуции каждого лексико-семантического варианта по данным такого словаря представляется невыполнимой задачей;

в) переводной словарь объединяет употребление слов в самых различных стилях и жанрах; между тем, в современных прикладных задачах ограничиваются, как правило, одной областью языка (подъязыком);

г) в переводных словарях нет статистических данных, которые позволили бы выбрать главное значение, отделить его от второстепенного, определить наиболее частое значение и пр.

Ни один из рассмотренных ранее «легких» путей разрешения лексической многозначности не ведет к результатам, нужным, чтобы построить автоматический словарь для снятия полисемии

Таблица 17. Коэффициент многозначности

Лексико-грамматический класс	Коэффициент многозначности
Существительные	0,21
Прилагательные	0,45
Глаголы	0,51
Наречия	0,21
Предлоги и союзы	1,0

многозначных слов. В перспективе остаются только конкордансы, составляемые на основании билингв, уточняемые посредством обычных двуязычных словарей и конкордансов и проверяемые с помощью ЭВМ. Соответственно возрастает значение методики составления автоматических словарей такого типа, и, в частности, аппроксимационной методики, предложенной выше.

Опишем полностью начало аппроксимации.

Исходный массив алгоритмов перевода составлялся на материале английских газетных информационных и публицистических текстов на английском языке общим объемом около 700 000 словоупотреблений. Многозначные слова распределены по лексико-грамматическим классам.

Больше всего в словаре многозначных существительных, затем глаголов и прилагательных. Многозначные слова составляют около 30% всех слов входного словаря. Внутри отдельных лексико-грамматических классов число многозначных слов различно. Коэффициент многозначности — отношение числа многозначных слов данного лексико-грамматического класса к общему числу слов этого класса — имеет следующие значения (табл. 17).

Полная многозначность предлогов и союзов объясняется высокой степенью их грамматикализации. Существительных больше, чем прилагательных, однако многозначных слов среди них меньше, чем среди прилагательных. Наиболее многозначны глаголы.

Перевод с исследуемого языка на эталонный производит переводчик, знающий английский и русский языки (последний как родной) и предмет, о котором идет речь в текстах. Переводчик получает конкорданс; пользуясь своим знанием языка и предмета и в случае необходимости консультируясь с другими переводчиками и информантами, а также обращаясь к толковым и переводным словарям, справочным пособиям (словарям сокращений, географических названий, терминологическим словарям и пр.) переводчик дает перевод слова в каждом конкретном его употреблении, зафиксированном в конкордансе, и, если необходимо, указывает на значения слова, почему-либо там не отраженные.

После того как произведен перевод, составитель алгоритма выполняет следующие операции:

— отделяет однозначные слова от многозначных и в дальнейшем рассматривает только последние;

— перечисляет в выходном языке все слова, которые в заданном конкордансе могут служить при переводе лексическими эквивалентами исследуемого¹;

— устанавливает, какие факторы влияют на выбор лексического эквивалента в каждом конкретном случае;

— определяет, как эти факторы взаимосвязаны между собой;

— указывает типы дистрибуции, по которым устанавливается однозначное соответствие между исследуемым и переводящим словом;

— представляет процедуру идентификации типа дистрибуции и выбора лексического эквивалента для перевода в виде алгоритма перевода данного многозначного слова. Разрешение лексической многозначности происходит в дальнейшем как работа на тексте составленного таким образом алгоритма, разрешающего многозначность данного слова. Совокупность алгоритмов и образует контекстологический словарь.

Из текста до этапа перевода получается следующая информация:

1. Морфологические признаки слов. Сами эти признаки (аффиксы и окончания) и информация о них являются результатом морфологического анализа, осуществляемого на этапе поиска входной словоформы по словарю.

2. Информация о принадлежности слова к лексико-грамматическому классу. Однозначность этой информации обеспечивается решением проблем омографии.

3. Сведения о синтаксических функциях слов в предложении.

Лексические значения многозначных слов детерминируются информацией о следующих явлениях: а) сочетаемости многозначного слова в исследуемом языке со словами определенных лексико-грамматических классов (подклассов); б) сочетаемости слова со словами определенных списков или отдельными словами; в) участии слова в синтаксических конструкциях или СФ самого слова; г) наличии у слова морфологических признаков или морфем; или комбинации из перечисленных явлений, каждая из которых однозначно определяет один перевод.

Назовем детерминантой лексико-грамматический класс или подкласс, список слов или отдельное слово, синтаксический, морфологический признак или морфему, которые по отдельности или в некоторой совокупности однозначно определяют одно значение многозначного слова как требующее отдельного лексического эквивалента в выходном языке.

Выделим следующие виды детерминант.

¹ Разрешается в порядке исключения добавлять только весьма распространенные переводы, почему-либо отсутствующие.

Лексико-грамматическая детерминанта (ЛГ-детерминанта) представляет собой лексико-грамматический класс или подкласс слов, характеризующихся общим семантическим содержанием и общими особенностями дистрибуции.

Лексическая детерминанта (Л-детерминанта) — отдельное слово или список слов, не образующих класс или подкласс. Л-детерминанта характеризует индивидуальную сочетаемость слова.

Для ЛГ-детерминанты и Л-детерминанты определен признак «направление». Детерминанты, находящиеся справа от исследуемого слова, называются правыми, слева — левыми.

Синтаксическая детерминанта (С-детерминанта) представляет собой синтаксическую функцию слова в предложении. В качестве С-детерминанты может выступать либо СФ самого слова, либо СФ ЛГ- или Л-детерминанты, поэтому для С-детерминанты, как и для двух последующих, определен признак «принадлежность».

Морфемная детерминанта (М-детерминанта) — морфема, позволяющая различать лексическое значение. Чаще всего в виде М-детерминанты выступают суффиксы и окончания.

Морфологическая детерминанта (МЛ-детерминанта) представляет собой грамматическую категорию в эталонном языке, релевантную для выбора лексического эквивалента в этом языке. Например, если глаголы, образующие видовую пару в русском языке, записаны в выходном словаре этого языка отдельно, то грамматическая категория «вид», определенная до перевода, позволяет сделать правильный выбор этого эквивалента.

Пять перечисленных выше детерминант называются детерминантами первого порядка, или простыми детерминантами. Приведем примеры детерминант первого порядка.

ЛГ-детерминанта. Английское существительное *growth* при наличии справа от него географического названия переводится 'развитие'². В других контекстах переводы другие. Географическое название является правой ЛГ-детерминантой существительного *growth*.

Л-детерминанта. Глагол *to inflate* в сочетании с последующим существительным *price* переводится 'вздутие', во всех других контекстах — 'произвести инфляцию'. Существительное *price* является правой лексической детерминантой глагола *to inflate* в этом переводе.

М-детерминанта. Существительное *consequence* с окончанием *s* переводится 'последствия', в остальных случаях переводы другие. Окончание множественного числа *s* является М-детерминантой для существительного *consequence* в этом переводе.

МЛ-детерминанта. Глагол *to guide* с признаком «пассив» переводится 'руководствоваться', без такого признака — 'руко-

водить'. Признак «пассив» является МЛ-детерминантой глагола *to guide* в указанном переводе.

ЛГ- и Л-детерминанты будем называть «лексически-выраженными», а три остальные — «грамматически-выраженными». Свойство эквивалентности заключается в том, что две разные детерминанты дают один и тот же результат. Например, М-детерминанта *s* для существительного *consequence* эквивалентна МЛ-детерминанте «число» для этого же перевода. Свойство перемещения характерно для лексически-выраженных детерминант. По этому свойству, например, правая лексическая детерминанта активных конструкций становится для пассивных конструкций левой.

Обнаружение детерминанты первого порядка соответствует определенной стандартной лингвистической операции, которая в алгоритме выполняется стандартным оператором.

Детерминанты второго порядка представляют собой конъюнкции детерминант первого порядка. Установлено, что число детерминант в таких конъюнкциях редко превышает четыре. Максимальное число детерминант требуется для нахождения всех компонентов идиоматических словосочетаний, не вошедших в словарь оборотов. Наиболее часты детерминанты второго порядка, составленные из двух простых детерминант, реже встречаются детерминанты из трех. Среди детерминант второго порядка будем различать составные детерминанты, являющиеся объединениями только лексически-выраженных детерминант, и сложные, компонентами которых могут быть любые простые детерминанты. Чаще всего сложные детерминанты состоят из двух компонентов. У сложных детерминант первого типа грамматически выраженные детерминанты, входящие в объединение, записаны исследуемому слову. У сложных детерминант второго типа эти детерминанты относятся к лексически-выраженным детерминантам. Приведем примеры детерминант второго порядка.

Составная детерминанта: глагол *to fall* в значении 'солидаризоваться' имеет составную детерминанту *in+line* (Л-детерминанта + Л-детерминанта).

Сложная детерминанта первого типа: перевод существительного *effect* 'вещи' детерминируется морфологической детерминантой *s* и левой лексической детерминантой в виде списка (М-детерминанта + Л-детерминанта).

Сложная детерминанта второго типа: прилагательное *evil* в переводе 'злонамеренный' детерминируется последующим глаголом *mind* и окончанием *-ed* этого глагола (Л-детерминанта + М-детерминанта).

Существуют и другие виды сложного детерминирования. Многие из них допускают замены или имеют эквивалентное детерминирование. Первыми во избежание ложного детерминирования должны следовать детерминанты второго порядка (принцип *longest match*); однако, строго говоря, это обязательно

² Приводятся примеры из исходного конкорданса (Контекстологический словарь 1976).

лишь в том случае, если имеются омонимичные детерминанты к детерминантам второго порядка; если таковых нет, то порядок следования детерминант произвольный.

При поиске детерминант в тексте действуют правила пропуска, описанные ранее, а также приведенные во введении к контекстологическому словарю.

5.2. Состав и структура словаря

Алгоритм (схема) перевода многозначного слова состоит из последовательности детерминант и переводов, завершаемой так называемым общим выходом алгоритма, который отражает недетерминируемый контекстом общий перевод данного слова. Пример схемы перевода (существительного *award*) дан ниже.

Обозначения в столбцах таблицы следующие: в первом столбце порядковый номер оператора, во втором — код оператора (проверки вправо, влево, перевод), в третьем — номер правила пропуска, в четвертом — детерминанта или перевод, в пятом — номер оператора отхода по «нет» при проверке. Слова, образующие список, даются отдельно, в правой стороне таблицы.

Контекстологический словарь (КС) содержит алгоритмы перевода многозначных существительных, прилагательных, глаголов и наречий. Всего в КС имеется около 3 тыс. схем перевода слов. Первый вариант словаря был составлен для сравнительно ограниченной, хотя и достаточно широкой области современного английского языка — для подязыка газетных сообщений на политические, публицистические, экономические и научно-технические темы, затем словарь был пополнен терминами по вычислительной технике и программированию. Идномы в словарь не входят, поскольку они образуют словарь оборотов. Однако вследствие отсутствия в языкознании и тем более в переводческой практике четких критериев отделения фразеологических единиц от свободных словосочетаний часть оборотов может быть найдена в словаре. Заметим в связи с этим, что степень фразеологической связанности компонентов словосочетаний нерелевантна для разрешения лексической многозначности методом детерминант.

Таблица 18

AWARD

1. lft	contract (n) 3	7. lft 1	ceromony 9
2. tr 1	сдача подряда	8. tr	присуждение наград
3. lft	денежная единица 5	9. tr	присуждение
4. tr	субсидия		List (1)
5. lft 1	list (1) 7	1. term (n)	
6. tr	субсидия	2. account (n)	

Структура и содержание КС делают возможными его разнообразные применения, главными из которых являются:

— непосредственное использование в системах автоматической обработки текстов, в МП или информационном поиске, для разрешения многозначности ключевых слов или для перевода многозначных слов на русский язык;

— использование в качестве справочного пособия при обучении английскому языку, поскольку словарь отражает наиболее актуальную сочетаемость английских слов, а также в качестве пособия переводчикам по публицистической (газетной) тематике;

— использование в качестве основы для программированного обучения английскому языку. Для такого использования удобна нынешняя форма словаря;

— использование для научно-исследовательской работы лингвистов в области современной английской лексикологии, лексикографии и машинной обработки текстов на естественных языках.

Исследование общих свойств схем перевода многозначных слов, которое было предпринято нами в (Марчук 1973), имело три основных направления:

— типизация схем перевода. В частности, были выявлены общие и частные структурные типы и подтипы схем;

— стандартизация и формализация операций по переводу слова и связанных с этим преобразований. На этом направлении были введены элементы контекстуальной модели перевода и модели перевода. В результате появилась возможность с помощью структурных типов и подтипов алгоритмов, а также контекстуальной модели перевода полностью задать в обобщенном виде любой алгоритм перевода. Алгоритм как таковой делался независимым от таблиц детерминант и переводов, что облегчает пополнение и изменения содержания словаря;

— изучение возможностей обобщения детерминант, особенно лексических и лексико-грамматических, с тем, чтобы добиться максимальной общности в разрешении лексической многозначности и перейти от отдельных контекстных явлений к некоторым классам явлений, если это возможно (и в какой мере это возможно).

Составление и эксплуатация полного словаря в режиме экспериментального МП, формулировка новых условий его работы в связи с диалоговым режимом дают возможность вновь рассмотреть указанные задачи уже на новом материале и в свете нового подхода к цели исследования.

Метод детерминант позволяет описать алгоритм перевода многозначного слова (ПМС) в обобщенном виде. Система, в которой определен алгоритм, задается пересечением двух осей: структурной, в которой действуют понятия «структурный тип» (СТ) и «структурный подтип», и операционной, где алгоритм характеризуется контекстуальной моделью перевода и индексом

контекстуальной модели. Индекс контекстуальной модели перевода показывает распределение элементов модели по детерминантам алгоритма. Конкретный алгоритм перевода каждого многозначного слова задается таблицей детерминант-переводов. При переходе к многозначным словам других лексико-грамматических классов (от глаголов к существительным, прилагательным и наречиям) структурная ось остается неизменной³. Контекстуальные модели перевода, в свою очередь, будут различны для каждой части речи (лексико-грамматического класса).

Контекстуальная модель перевода (КМП) есть сумма элементов контекстуальной модели перевода (элементов КМП). Элемент КМП есть некоторый типовой набор информации о слове-детерминанте, основном слове и его переводе. В этот набор входят следующие данные:

- а) принадлежность детерминанты к лексико-грамматическому классу;
- б) правила поиска детерминанты (какие слова находятся между основным словом и детерминантой);
- в) участие в трансформациях;
- г) состав лексической единицы-эквивалента в выходном языке;
- д) указание, переводится ли детерминанта вместе с основным словом. Для грамматически-выраженных детерминант пункты а), б), д) нерелевантны.

Пункты а) и б) позволяют выбрать правило пропуска. Пункт в) определяет лексические и, возможно, синтаксические преобразования, связанные с переводом данной лексической единицы, а также направление поиска детерминант (например, перемена направления детерминирования для глагольных конструкций с пассивом, когда правая лексическая детерминанта становится левой). Пункт г) нужен для синтеза переводящей словоформы, пункт д) — для правильного выбора оператора перевода. В целом элемент КМП описывает собственно лексические связи переводимого слова с контекстом с точки зрения их качественного характера и синтаксическую связанность каждого лексического значения в ее алгоритмическом выражении. Поэтому совокупность элементов КМП для слов лексико-грамматических классов, рассмотренных в данной грамматике, дает наглядную картину контекстных зависимостей многозначных слов в их формальном проявлении. Таким образом, наиболее интересные данные, которые можно получить, исследуя полный состав словаря в противопоставление выборке, изученной ранее, прежде всего относятся к характеристикам контекстуальных моделей перевода многозначных слов.

Рассмотрим основные элементы КМП для глаголов, существительных, прилагательных и наречий КС. Каждый элемент описывает одну или несколько детерминант в схеме.

Для глаголов исходной служила выборка, описанная в (Марчук 1973)⁴. Элементы КМП первого типа имеют в качестве детерминант существительные. При поиске пропускаются слова, которым приписан признак СФ «определение» — прилагательные, причастия и пр. Для всех элементов этого типа характерна трансформация пассива. Перевод производится либо одним русским глаголом, либо сочетанием с предлогом, либо фразеологическим сочетанием. Детерминанта в большинстве случаев не переводится (см. табл. 19).

Элементы КМП второго типа характеризуются наречием или предлогом в качестве лексической детерминанты. С алгоритмической точки зрения оба эти класса эквивалентны, поэтому для них принят один термин «послелог». Правила поиска разнообразны. В трансформациях детерминанта не участвует. Перевод производится либо одним глаголом, либо сочетанием глагола с предлогом. Детерминанта переводится вместе с основным глаголом (см. табл. 19).

Элементы КМП третьего типа, в отличие от первых двух, имеют составную детерминанту из существительного и послелога. Правила поиска варьируются. Некоторые виды элементов предусматривают трансформацию пассива. Детерминанты переводятся вместе со словом. Как видно из табл. 19, в большинстве случаев эти детерминанты описывают устойчивые фразеологические сочетания.

Для определения элементов КМП существительных исходная выборка была получена следующим образом. Из схем перевода существительных в КС были взяты первые десять по каждой букве алфавита. Если учесть, что по некоторым буквам число схем менее десяти (в этих случаях схемы брались из следующей буквы дополнением до десяти), то общая выборка составила, таким образом, около 200 схем.

В наблюдаемых элементах КМП для существительных участие в трансформациях ограничивалось трансформацией определения. Существительное в препозиции другому получает СФ определение и переводится однокорневым прилагательным. Если однокорневого прилагательного нет и требуется перевод определения в форме родительного падежа, соответствующая информация указывается в таблице «существительные-определения», а также в ИЯ в виде признака «нет определения». Если требуется не однокорневое прилагательное, а какое-либо другое, то такой случай специально предусматривается в схеме ПМС и образует особую разновидность элемента КМП.

³ Структурные типы и подтипы устанавливались на выборке из 100 алгоритмов каждого лексико-грамматического класса, поэтому они отражали универсальные свойства алгоритмов перевода в отношении состава и структуры.

⁴ Было взято 200 алгоритмов перевода глаголов частотности выше средней (самые частые глаголы исключены) и средней. Из них 70 алгоритмов были исследованы для выделения элементов КМП.

Таблица 19

Чем выражена детерминанта. Обозначение	Промежуточные слова	Участие в трансформациях	Чем переводится	Переводится ли детерминанта	Пример
Элементы КМП первого типа (глаголы)					
Существ. ЭС 1	Прилаг. или причаст.	Пассив	Одним русским словом	Нет	<i>to guide story</i> 'инспирировать' <i>to heighten tension</i> 'усилить'
Существ. ЭС 2	»	»	Сочетанием слов	»	<i>to injure interest</i> 'нанести ущерб' <i>to invite suggestion</i> 'предложить прислать'
Существ. ЭС 3	»	»	Одним глаголом с предлогом	Нет	<i>to grant point</i> 'согласиться с'
Существ. ЭС 4	»	»	Одним глаголом или идиомой	Да	<i>to handle</i> 'обращаться с' <i>to gain ground</i> 'укрепиться' <i>to gain stature</i> 'занять свое место'
Элементы КМП второго типа (глаголы)					
Послелог ЭП 1	Нет	Нет	Одним глаголом	Да	<i>to hand over</i> 'передать' <i>to furnish with</i> 'снабдить'
Послелог ЭП 2	Существ. и прилагат. или причастие	»	Одним глаголом с предлогом	Да	<i>to identify with</i> 'отождествить с' <i>to tack on</i> 'добавить к'
Элементы КМП третьего типа (глаголы)					
Послелог+ +существ. ЭПС 1	Нет	Нет	Идиомой	Да	<i>to jump to conclusions</i> 'делать поспешные выводы'
Послелог+ +существ. ЭПС 2	Нет	Нет	Одним глаголом	Да	<i>to kick over traces</i> 'взбунтоваться'
Существ.+ +послелог ЭПС 1	Прилаг.	Пассив	Словосочетанием	Да	<i>to lose sight of</i> 'упустить из виду' <i>to lose hold of</i> 'выпустить из рук'

Таблица 19 (продолжение)

Чем выражена детерминанта. Обозначение	Промежуточные слова	Участие в трансформациях	Чем переводится	Переводится ли детерминанта	Пример
Элементы КМП первого типа (существительные)					
Существ. ЭС 1	Прилаг. причастие	Определен.	Одним существ.	Нет	<i>abandonment</i> 'утрата'
Существ. ЭС 2	»	»	Одним прилагат.	»	<i>background</i> 'основной'
Элементы КМП второго типа (существительные)					
Предлог ЭП 1	Нет	Нет	Одним существ.	Нет	<i>adjustment</i> 'приспособление'
Предлог ЭП 2	Прилаг.	Нет	Одним существ.	Да	<i>against background</i> 'на фоне'
Элементы КМП третьего типа (существительные)					
Прилагат. ЭПр 1	Нет	Нет	Одним существ.	Да	<i>productive capacity</i> 'производительность'
Прилагат. ЭПр 2	Нет	Нет	Одним существ.	Нет	<i>deduction</i> 'заключение'
Элементы КМП четвертого типа (существительные)					
Глагол ЭГ 1	Нет	Нет	Сочетанием с глаголом	Да	<i>to carry ball</i> 'проявлять активность'
Глагол ЭГ 2	Нет	Нет	Одним существ.	Нет	<i>effort</i> 'попытка'
Элементы КМП пятого типа (существительные)					
Послелог+ +глагол ЭПГ 1	Нет	Нет	Одним существ.	Да	<i>build-up capability</i> 'боеспособность'
Прилаг.+ +существ. ЭПрС	Нет	Трансфор. с предлогом	»	Нет	<i>good will gesture</i> 'проявление'
Элементы КМП первого типа (прилагательные)					
Существ. ЭС 1	Нет	Нет	Одним прилагат.	Нет	<i>absolute</i> 'полный'
Элементы КМП второго типа (прилагательные)					
Существ. ЭС 2	Нет	Нет	Прилагат. с наречием	Нет	<i>adaptable</i> 'легко приспособляемый'
Элементы КМП третьего типа (прилагательные)					
Существ. ЭС 3	Нет	Нет	Существ.	Да	<i>armoured concrete</i> 'железобетон'

Таблица 19 (окончание)

Чем выражена детерминанта. Обозначение	Промежуточные слова	Участие в трансформациях	Чем переводится	Переводится ли детерминанта	Пример
Элементы КМП первого типа (наречия)					
Прилагат. ЭПр	Нет	Нет	Одним наречием	Нет	<i>adequately</i> 'достаточно'
Элементы КМП второго типа (наречия)					
Глагол ЭГ 1	Нет	Нет	Одним наречием	Нет	<i>affirmatively</i> 'твёрдо'
Элементы КМП третьего типа (наречия)					
Существ. ЭС 1	Нет	Нет	Одним наречием	Нет	<i>alone</i> 'только'

Элементы КМП для существительных отличаются большим разнообразием, как видно из табл. 19.

КМП прилагательных отличаются однообразием: в основном это модель КМП 1, состоящая из ЭС 1.

Элементы КМП для наречий состоят из трех основных типов.

Для определения элементов КМП прилагательных и наречий исходная выборка была получена тем же способом, что и для существительных.

Для прилагательных характерны элементы КМП — табл. 19.

Модели перевода для наречий не отличаются разнообразием, однако их больше, чем КМП для прилагательных.

Более тщательное исследование КМП для слов основных лексикограмматических классов, представленных в КС, даст возможность расширить список как элементов КМП, так и моделей в целом. Однако основные закономерности хорошо проявляются уже на обследованном материале.

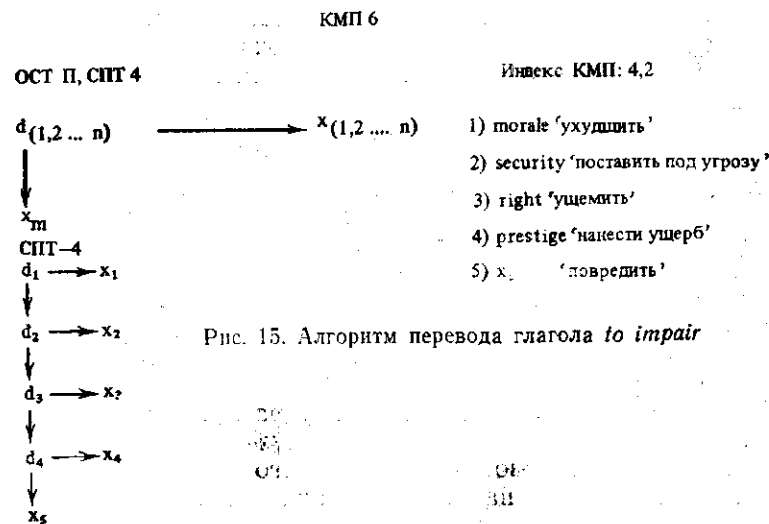
Можно считать установленным, что инвентарь контекстных зависимостей и связанных с ними преобразований для нахождения переводных соответствий наиболее обширен у глаголов (что соответствует интуитивным предположениям), а также у существительных (что трудно отнести за счет каких-нибудь специфических синтаксических характеристик существительного и можно лишь полагать вызываемым особенностями английского существительного в предложной системе при отсутствии положительных формальных признаков для существительного в тексте). Бедность контекстных связей прилагательного в переводе объясняется довольно четко фиксированной ролью прилагательного. Наконец, разнообразие элементов КМП для наречий точно соответствует известному из обычной грамматики принципу соотносительности наречий с определяемыми ими частями речи.

Аппарат КМП дает возможность достаточно подробно отразить контекстные связи подлежащих переводу элементов пред-

ложения. В рамках трансляционной грамматики и алгоритма возможности, предоставляемые этим аппаратом, позволяют передавать довольно разнообразную гамму переводных соответствий в том случае, если их можно фиксировать на лексическом уровне.

Рассмотрение алгоритмов перевода многозначных слов показало следующее.

1) Множество алгоритмов перевода многозначных слов каждого лексико-грамматического класса содержит алгоритмы,

Рис. 15. Алгоритм перевода глагола *to impair*

структура которых характеризуется одним СТ. Все множество алгоритмов перевода слов одного класса описывается конечным набором структурных типов. Структурные типы алгоритмов для слов одного лексико-грамматического класса назовем частными структурными типами (ЧСТ).

2) Имеются СТ, общие для алгоритмов перевода слов, принадлежащих к разным лексико-грамматическим классам. Такие СТ назовем общими структурными типами (ОСТ).

В структурном подтипе структурного типа (СПТ) обозначена каждая детерминанта и каждый перевод. Число детерминант и переводов соответствует их числу в индивидуальном алгоритме. Указанием на СПТ описывается с точностью до одной детерминанты и одного перевода структура алгоритма перевода многозначного слова.

Множество алгоритмов перевода, имеющих один структурный тип, распадается на подмножества, имеющие общий структурный подтип, так что число структурных подтипов значительно меньше числа индивидуальных алгоритмов перевода слов данного лексико-грамматического класса.

Структурный тип, подтип, индекс контекстуальной модели перевода и таблица детерминант-переводов однозначно задают индивидуальный алгоритм перевода многозначного слова. На рис. 15 изображен алгоритм перевода многозначного английского глагола *to impair*.

Проверка гипотезы о наличии общих и частных структурных типов на материале выборки по 100 алгоритмов перевода многозначных слов на каждого лексико-грамматического класса описана нами в (Марчук 1973). В полученной выборке были выделены типы структур алгоритмов. Были обнаружены три структурных типа, общие для всех лексико-грамматических классов. Кроме общих структурных типов у каждого лексико-грамматического класса оказалось в наличии несколько частных структурных типов.

Проведенное обследование позволяет считать подтвердившейся гипотезу о том, что алгоритмы перевода имеют общие структурные типы, а число подтипов значительно меньше числа индивидуальных алгоритмов перевода. Явным образом преобладают ОСТ, на их долю приходится в среднем 60—80% всех алгоритмов. ОСТ составлены только из лексических детерминант. Два типа односоставны, один двухсоставен. Это говорит о том, что для основных лексико-грамматических классов и основной массы многозначных слов лексическая многозначность разрешается относительно простыми схемами контекстного анализа.

Полученные результаты позволяют развить дальше выводы, к которым пришел А. Каплан, исследуя разрешающую силу контекста в рамках своего ограниченного эксперимента (Kaplan 1955). По нашим наблюдениям оказывается, что разрешающая сила контекста — точнее, объем разрешающего контекста, — зависит от лексико-грамматического класса исследуемого слова. Для глаголов и прилагательных наибольшей разрешающей силой обладает контекст с правой лексической детерминантой, в то время как для существительных и наречий такой силой обладает контекст из одного предшествующего и одного последующего слова.

Тотальное исследование схем КС показывает, что наиболее частые многозначные слова разрешают многозначность с помощью алгоритмов, устанавливающих относительно небольшое количество видов связи. Простая схема запроса контекста позволяет разрешить многозначность значительного числа слов в достаточно большом количестве случаев употребления.

На представленном материале не прослеживается также сколько-нибудь явной зависимости между числом переводов, составностью структуры и частотой многозначного слова⁵.

⁵ Составность структуры — количество разных видов проверок контекста в структурном типе и подтипе.

5.3. Комплементарные методы

При построении алгоритмов распознавания значений многозначных слов возникает ряд случаев, не разрешимых в рамках метода детерминант. Эти случаи можно систематизировать в три типа. Первый можно назвать проблемой общего выхода алгоритма, второй — проблемой фразеологических единств, третий — проблемой метатаксных преобразований (Tesnière, 1959). Рассмотрим методы, с помощью которых указанные случаи можно было бы решить в рамках теории детерминант.

Связь между частотностью и дистрибуцией неоднократно отмечалась многими исследователями. При последовательном проведении принципа «каждому значению отвечает определенная дистрибуция» нужно и для главного (основного) значения указать тип дистрибуции. Но каким принципом при этом руководствоваться? Например, можно предположить, что формула дистрибуции для такого значения должна быть наименее специализирована и наиболее частотна. Однако на самом деле зависимость между специализированностью и частотностью является весьма сложной.

Рассмотрим пример.

Глагол *to put* в трех своих значениях может быть описан следующими дистрибутивными формулами:

- 1) 'примириться': $P + put + up + with + C$. *He never put up with injustice.*
- 2) 'положить': $P + put + up + C_1 + on + C_2$. *She put up the book on the shelf.*
- 3) 'отложить': $P + put + off + C$. *They put off their visit for a week.*

Которая из этих формул наиболее специализированная? Какая наиболее частотна? Если понимать под специализацией усложненность формул дистрибуции в смысле количества компонентов, то наиболее специализирована формула 2. Однако далеко не очевидно, что она менее частотна, чем формулы 1 или 3. Во всяком случае частотность не является легко определяемым признаком для каждого из возможных переводов глагола, потому что трудно определить масштаб выборки, в которой все частотные характеристики всех входящих в алгоритм переводов были бы достаточно достоверны, так как в разных наборах контекстов разные значения одного и того же глагола непременно имеют разные частоты. Затруднительна также (хотя и не невозможна) частотная характеристика главного значения. Несмотря на то, что при достаточно большом конкордансе сведения будут устойчивые, переход, например, на новую тематику потребует заново производить такое определение значения.

Выбор основного значения по степени «общности» также представляет трудности, так как «общность» можно понимать как: 1) первое, основное значение слова в обычном переводном словаре; 2) основное, наиболее распространенное значение слова в данном конкордансе; 3) основное значение данного слова в генеральной совокупности текстов, известное составителю алгоритма, но почему-либо не совпадающее со значениями по пп. 1 и 2.

В КС данная проблема имеет следующее рабочее решение. Составитель алгоритма имеет право действовать как информант и производить выбор недетерминируемого значения в качестве общего выхода по любому из трех приведенных выше соображений. Все другие значения признаются частными и детерминируются. Выбор общего значения может, в частности, определяться именно этим последним обстоятельством, т. е. необходимостью детерминировать остальные значения.

Таким образом признается, что каждое многозначное слово имеет одно главное значение, которое специально не детерминируется, хотя оно и может иметь, строго говоря, специальную дистрибутивную формулу, которая не подходит для частных значений.

Это значение является так называемым общим выходом, в который текстовое слово попадает при последовательном отходе по «нет» от всех основных ветвей алгоритма. Детерминируются значения, отличающиеся от общего выхода. Это значит, что к методу детерминант подключается метод конвенций по вопросу об иерархии значений полисемантического слова и отбору главного значения.

Принцип контекстной детерминации применительно к лексике имеет некоторые отличия от детерминации в грамматическом анализе. Есть отличия в решении вопроса об общем выходе. Во-первых, при анализе с целью определения грамматической информации, например СФ, общий выход в большей степени зависит от частоты и соображений простоты дистрибуции, чем при переводе многозначных слов. Во-вторых, решение о переводе слова не подлежит пересмотру. Решения же относительно грамматической информации могут быть скорректированы на последующих этапах анализа.

Относительно фразеологических единств следует иметь в виду, что степень фразеологической связанности с алгоритмической точки зрения не релевантна для процесса разрешения многозначности. Можно, таким образом, утверждать, что метод детерминант сочетается с методом построения словаря оборотов. Обороты, включенные в словарь оборотов, не должны входить в КС, и наоборот.

Рассмотрим вопрос о метатаксных преобразованиях. Некоторые части речи (лексико-грамматические классы) характеризуются особой синтаксической ролью в предложении. Так, активная роль глагола в синтаксисе и специфика алгоритмического

разрешения многозначности (см. блок-схему общего алгоритма) приводят к затруднению, выражающемуся в том, что перевод отдельных глаголов (равно и некоторых других частей речи) требует выяснения того, в каких синтаксических конструкциях (переводимой и переводящей) данное слово должно участвовать; обратно, применение переводящей синтаксической конструкции зависит от того, какой перевод получит образующий конструкцию глагол. Метод детерминант допускает анализ синтаксических явлений и включение соответствующих правил в алгоритм (с помощью лексико-грамматических и морфологических детерминант). Однако включение сложных преобразований синтаксических структур в схемы перевода многозначных слов может затемнить общую концепцию системы. Целесообразно анализировать и переводить конструкции, локализованные при отдельных словах, на этапе «трансформационный анализ и синтез».

Для англо-русского МП ситуации подобного рода обозримы и могут быть заданы перечислением.

1) а) в исходном списке отмечено пять глаголов, а именно *to like, lack, need, preface, seem*, которые во всех своих переводах, зафиксированных в конкордансе, требуют изменения синтаксической структуры предложения;

б) отмечено четыре глагола, которые в пассиве требуют для перевода особых конструкций, неконгруэнтных английским. Это глаголы *to quote, say, report, know*;

в) в особых конструкциях участвует английский глагол *to know* в пассивных оборотах. *He is known to have said. They were known to have been living near Paris, etc.*

2) Требуют специального анализа конструкции со вторичной предикацией. В большинстве случаев такого рода требуется членение исходного предложения и включение в него придаточных (на языке перевода).

Для каждого случая употребления глагола названного выше перечня существует конечная последовательность операций, которая приводит к правильному идиоматическому переводу. Так, рассмотрим предложение с глаголом *lack*. *He lacks this quality*. Перевод: 'у него отсутствует это качество'. Возможная алгоритмическая последовательность действий: 1) перевести глагол *lack* глаголом 'отсутствовать'; 2) подлежащее оставить на своем месте, но приписать ему признак «родительный падеж»; 3) перед подлежащим восстановить предлог 'у'; 4) дальнейшие действия производить согласно предусмотренным в общем алгоритме. Таким же способом можно описать анализ и перевод других конструкций с подобными словами. Ясно, что в совокупности таких алгоритмических описаний можно будет выделить общие части и в этом случае алгоритм в целом получится более экономным и универсальным, чем сумма отдельных преобразований.

3. Лексически-выраженная детерминанта может находиться так далеко от исследуемого слова, что с помощью существующих

правил пропуска ее обнаружить не удастся без перехода в соседнее предложение, что возможно, но нежелательно из-за опасности ложного детерминирования. Пример: (правая лексическая детерминанта глагола *to take* находится в придаточном предложении, в то время как сам глагол в главном)... *positions they will take in Paris*. Такие случаи, хотя они кажутся довольно естественными, достаточно редки: на 700 употреблений глагола *to take* в текстах таких случаев всего 4.

4) Для пассивной конструкции правая лексически выраженная детерминанта становится левой. Может быть осуществлена несложная операция повторения поиска правой лексической детерминанты влево: сигналом к такой операции будет наличие признака «пассив», который сообщается глаголу специальной схемой анализа.

5.4. Типизация детерминант и общие свойства КС

Материал КС дает широкие возможности для изучения типизации детерминант. ЛГ-детерминанты в большинстве случаев определяют значения, которые наиболее далеки друг от друга, а Л-детерминанты — все прочие значения, которые по степени близости не дифференцируются. ЛГ-детерминанты в схемах перевода других многозначных слов также дают возможность сделать соответствующие обобщения. Однако путь расширения алгоритма в смысле замены одних Л-детерминант на более широкие классы и подклассы слов начинается, по-видимому, от рассмотрения конкорданса для каждого многозначного слова, а замена и обобщение детерминант возможны лишь после изучения конкордансов и алгоритмов перевода всех слов, для которых такая замена предполагается. От текста и от реально наблюдаемой лексической и грамматической сочетаемости, релевантной переводу, можно переходить к составлению списков слов, а затем к расширению этих списков до классов и подклассов, которые могли бы служить в качестве детерминант.

Три основных направления исследования общих свойств схем перевода многозначных слов, оцененных по своим результатам на материале всего КС и в свете новых возможностей использования словаря (в диалоговом режиме с интер- и постредактированием), можно на настоящем этапе исследования охарактеризовать следующим распределением акцентов.

Типизация схем перевода и выделение общих структурных типов и подтипов относятся к алгоритму перевода многозначных слов по контексту в собственном смысле, как к последовательности действий определенного характера. Существование типов позволяет отвергнуть взгляд на контекстное определение многозначности как ведущее к созданию необозримого количества сугубо индивидуальных алгоритмов, что якобы препятствует машинному решению проблемы.

Обобщение и типизация детерминант по степени общности показывают, что в принципе можно разрешать лексическую многозначность в пересечении крупных классов слов. Однако эти классы не будут синтаксическими. Они могут быть лексико-грамматическими (части речи), но только далеко не для всех случаев разрешения многозначности. Основная масса лексических контекстных связей многозначных слов индивидуальна. Это не мешает возможности использовать эти связи для выявления переводных соответствий на лексическом уровне. Типизация детерминант в виде перехода от отдельных слов-детерминант к детерминантам-спискам, классам и подклассам слов — необходимое и в какой-то мере неизбежное явление по мере расширения исходного конкорданса и уменьшения неформализуемого остатка. Однако эта типизация не может рассматриваться как самоцель в построении КС инженерного типа: типизация должна проводиться не ради типизации, а для достижения технологических или лингвистических показателей, которые без нее невозможны.

Наиболее важным представляется направление исследований, связанное со стандартизацией и формализацией операций по переводу слова, отражаемых в контекстуальной модели перевода. Именно в этом аспекте выявляются релевантные переводу свойства схем, и возможности установления переводческих соответствий на разных уровнях языковой структуры в ракурсе лексического перевода отражаются в элементах КМП. Выявление элементов КМП, проведенное в настоящем исследовании, может быть значительно пополнено тотальным обследованием содержащихся в КС алгоритмов. Дело, однако, не только в выявлении использованных моделей, но и в том, что создание новых элементов КМП и КМП в целом может существенно расширить возможности передачи переводческих соответствий. Кроме того, здесь весьма показательные результаты дала бы тотальная статистика введенного в память ЭВМ КС, и статистическое соотношение типов зависимостей позволило бы обратить внимание на наиболее актуальные типы связей. Множество контекстных, часто перекрещивающихся, связей многозначных слов, с разнообразными факторами и типами зависимостей, на самом деле, благодаря предложенному формализму, достаточно хорошо упорядочивается, не теряя при этом разрешающих свойств.

Дальнейшее детальное изучение КМП целесообразно принимать на уровне конкретного подязыка, подлежащего переводу. Не требует особых доказательств положение о том, что чем шире подязык, тем разнообразнее набор контекстных связей и зависимостей, подлежащих учету в рамках переводных соответствий. Подязыковая ориентировка набора КМП позволит повысить эффективность системы МП в достаточной для практического использования степени.

5.5. Эффективность перевода слов контекстологическим словарем

Учитывая методику составления алгоритмов перевода многозначных слов, можно утверждать, что реальной проверке эти алгоритмы подвергнутся тогда, когда объем переведенных текстов будет соизмерим с объемом исходного конкорданса.

Проверка алгоритмов на новом материале, не использованном при их составлении, имеет две цели: пополнение алгоритмов для повышения их разрешающей способности и выявление основных типов ошибок, с помощью чего можно оценить возможности метода.

Таблица 20

Тип ошибки	Число ошибок в обследованном объеме текстов, в %
В алгоритме отсутствует лексическая детерминанта. Перевод есть и легко детерминируется	33
В алгоритме нет детерминанты и нет перевода. Перевод легко детерминируется лексически	34
В алгоритме нет перевода и нет детерминанты. Перевод детерминируется грамматически	12
В алгоритме не детерминирован выбор глагола из пары возвратный—невозвратный (частный случай детерминирования, возможно грамматическое детерминирование)	7
Неправильен общий выход алгоритма перевода	2
Детерминирование в рамках метода невозможно	12
Итого:	100

Алгоритмы КС неоднократно проверялись в ходе машинных и ручных экспериментов. При машинном эксперименте переводятся подряд все многозначные слова, встретившиеся в данном массиве текстов; при этом действует закон частотности и в большинстве случаев переводятся наиболее распространенные слова, которым приписываются правильные переводы, однако из одного и того же диапазона ЦЭ. Поэтому более показательны результаты ручной проверки, в ходе которой можно варьировать проверяемый материал, т. е. подбирать контексты, в которых встречались бы разные многозначные слова.

Результаты проверки 200 алгоритмов перевода многозначных глаголов «средней» многозначности на текстах общим объемом 30 000 словоупотреблений показывают, что из 250 употреблений глаголов неправильно перевелось 58, т. е. около 20%. Анализ

неправильных переводов дает результаты, сведенные в таблицу 20.

Таким образом, расширение конкорданса устраняет ошибку в 88% случаев. Если учесть, что многозначные глаголы, работа схем которых проверялась, наиболее тесно связаны с синтаксической и семантической структурой предложения, т. е. представляют собой наибольшую сложность для перевода методом детерминант, то полученную эффективность следует рассматривать как некоторый нижний предел, поскольку для многозначных слов других частей речи можно ожидать лучшей эффективности по причине более простых контекстных связей. В среднем для многозначных слов всех частей речи метод дает эффективность до 90%, с учетом стандартизации терминологии и узкой специализации текстов — до 95%.

Анализ перевода многозначных слов научно-технических подязыков, выполненный на материале переведенных системой АМПАР текстов, показал, что на этапе до комплексной отладки и настройки системы типы ошибок сохраняются примерно в той же пропорции. Следующим этапом проверки КС должно быть обследование результатов перевода устойчивых массивов текстов заданного подязыка. В настоящее время эта работа проводится на текстах по программированию и вычислительной технике.

ПЕРСПЕКТИВЫ МАШИННОГО ПЕРЕВОДА

6.1. Качество машинного перевода

Вопрос о качестве машинного перевода важен не только сам по себе. Он в значительной мере стимулировал исследования по прагматическим аспектам перевода вообще¹. Основным вопросом для МП является приемлемость его для потребителей. На заседании специальной комиссии Европейского экономического сообщества в 1978 г. вопрос был рассмотрен именно с этой точки зрения. Было установлено, что следует проводить различие между глобальной, или макро-оценкой, главными критериями для которой должны быть общая понятность (intelligibility) и время редактирования, и аналитической, или микро-оценкой. Оценка общей понятности может быть дана экспертами; однако время редактирования оказалось весьма субъективным фактором, зависящим от компетенции, подготовки и доброй воли редактора. Микрооценка состоит в определении количества ошибок различного типа, содержащихся в переводе; эта оценка обеспечивает необходимую обратную связь для постоянного улучшения словарей и программ системы. Скорость редактирования, которую можно измерять с помощью количества слов, изменяемых в процессе редактирования, также является хорошим показателем качества. Различные потребители будут платить по-разному за различные свойства получаемого ими перевода. Таким образом, уровень качества должен определяться в зависимости от целей, которые ставятся перед системой МП в конкретной задаче информационного обслуживания (Rolling, 1978; Bruderer, 1978).

Еще в 1956 г. американские психологи Миллер и Биб-Сентер предложили следующие методы оценки качества перевода:

1) проставление оценки качества перевода испытуемыми по стабильной шкале;

2) подсчет процента слов, совпадающих в неотредактированном и отредактированном переводе;

3) изучение ответов человека, читавшего только перевод, на вопросы, поставленные по тексту оригинала.

Психолог Ш. Пфаффлин использует для оценки качества перевода следующие два критерия:

— оценка понимания перевода путем постановки вопросов по тексту;

— оценка ясности смысла перевода по шкале «ясно», «неясно», «бессмысленно» (Pfafflin, 1965).

Ни один из этих методов не дает четкого ответа на вопрос о том, каковы критерии качества машинного перевода и каков допустимый уровень качества. Более подробная шкала разработана в Национальной физической лаборатории в Теддингтоне, Англия; эта шкала построена на основании соображений о том, что экспертам следует предоставить полную свободу для решения того, насколько нужен МП. Шкала следующая:

8. Перевод вполне адекватный. Смысл сразу ясен, хотя и не всегда выражен вполне правильно.

7.²

6. Перевод в основном хороший, неясны лишь несколько фраз, так что может быть упущено что-нибудь существенное, но, как правило, перевод в целом достаточно ясен и понятен.

5.

4. Неплохо. Однако требуется немало времени, чтобы разобраться в смысле, причем все равно нет полной уверенности в правильности понимания.

3.

2. Плохо. Перевод может быть полезен только тому, кто готов преодолевать значительные трудности, но даже и такой человек может быть разочарован.

1.

0. Перевод бесполезен (Мак-Дэннел, Прайс, Зансер, Пейтс, 1971).

Эта шкала дает более широкие границы для оценки МП.

В советской литературе исследования по качеству МП крайне немногочисленны. В опубликованных работах предлагалось различать три отдельные его характеристики: понятность, верность оригиналу и приемлемость³. Каждая из этих характеристик измеряется по особой шкале, в рамках которой используются экспертные оценки специалистов. Интересно, что советский опыт оценки качества МП дал в одном аспекте результаты, сходные с оценкой Б. Достерта, а именно: на качество и понятность МП влияют в основном семантические (лексические) свойства перевода (непереведенная лексика и пр.), а синтаксические (грамматические) свойства текста в этом отношении менее важны.

Обычный перевод может обладать массой недостатков, отсутствующих в МП (например, гипертрофия черт индивидуального стиля человека-переводчика и т. п.). С другой стороны, «гладкость» грамматической структуры обычного перевода, даже при его непонятности и ошибках в содержании, может повлиять на его оценку в сторону завышения. Наоборот, шероховатость грамматической структуры МП может привести к занижению его оценки.

² Пустые рубрики введены для промежуточных оценок.

³ Обобщающая работа на эту тему, по-видимому, в советской литературе всего одна, а именно (Москович, Вишнякова 1970).

¹ По выражению теоретика перевода Е. Найды на вопрос: «Правильен ли этот перевод?» следует отвечать другим вопросом, а именно: «Для кого?» (Цит. по: Dostert 1973).

Рассматривая существующие способы оценки качества МП, можно сделать следующие выводы:

1. Нет четкой дифференциации в методах оценки машинного и обычного перевода. В результате методы оценки МП не отличаются от методов оценки, которые могли бы быть применены к обычному переводу.

2. Оценка прагматической стороны смешивается с оценкой качества. Представляется нецелесообразным вводить в оценку качества положение о том, кому и зачем он нужен. Последний аспект следовало бы оценивать отдельно, т. е. сначала, видимо, нужно определить цель, для которой осуществляется МП, а затем устанавливать, каково качество выполнения поставленных заданий. При оценке качества обычного перевода, как правило, уже исходят именно из такого положения.

3. Понятность перевода и верность оригиналу не следует рассматривать независимо друг от друга. Вообще говоря, можно предположить, что текст, смысл которого абсолютно неясен, будет сохранять верность оригиналу; или наоборот, перевод совершенно не передает смысла оригинала, но смысл всего текста перевода ясен и не вызывает сомнений. Однако в таких ситуациях положительные оценки не имеют значения, они не касаются подлинной ценности машинного перевода или вскрывают какие-то совершенно особые случаи.

4. В существующих способах оценки, если мы зададимся целью ориентироваться на более точную оценку качества МП после того как определено его целевое назначение, оказываются отсутствующими по крайней мере два формальных параметра, которые в значительной мере могут определять пригодность МП и его качественную оценку: объем текста перевода и сложность его с точки зрения используемых грамматических конструкций и лексики.

Переходя к оценке качества перевода, целесообразно обратиться к представлениям теории перевода в той части, в которой говорится об эквивалентности перевода по уровням. Представляется, что в оценке качества МП была незаслуженно игнорирована структура достижения эквивалентности по уровням «перевод слов» — «перевод словосочетаний» — «перевод предложений» — «перевод текста». Машинный перевод можно достаточно точно оценивать на каждом из этих уровней, даже включая уровень морфем. Однако представляется, что для прагматических целей наиболее удобно начать рассмотрение от уровня текста.

Во всех рассмотренных способах оценки понятие «текст перевода» отсутствует. Между тем очевидно, что этот параметр имеет определенную прагматическую ценность. Так, если правильно переведен текст в одно предложение, то оценка его должна заведомо отличаться от оценки текста в два предложения, из которых одно переведено правильно, а второе — нет, или от текста в три предложения, каждое из которых может быть переведено либо правильно, либо неправильно.

Точную меру композиционной сложности текста установить достаточно трудно, поскольку неизвестно, какие грамматические структуры и какой выбор лексических средств выражения считать «простым» и отличать от «сложного». Понятие стиля слишком расплывчато, чтобы можно было пользоваться им применительно к такой точной задаче, какой является определение качества МП. Однако можно ввести некоторые общие соображения, помогающие принять рабочее решение. Так, можно предположить, что сложность переводимой фразы тем больше, чем больше длина ее в словах; для этой длины можно установить некоторый порог, в соответствии с которым назначать оценку качества.

Качество перевода должно обуславливаться его назначением. Рассмотрим градацию переводов по О. Каде: «черновой перевод», «рабочий перевод», «перевод для опубликования». Эта градация сделана применительно к научно-техническому переводу, в котором прагматический аспект в настоящее время выражен наиболее четко. В соответствии с практикой научно-технических переводов в СССР эти категории понимаются следующим образом:

— черновой перевод есть некоторый промежуточный продукт, представляющий собой первый, начальный вариант перевода. Он содержит новые (незнакомые) слова, которые не переведены, может содержать слова, неправильно переведенные, ошибки в грамматическом согласовании, неправильный перевод отдельных частей предложения, при сохранении смысла предложения в целом. Переводчик (редактор) исправляет черновой вариант, уточняя значения слов с помощью словаря, прибегая к помощи специалистов при неясности для него смыслового содержания или перевода отдельных слов-терминов. Примером чернового перевода в «человеческой» практике является перевод с листа в консультации со специалистом. Черновой перевод может быть и конечным продуктом в тех случаях, когда он используется как сигнальная информация, по которой специалист принимает решение о необходимости производства «рабочего» перевода или перевода «для опубликования»;

— рабочий перевод есть отредактированный черновой. Он не содержит смысловых ошибок. Новых и непереведенных слов в нем нет. Могут быть отдельные стилистические ошибки или вариативность при выражении одного и того же смысла. В выполнении рабочего перевода принимают участие, как правило, переводчик и редактор; в тех случаях, когда перевод осуществляется без редактирования, переводчик должен быть высококвалифицированным специалистом в соответствующей области знания;

— перевод «для опубликования» есть перевод, тщательно отредактированный, полностью отработанный стилистически, адекватный оригиналу, публикуемый государственными издательствами.

В указанной градации машинный перевод следует рассматривать как черновой.

Качество МП может быть определено по затратам труда, который необходим для превращения перевода из чернового в рабочий.

МП следует понимать как продукт, на уровне «чернового» перевода не подвергавшийся редактированию: черновой МП получается без участия человека. Уже на этом уровне он может быть использован в качестве сигнальной информации, в этом случае он редактированию не подвергается и применяется как конечный продукт. Интерес представляет МП, поступающий на редактирование человеку-редактору с целью превращения в рабочий перевод. Определенная часть редакторского труда может быть снята благодаря некоторой конвенции, получившей определенное применение; так, на переводе, выполненном на ЭВМ, может ставиться штамп «МП» или «машинный перевод», благодаря чему могут быть сняты претензии за грамматические неточности, стилистические шероховатости и пр., которые не искажают смысла. Поэтому целесообразно не вводить в число обязательных требований к МП полное сохранение грамматической правильности там, где известная избыточность языкового выражения, например, в русском языке, делает излишними некоторые виды грамматического оформления. Например, фразу такого вида: «В течение последних *несколько* лет отношения между Советским Союзом и молодыми национальными государствами стали даже более *активными* и более *тесные*» можно считать правильно переведенной, хотя по нормам русского языка следует в подчеркнутых словах исправить морфологические ошибки⁴. Сохранение смысла обеспечивается отсутствием двух типов ошибок, которые можно классифицировать как лексические и грамматические. Проиллюстрируем эти ошибки примерами из текста:

The problem of automation must become one of increasing concern to the Negro American Labor Council.

Thus they put forth a perspective of continued inequality until the problem for all is solved.

Если первое предложение все же понятно, хотя и после некоторого размышления, второе непонятно совершенно, вследствие того, что неправильно определена предикативность. Ошибки, затрудняющие понимание первого предложения:

— отсутствие перевода у одного слова. В данном случае оно оказалось общеевропейского корня и при выдаче латинским шрифтом в тексте перевода затруднений не вызывает: однако надо иметь в виду необходимость перевода всех слов;

— непонятно сказуемое предложения;

⁴ Морфологические ошибки подобного рода, видимо, можно будет исправлять автоматически с помощью программы-редактора, которая сможет работать без участия человека.

— неправилен перевод собственного имени, несущего значительную информационную нагрузку.

Действия редактора по превращению этого перевода в понятный должны быть следующие:

— перевести слово *automation* 'автоматизация';

— правильно перевести сказуемое, например так: 'должна вызывать все большую озабоченность';

— правильно перевести (во всяком случае однообразно по всему тексту перевода) название организации: 'Негритянский американский рабочий совет'⁵.

Таким образом, здесь три ошибки, требующие вмешательства редактора: одна грамматическая и две лексических. Предложение после их исправления имеет следующий вид: 'Проблема автоматизации должна вызывать все большую озабоченность Негритянского американского рабочего совета'.

Во втором предложении коррекция касается главным образом грамматических проблем. Для того, чтобы предложение сделать правильным и понятным, требуется:

— слово 'перспективу' поставить в правильном падеже;

— 'продолжающееся' лучше перевести как 'продолжение';

— вместо 'для всего' лучше перевести 'для всех' и поменять слова местами. Тогда правильный и понятный перевод будет выглядеть следующим образом: 'Таким образом они изложили перспективу продолжения неравенства до того, как проблема разрешается для всех'.

Здесь 'до того' можно было бы заменить на 'до тех пор, пока', 'разрешается' на 'пока проблема не будет разрешена' и т. д. Однако нетрудно убедиться, что все эти возможные преобразования отнюдь не являются необходимыми. Полный смысл предложений становится понятным и без них, гладкость стиля здесь может быть с полным основанием скомпенсирована штампом «машинный перевод», доверие к которому сохраняется правильной и исчерпывающей передачей смысла.

Ошибки, которые были исправлены во втором предложении, следует расценивать как грамматические.

Введем теперь классификацию ошибок, которые необходимо исправлять для превращения чернового машинного перевода в рабочий со штампом.

Лексические ошибки: перевод новых слов, перевод слов, несущих основную смысловую нагрузку, в тех случаях, когда неправильный перевод ведет к искажению смысла;

— грамматические ошибки по характеру последующей коррекции следует различать двух родов: коррекция предикативности основного высказывания и коррекция предикативности периферийных синтагм. Грамматические ошибки на согласование

⁵ Заметим, что здесь требуется большая работа по розыску реалии. После соответствующей обработки словосочетание вносится в словарь и далее его перевод затруднений не вызывает.

и пр. не исправляются (кроме тех случаев, когда они существенно влияют на правильность передачи смысла, например, ошибки в числе).

Рассмотрим теперь представления о сложности фраз и длине текста, с целью правильного учета этих параметров при определении качества перевода. Предварительно заметим, что набор оценок и шкалы не следует делать слишком дробными или содержащими слишком большой набор характеристик, так как это затруднило бы оценку.

Машинный перевод отдельно взятых фраз не имеет практической ценности, если при этом фразы подвергаются редактированию или искусственному сокращению длины, как это имеет место в ряде экспериментальных алгоритмов МП. Имеет смысл машинный перевод текстов, и качество МП мы будем оценивать по переводу текстов. Определим, что представляет собой текст. Исходя из средней длины английского предложения (фразы) в 20 слов, будем считать, что текст образуется тогда, когда имеет место совокупность фраз не менее чем 5 (из условия объема кратковременной памяти человека). Будем, следовательно, считать текстом отрезок минимум в пять фраз, каждая из которых состоит из 20 слов, или текст объемом в 100 слов. Это налагает определенные жесткие требования на условия оценки качества.

Фразы длиной более 20 слов будем считать сложными, так как в них вероятны придаточные предложения, конструкции вторичной предикацией и прочие сложные периферийные синтагмы. Фразы объемом до 20 слов будем считать простыми.

Будем определять сложность или простоту текстов в зависимости от процентного соотношения в нем простых и сложных фраз. Если текст содержит 50% или более простых фраз, будем считать его простым, если текст имеет сложных фраз более чем 51%, то сложным. Будем считать, что текст переведен правильно и понятно, если 50% или более фраз в нем переведены правильно и понятно.

Исходя из введенных определений, можно установить меру качества перевода отдельно для простых и сложных фраз.

Предварительно уточним классификацию текстов. Текст минимального объема будем называть малым: это тексты объемом от 100 до 400 слов. Тексты в 400—500 слов будем называть средними по объему (100×5 по объему оперативной памяти). Тексты свыше 500 слов будем называть большими по объему.

Для оценки качества перевода предложений, фраз и текстов, будем применять трехбалльную систему из трех оценок: хорошо, удовлетворительно и плохо.

Оценка качества перевода фраз производится отдельно для фраз длиной до и свыше 20 слов. Простые фразы (длиной до 20 слов) оцениваются следующим образом:

«Хорошо» — во фразе не более двух лексических ошибок или одной грамматической (периферийная синтагма).

«Удовлетворительно» — во фразе не более трех лексических ошибок или две лексических + одна грамматическая (периферийная синтагма);

«Плохо» — во фразе более трех лексических ошибок или две грамматических (периферийная синтагма) или одна грамматическая (основная предикативная синтагма).

Фразы длиной более 20 слов, не имеющие придаточных предложений, оцениваются следующим образом:

«Хорошо» — те же условия;

«Удовлетворительно» — допускается еще одна лексическая ошибка;

«Плохо» — те же условия.

Фразы длиной более 20 слов, но с придаточными предложениями, оцениваются по совокупности оценок для предложений, составляющих фразу, по правилам сложения оценок для фраз.

Отдельно производится оценка перевода каждого простого предложения, образующего сложное, и затем общая оценка перевода фразы выводится как среднее арифметическое оценок перевода составляющих предложений, с той поправкой, что переводу главного предложения приписывается удвоенная оценка. Так, если фраза состоит из главного и придаточного, и если главное переведено «плохо», а придаточное «хорошо», то арифметическое среднее следует вычислять из следующей суммы оценок: «плохо» + «плохо» + «хорошо». При этом округления в случае дробных результатов осуществляются в сторону ухудшенной оценки.

Приведем пример оценки качества перевода фразы с числом слов более 20.

Английская фраза

Перевод

Оценка

Moscow: the President of USSR Supreme Soviet, said here yesterday that «in recent years the relations between the USSR and young national states have become even more active and closer».

Москва: Президент Советского Союза Верховный Совет сказал здесь вчера, что «в течение последних несколько лет отношения между Советским Союзом и молодыми национальными государствами стали даже более активным и более тесным».

Перевод главного предложения «хорошо», оценка удваивается, перевод придаточного «хорошо», суммарная «хорошо».

Throughout these years the Soviet Union has shown limitless patience, understanding and flexibility to give every practical help to Indian progressive development and every moral and diplomatic support to the line of nonaligned independence inaugurated by Nehru.

Все время этих годов Советский Союз показал безграничному терпению, взаимопонимание и гибкость оказать каждую практическую помощь в индийском прогрессивном развитии, и каждой моральной и дипломатической поддержки к линии неприсоединившейся независимости торжественно открытой Nehru.

Суммарная оценка «удовлетворительно».

Английская фраза	Перевод	Оценка
While the Western powers sought to make India dependent on charity, it was the Soviet Union which led the way in the construction of a giant steelworks as an indispensable foundation for economic independence, and it was only as a result of this that the western powers were shamed into subsequently also building steelworks.	В то время как западные власти попытались сделать индийского иждивенца на charity, этом был Советским Союзом который приводимый путь к строительству гигантских steelworks для экономического независимости, и этого был только в результате это что западные власти shamed в впоследствии также строительства steelworks.	Перевод первого придаточного «удовлетворительно», главного — «плохо», оценка удваивается, второго придаточного «плохо». Суммарная оценка «плохо».

Поскольку текст состоит из фраз, можно вывести общую оценку для машинного перевода текста в целом.

Очевидно, что работа редактора в редактировании чернового перевода с целью превращения его в рабочий состоит из двух основных этапов. Первый этап заключается в устранении лексических и грамматических ошибок указанного выше типа. Для работы этого этапа редактор должен располагать как английским, так и русским текстом и использовать при работе оба текста. В этом отношении работа редактора машинного перевода равноценна работе редактора по редактированию обычного перевода, выполненного человеком. Ясно, что трудоемкость этого редакторского труда зависит от качества МП в определенном нами выше смысле.

Второй этап состоит в редактировании полученного правильного по смыслу текста уже только в морфологической части, а также во внесении в него стилистических изменений, если это необходимо.

В приложении содержится образец перевода, полученного полностью на ЭВМ и затем отредактированного девятью высококвалифицированными редакторами ВЦП. Текст взят из бюллетеня ЮНИСИСТ. Машинный перевод текста относится к категории «удовлетворительно». Обращает на себя внимание большое количество вариантов, которые в подавляющем большинстве случаев могут быть отнесены к так называемой «вкусовой» правке. Значительны также отклонения в переводе терминологии. Так, наименование организации в начальной строке текста исправлено всеми редакторами по-разному, в результате название European Co-ordination Centre for Research and Documentation in the Social Sciences (Vienna Centre) переведено:

— Европейский координационный центр исследований и документации в области общественных наук (Вена);

— Европейский центр по координации научных исследований и документации в области общественных наук (Венский центр);

— Европейский центр по координации исследований и документации в области общественных наук (Венский центр);

— Европейский координационный центр по исследованиям и документации в общественных науках (Вена, Центр);

— находящийся в Вене Европейский координационный центр по исследованиям и документации в области общественных наук;

— Европейский координационный центр по вопросам исследований и документации в общественных науках (Вена) и т. п., практически нет двух совпавших наименований.

Интересно отметить, что и в техническом тексте, где, как казалось бы, меньше возможностей для осуществления «произвола» переводчика, число вариантов перевода может быть не меньшим. Бюллетень по русско-английскому переводу, распространяемый Американской ассоциацией переводчиков, имеет регулярный раздел, содержащий различные варианты перевода коротких текстов в одно—два предложения или разные варианты русско-английского и англо-русского перевода научно-технической терминологии (RETX 1980).

6.2. Роль и место редактора в редактировании МП

Кооперация человека с ЭВМ в процессе перевода рассматривалась достаточно давно. В 1967 г. М. Мастерман была опубликована некоторая общая концепция МП с помощью человека (man-aided computer translation). На вход машины подается размеченная синтагма, в которой выделены главные члены и указан некоторый минимум сведений синтаксического характера о входящих в синтагму словах. Получив схему синтагмы, машина запрашивает у оператора более подробную контекстную информацию, которой последний располагает благодаря знакомству с текстом в целом. Стратегия взаимодействия состоит в том, чтобы любой ценой избежать постредактирования, допуская максимальное предредактирование входного текста; последнее выполняется в технике вопросов и ответов исключительно на входном языке. Главный аргумент против постредактирования состоит в том, что оно бывает либо чисто формальным, как, например, обеспечение согласования по роду, и тогда оно вполне может быть передано машине, либо имеет характер интуитивной и даже творческой деятельности, и тогда оно невозможно без постоянного обращения к входному тексту, а в этом случае постредактор может с таким же успехом сам переводить текст (Мастерман 1971).

Вопрос о редактировании довольно важен, когда мы рассматриваем практическое использование машинного перевода. Нам представляется, что экономическая целесообразность МП будет существовать, если он будет использован по схеме в аналогии с обработкой перевода, выполненного человеком (см. рис. 16).

При такой схеме некоторое увеличение труда редактора при редактировании машинного перевода (этот труд будет прогрес-

сивно уменьшаться по мере совершенствования алгоритма перевода и словарей на текстах данной тематики) вполне компенсируется более быстрым получением МП по сравнению с человеческим на первой стадии. Нет также нужды в сколько-нибудь существенном предредактировании.

Ввиду того, что система собственно автоматического перевода в модели МПС дает достаточно высокое качество перевода, роль интерредактора сводится к анализу и переводу новых слов, число которых невелико. Постредактор на первых порах затрачивает

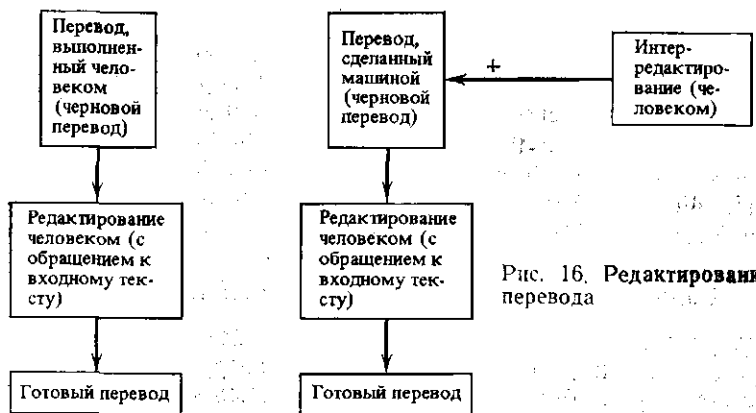


Рис. 16. Редактирование перевода

значительное количество труда на редактирование полностью готового продукта. В связи с этим место редакторов в системе МП определяется следующим порядком следования этапов:

— подготовка текста (определение необходимого качества перевода, разметка, перфорация и контроль отперфорированного материала);

— ввод текста в ЭВМ, поиск слова в словаре, выявление и анализ новых слов, формирование информации, необходимой для МП;

— интерредактирование;

— автоматический перевод текстов;

— постредактирование;

— корректировка словаря и алгоритма МП.

На первом этапе текст переносится на промежуточный носитель и вводится в ЭВМ, где производится автоматическое выделение слов, сравнение их со словарем автоматического перевода, фиксация слов, отсутствующих в словаре, и присвоение словам грамматических признаков данного этапа. После этого каждое новое слово с грамматическими признаками и окружающими его словами выдается на экран дисплея интерредактору. Интерредактор подтверждает (или исправляет) правильность грамматической информации данного слова, а также может производить другие операции (исправлять ошибки ввода, приписывать новый переводной эквивалент и пр.).

Постредактор производит пофразовый контроль автоматического перевода. Когда перевод фразы не удовлетворяет постредактора, он редактирует этот перевод с помощью дисплея и при необходимости отправляет исходную фразу и ее перевод в специальное запоминающее устройство для последующего анализа с целью совершенствования алгоритма. Возникающие при переводе ошибки можно условно разделить на две группы: ошибки, для исправления которых редактору необходима фраза исходного текста и возможные переводные эквиваленты отдельных слов из этой фразы, и ошибки, исправляемые редактором без обращения к исходному тексту. Для исправления ошибок первого типа система имеет специальный словарь автоматического редактирования, в котором по результатам редактирования текста происходит накапливание слов и их переводных эквивалентов, которых нет в словаре автоматического перевода. По запросу редактора система выдает фразу исходного текста и имеющиеся в словаре редактирования переводные эквиваленты запрашиваемых слов.

Таким образом, словарь автоматического перевода (САП) — исходный словарь, и словарь автоматического редактирования (САР) — промежуточный словарь — функционально разные вещи в системе МП. Существенное отличие между САР и САП состоит в следующем:

— форма записи и объем информации при слове в САП зависят от алгоритма перевода. Форма записи и информации в САР аналогичны записи в обычном словаре и не зависят от алгоритма;

— для хорошо организованной системы перевода объем САР невелик по сравнению с объемом САП;

САР содержит только те слова и переводные эквиваленты, которые отсутствуют в САП;

— пополнение и изменение САП производится лингвистом на основании анализа результатов редактирования и пополнения САР. САП пополняется постредактором (специалистом, знакомым с предметом перевода).

Введение этапов полуавтоматической коррекции промежуточных и окончательных результатов перевода существенно повысит качество перевода без значительного увеличения общего времени перевода. Полуавтоматический сбор статистических данных о частоте и характере ошибок перевода ускорит работу лингвистов по корректировке алгоритма и словаря автоматического перевода.

Интерредактор должен быть лингвистом, знать входной язык и оперативно приписывать новым словам переводы, а также некоторый набор грамматической информации. Постредактор должен быть лингвистом и одновременно специалистом в данной области, что необходимо для уверенной ориентации в предметной области, подлежащей переводу; кроме того, ему необходимы знания принципов устройства системы МП.

6.3. Техничко-экономическая эффективность МП

Технические преимущества машинного перевода связаны с тем, что он обеспечивает правильную и единообразную передачу технической и научной терминологии, выгодно отличаясь этим от человека-переводчика и редактора, которые имеют склонность варьировать даже лексические переводные эквиваленты. В таблице 21 суммированы качества МП, не связанные со стоимостью.

Из данных этой таблицы следует особо отметить точность перевода терминологии, которой зарубежные фирмы придают большое значение, и высокую скорость выполнения переводов (Sereda, Levesque, 1977).

Факторы, определяющие стоимость немашинного, человеческого перевода, следующие: а) сам перевод, в письменной форме или надиктованный на магнитофон; б) печатание переведенного текста; в) редактирование, самим переводчиком и/или редактором; печатание отредактированного текста. Факторы а) и в) имплицитно включают расходы на профессиональную подготовку переводчиков и редакторов. Недостатки в такой подготовке должны быть скомпенсированы временем (и деньгами), затраченным на обращение к специальным словарям или к терминологам во время процесса перевода и редактирования.

Стоимость машинного перевода может быть удобно разделена на затраты на разработку и затраты на эксплуатацию.

Затраты на разработку: а) создание программного обеспечения системы; б) создание двуязычного словаря данного подязыка. Затраты на эксплуатацию: в) пре-редактирование, осуществляется не специалистами; г) ввод текста с помощью перфорации, с магнитной ленты или оптического считывающего устройства; или перевод текста, записанного на магнитную ленту в специальный формат с помощью программы—интерфейса; д) перевод и распечатка текста компьютером; е) пост-редактирование, осуществляется лингвистом; ж) перепечатывание (или фотопечать, или распечатка на ЭВМ) отредактированного текста перевода.

На материалах, переведенных системой СИСТРАН с английского языка на французский, была определена стоимость машинного перевода, которая затем была сравнена со стоимостью немашинного человеческого перевода. Машина перевела тексты общим объемом 10 000 слов (рефераты, статьи из научных журналов, внутренние документы Европейского экономического сообщества в области пищевой промышленности и технологии). Определялась стоимость перевода одного слова, для системы СИСТРАН и для переводов, выполненных переводчиками как штатными, так и нештатными. Результаты показывают:

— неотредактированный МП значительно дешевле, чем любой человеческий перевод;

— отредактированный МП дешевле, чем отредактированный человеческий перевод, производимый службой переводов ЕЭС;

Таблица 21

Критерии	Немашинный («человеческий») перевод	Машинный перевод
Единообразие перевода терминологии	Даже в наилучших организациях невозможно добиться полного единообразия: 1) переводчики различаются по стилю; 2) главной проблемой является отсутствие стандартной терминологии; 3) некоторые редакторы имеют свои предпочтения	Стопроцентное единообразие
Техническая точность	Если переводчик не является одновременно опытным специалистом в данной технической области, достижение технической точности в переводе весьма сомнительно. Между редактором и переводчиком должен постоянно поддерживаться тесный контакт. Машина и человек в этом аспекте равноценны	Правильное кодирование с полной информацией значительно увеличивает уровень технической точности. Консультации между редактором и ЭВМ имеют место в случаях крайней неоднозначности
Грамматика	Грамматические ошибки всегда возможны	Правильное кодирование устраняет грамматические ошибки
Стиль	Переводчик превосходит компьютер в отношении стиля. Однако в технических переводах стиль не является главной характеристикой Примечание: эксперимент показал, что в ряде случаев МП более приемлем для специалистов, чем не машинный перевод.	Правильное кодирование обеспечивает приемлемый перевод, в общем хорошо понятный специалистам
Скорость перевода	Хорошая норма работы человека-переводчика — в среднем 1200—1500 слов в день	Теоретически: 300 000 слов в час, однако при загруженности ЭВМ другими задачами на МП отводится 4 часа в день; таким образом, 1 200 000 слов в день
Редактирование	В среднем от 1000 до 5000 слов в день, кроме случаев, когда переводчик является одновременно высококвалифицированным специалистом и его работу редактировать не нужно. Редактор должен обладать высокой квалификацией, работать внимательно; процесс медленный	Возможен объем работы в 20 000—25 000 слов в день, при совершенствовании машинного словаря: 1) нет необходимости сверять терминологию, нужны только мелкие поправки; 2) высокая скорость

— отредактированный МП дороже, чем неотредактированный перевод, сделанный штатными переводчиками;

— отредактированный МП уравнивается по стоимости с неотредактированным переводом штатных переводчиков, если тексты представляются в машинночитаемой форме, что избавляет от необходимости перфорировать.

Оценка также показывает, что значительная экономия может быть достигнута в будущем за счет:

— улучшения качества начального продукта МП посредством обратной связи и снижения редакторских требований;

— совершенствования механизмов ввода и вывода.

Затраты на создание систем машинного перевода и соответствующих словарей могут окупиться за сравнительно короткое время, если переводятся большие объемы текстов. Например, стоимость создания англо-французской системы СИСТРАН для ЭЭС может быть полностью возмещена за один год, если переводить этой системой весь объем работ ЭЭС в этой области, т. е. около 20 млн слов в год (Chaumier, 1977).

Технико-экономическая эффективность системы МП в условиях СССР в настоящее время может быть определена лишь приблизительно, в связи с тем, что точная стоимость переводческого обслуживания может быть рассчитана лишь для достаточно конкретных информационных потоков, ориентированных на отдельных потребителей. Примерный расчет технико-экономической эффективности можно базировать на определении времени окупаемости системы. Время окупаемости можно вычислить по формуле:

$$T = \frac{B}{W_{\text{год}}},$$

где T — время окупаемости в годах, B — стоимость проектирования, создания и опытной эксплуатации системы, $W_{\text{год}}$ — годовая дополнительная прибыль.

Годовая дополнительная прибыль может быть определена как разность

$$W_{\text{год}} = W_{\text{м}} - W_{\text{р}},$$

где $W_{\text{м}}$ — общая годовая прибыль от переводов, произведенных машинным способом, $W_{\text{р}}$ — общая годовая прибыль, которая может быть получена от переводов того же объема при «ручном» исполнении.

Расчет, произведенный для системы АМПАР, дает срок окупаемости в 2,5 года, который можно признать удовлетворительным.

6.4. Машинный перевод в системе научно-технической информации

Машинный перевод в свете изложенного выше следует рассматривать как грубый продукт, «черновой перевод», получаемый достаточно быстро. Он обладает положительными свойствами в виде точности перевода терминологии, доступности в виде сигнальной информации и т. п. Очевидным и не требующим особых доказательств является также и то, что для перевода спецификаций, инвентарных списков, стандартизованных описаний, выказываний в рамках ограниченных и специализированных подязыков системы МП, особенно построенные на принципах переводных двуязычных соответствий, промышленно пригодны уже сейчас. На информационном рынке все более проявляется растущая потребность в переводах, в их объемах и особенно в ускорении сроков. Однако если учесть, что современный МП требует достаточно больших затрат и участия крупных коллективов, оснащенных большим количеством современной техники, представляется совершенно необходимым специально изучить роль и место МП в современных системах научно-технической информации.

Роль научно-технического перевода в системе научно-технических коммуникаций современного мира вообще изучена крайне слабо. Как правило, констатируется, что научно-технические материалы, требующие перевода с одного языка на другой, переводятся, организуется хранение и распространение переводов, существуют центры переводов («клирингхаузы»), известны расходы на переводческие службы и т. п. Однако когда мы рассматриваем реальные расходы на внедрение действующих систем МП, требуется на самом деле более тщательный анализ и учет множества иных, обычно не учитываемых факторов. Не претендуя на подробный анализ и изучение затрагиваемой проблемы, назовем некоторые существенные факторы, определяющие место МП в системе научно-технической информации.

Ценность МП как сигнальной информации, безусловно, определяется языковой ситуацией в стране. Комиссия по машинному переводу американского национального научного фонда, изучая положение с МП в США, в первую очередь обратила внимание на то, с каких и на какие языки переводится научная литература. Большое, если не преобладающее количество научно-технических публикаций в мире делается на английском языке, поэтому для англоязычных стран перевод этой литературы не нужен. Специальные исследования на эту тему, проведенные в странах Западной Европы, показали, что переводы с английского языка в этих странах не актуальны, поскольку английский язык достаточно распространен среди ученых и специалистов. В СССР же, например, переводы с английского языка стоят по объему на первом месте. Актуальность перевода с какого-либо языка определяется количеством научной литературы на этом

языке, уровнем науки и техники, достигнутым данной страной, наличием переводчиков с данного языка и стоимостью их услуг. Так, для англоязычных стран большой интерес представляет литература на русском языке: однако Запад располагает достаточным количеством переводчиков с этого языка, сравнительно дешево продающих свои услуги. Переводчиков, например, с арабского или японского языков гораздо меньше, однако здесь вступает в силу другой фактор, а именно, с какого языка целесообразнее переводить.

Фактором, влияющим на целесообразность применения МП, является стабильность информационных потоков. Стабильность можно считать гарантированной при автоматизации перевода научно-технической документации, поскольку выпуск продукции, в том числе и новой, даже при смене номенклатуры, можно считать явлением постоянным. Стабилен также поток других информационных материалов, например научных отчетов, диссертаций, патентов, особенно при вынешних тенденциях к расширению информационного обмена. Однако при решении вопроса об МП требуется установить стабильность достаточно узкого информационного потока, например, патентов, диссертаций, отчетов по узкой тематике. Это связано с тем, что МП тем эффективнее, чем точнее его настройка на данный подъязык, и это безусловно касается как лексики, так и всего лингвистического обеспечения системы.

Машинный перевод как сигнальная информация неминуемо заставляет пересмотреть существующую систему снабжения разработчиков научно-техническими переводами. Потребность в переводе того или иного научно-технического материала обычно определяет бюро или отдел научно-технической информации (НТИ) на данном предприятии. При этом ясно, что один информатор БНТИ суммирует информационные потребности многих десятков разработчиков, и иногда не лучшим образом. Между разработчиками и литературой, таким образом, существует физический барьер в виде суммирования и неизбежного усреднения их потребностей. Второй барьер (в этом же месте) возникает из-за недостаточного знания информатором языка и предмета, о котором идет речь. Такая же ситуация имеет место и в том случае, если разработчик заказывает перевод по реферату. Здесь качество реферата, отражающее познания и компетентность референта в предмете, знание им языка, наконец, особенности и традиции реферативного издания также затрудняют знакомство разработчиков с информацией, отраженной в публикации. Исследования по использованию научно-технических переводов постоянно показывают, что значительная часть переводов, тщательно отработанных в языковом отношении, выполненных по оплаченным заказам потребителей, в том числе организаций и учреждений, используется разработчиками именно как сигнальная информация: бегло ознакомившись с переводом и поняв, «о чем идет речь», разработчик откладывает его на полку и

далее к нему не обращается. Это означает, что значительные расходы на перевод, редактирование, перепечатку, размножение, хранение и пересылку переводов оказываются затраченными напрасно. Если предположить, что разработчик, придя в библиотеку, получает вместе с новым иностранным журналом его «сигнальный перевод», то становятся ненужными посреднические услуги информатора и исчезнет дорогостоящее изготовление ненужных переводов. Тщательно изготавливаться будет только действительно потребный и многократно используемый перевод. Можно предположить, что при этом расходы на оперативный МП соответствующих текстов, выполняемый на месте или централизованно в достаточном количестве экземпляров, при решении проблемы ввода, например, с помощью магнитных лент и при неизбежном удешевлении машинного времени и неизбежном же дорожании человеческого труда переводчиков, будут постепенно приходить в некоторое равновесие с расходами на «человеческий» перевод, а в некотором будущем (не очень отдаленном) безусловно меньше.

Нет необходимости доказывать, что здесь действительно требуется тщательный анализ путей и средств появления, изготовления, распространения и экономических характеристик использования переводной научно-технической литературы и документации. С этой новой точки зрения можно будет вновь обратиться к старой проблеме реферирования. Хорошо известно, что два референта по-разному реферировуют один и тот же материал; если будет принята сигнальная информация МП в качестве некоторого реферата, то тогда «человеческий» реферат или аннотацию можно будет строить по иным и теперь уже совершенно четким правилам, в предположении, например, что МП дополнит то, чего может не быть в реферате.

Решение указанных проблем научно-технической информации будет тем эффективнее, чем больше реальных сфер информации будет им охвачено. Условия социалистического планового хозяйства, планомерного развития нашей науки и техники, решающие преимущества нашей системы над капиталистической позволяют строить сеть научно-технической информации максимально эффективным образом в масштабе всей страны. Увеличение капиталовложений в одном месте этой системы может и должно оцениваться с позиций общих характеристик этой системы. Так, МП безусловно эффективен как часть интегрированных информационных систем, содержащих, например, технические описания (например «Евронет»). Скорость и полнота информационного обслуживания, отказ от излишних «человеческих» переводов повысят вклад в общую рентабельность системы. Если в настоящее время многие учреждения и предприятия преодолевают языковые барьеры собственными кустарными способами, например, расширением штатов переводческих бюро, зачастую маскируемых, или отказом от участия в системе координации и регистрации переводов (быстрее вновь сделать пере-

вод, чем получить его из общегосударственного или ведомственного хранилища), то развитие информационной сети с инкорпорированным в ней МП сделает невыгодными такие частные решения вопросов.

Эти самые общие соображения, как нам кажется, показывают необходимость и своевременность рассмотрения проблемы о месте научно-технического перевода в системе научно-технической информации. Следует сказать, что это место определяет и требования к системе МП. Без учета роли МП в информационной сети требования к системе МП носят чрезвычайно общий характер: надо построить автомат, умеющий переводить с такого-то языка на такой-то. Между тем функциональная специализация помогает более точно задать технологические параметры системы и обеспечить тем самым достижение лучших результатов. Пронлюстрировать эту мысль можно на примере робототехники.

Работы тридцатых годов нашего столетия были некоторыми универсальными копиями человека. Считалось важным, чтобы робот умел ходить, говорить, воспринимать и исполнять простые команды, имел руки, ноги, внешнее сходство с человеком и пр. Этот универсализм сочетался с чрезвычайно малой, почти нулевой, пригодностью первых роботов. Современный робот узко специализирован. Он имеет четко очерченный круг задач — ставить колеса на автомобиль, красить детали в малярном цеху и др. Хорошо известна в настоящее время высокая эффективность промышленных роботов и их быстрое распространение в развитых странах. В то же время робот не потерял и своего моделирующего значения, он остается роботом, т. е. моделью многих и разных человеческих функций, он не есть станок с программным управлением. В этом смысле сама идея робота осталась, но ее реализация — это робот — специалист, а не робот, сходство которого с человеком ограничивается слишком внешними или несущественными для конкретной задачи чертами.

Ранние системы МП можно уподобить универсальным роботам начального периода робототехники. Проявляется определенная тенденция к специализации систем, при сохранении универсальной «способности переводить» эффективность и целесообразность робота определяется его местом в информационной системе. Подобно тому, как промышленный робот немислим в отсталом производстве, эффективная система МП может действовать лишь в модернизированной, пересмотренной с точки зрения современных требований сети научно-технической информации.

6.5. Задачи и перспективы машинного перевода

Развитие теории и практики МП позволяет в настоящее время рассматривать эту проблему как одну из проблем практического информационного обслуживания. Фундаментальные исследования в проблематике МП, разработанные и опробованные ал-

горитмы, словари, методические и технологические подходы к разным аспектам этой проблемы и практические результаты в виде действующих систем позволяют осуществлять практическое применение машинного перевода в научно-технической информации.

Однако в то же время нельзя утверждать, что все вопросы, касающиеся технологии построения действующих систем МП, решены и что такая технология установилась. Дело здесь не только в том, что каждая пара языков имеет свои проблемы, существенно определяющие подход к конструированию системы МП, особенно если мы рассматриваем уровень переводных соответствий, который остается при любом принципиальном подходе. Сравнительно неизученным остается круг проблем, связанных с переходом от общей задачи преобразования смысла при переходе от одного языка к другому к целевой систематизации фактов структуры языка в их текстовом проявлении. Модель МПС имеет свои ограничения, которые, возможно, могут быть сняты при реализации иных подходов в рамках принципиального использования переводных соответствий. Видимо, должен быть разработан инвентарь подобных рабочих моделей, занимающих как бы промежуточное место между тотальными теориями человеческой языковой деятельности и подлежащими систематизации фактами языка и текстов.

Мы придерживаемся мнения, что такие модели следует создавать, изучая работу человека по переводу, моделируя собственно переводческую деятельность человека. В переводческом сопоставлении языки изучаются особо, и эту специфику не только нельзя игнорировать, но наоборот, именно ее следует использовать и на нее опираться в построении модели МП. Ясно, однако, и то, что переводческая деятельность человека тоже многообразна, и могут существовать разные представления о том, как именно в мозгу человека осуществляется перевод. Из этого следует, что модели переводных соответствий тоже могут быть построены по-разному, с преимущественным вниманием к разным сторонам процесса.

С филологической точки зрения перевод есть авторская операция, которую человек-переводчик производит над текстом. Осознав характер задачи, действуя с определенной мерой осведомленности, напряжением внимания и воли переводчик достигает коммуникативного эффекта, предполагаемого автором оригинала. Переводчик формирует новый текст, который в идеале обладает всеми особенностями оригинала. Предварительное чтение переводчиком оригинала обеспечивает возможность приложения определенного «стилевого усилия», поскольку всякий раз текст, составленный человеком, содержит определенные ошибки против принятого норматива. Эти ошибки могут быть следствием разнообразных причин, но главным образом они суть следствие свободы воли говорящего, которой он пользуется для

достижения нового эффекта, также прилагая стилевое усиление. Из-за нарушений норматива любая формальная грамматика не будет полна и не будет давать полного покрытия текстов. Для читателя текста перевода, в такой же степени, как и для читателя текста оригинала, важно психолингвистическое осознание текста.

Машинный перевод есть перевод без авторства переводчика. В той мере, в какой создание и последующее пополнение алгоритмов и словарей есть работа коллектива, МП есть обезличенный коллективный перевод. МП базируется на чисто переводных, языковых соответствиях, никакого предварительного чтения текста система МП производить не может, поэтому МП исключает текстологический анализ. Передача смысла в МП достигается особым знаковым преобразованием, так как МП есть определенная операция над языком при условии идентичности содержания знака. Он осуществляет механические синонимические замещения на основе регулярных межъязыковых корреспонденций соотнесения. Эти регулярные корреспонденции не извлекаются автоматически из хранилищ или устройств, подобных словарям. Невозможно моделировать коммуникативную функцию естественного языка, если рассматривать его только как кодовую систему; переводные соответствия есть элемент динамики пользования языком, и удовлетворительно моделирующая перевод система не может не рассматривать переводные соответствия как всеобъемлющие, многоуровневые, но каждый раз новые образования. Должны быть формализованы не сами переводные соответствия как таковые, но закономерности, по которым они возникают в текстах и с помощью которых они могут быть обнаружены в деталях, важных и нужных для перевода. Такая формализация удается с помощью моделирования действий переводчика, при котором особое внимание уделяется семантико-синтаксическому анализу в терминах, близких к традиционному понятным всем пользователям языком.

Никаких стиливых усилий машина-переводчик не производит. Задача приложения этих усилий возлагается на читателя машинного перевода или редактора. Механические операции, которые производит ЭВМ над текстом, носят сложный характер: система осуществляет максимально полные рекомбинации знаков и комплексную переработку массовой информации в языке. Производится опознание, различение и перевод между разными категориями двух языков, разведение синонимов и омонимов, процессное развертывание членов предложения, в результате чего обеспечивается инвариантность существенных семантических признаков значения в передаче его на другой язык.

Применительно к задаче МП возможности машины не ограничиваются соответствиями на уровне графем и слов. Устанавливаемые по тексту с помощью аппарата трансляционной грамматики и иерархии алгоритмов переводные соответствия на уровне слов, словосочетаний и предложений обеспечивают передачу

содержания на уровне текста. При этом чрезвычайно важен системный подход к решению проблемы. Каждая отдельная «переводческая задача» может быть решена последовательностью алгоритмизуемых действий при достаточно четко описанной исходной информации. Однако совокупность частных решений и рецептов не образует систему. Система МП должна в первую очередь учитывать и оптимизировать разрешение наиболее частых и массовых лингвистических ситуаций в переводе, будучи готовой при этом к переходу на более тонкие методы анализа.

Исходя из концепции системной организации, особое внимание в МП уделяется словарю. Поскольку словарь для МП есть одновременно машинный словарь, который имеет аналоги в других системах автоматической обработки информации на естественном языке, возникает возможность сделать обобщения и говорить об общих свойствах машинных словарей. Эти общие свойства рассматриваются в противопоставлении свойствам обычных словарей, но не в абстрактном смысле (при котором такие противоречия могут стираться подобно тому, как стирается различие между человеком и мыслящей машиной), а в представлениях инженерной лингвистики — прикладной лингвистики, направленной на решение технических задач.

В вычислительной лексикографии особенно проявляется диалектическая зависимость между количеством и качеством применительно к описанию лексического уровня. Словарь небольшого подъязыка может вообще не содержать многозначных слов и удовлетворительно работать на текстах. Контекстологический словарь, как новый тип словаря, имеющий значение не только для МП, но и для лексикографии в целом, позволяет точно описать контекстные зависимости с количественной и качественной стороны для широкого языкового поля. Эффективность и экономичность словаря в машинном использовании служат достаточным подтверждением работоспособности идеи. Теория контекстной детерминации и ее частная теория — метод детерминант — позволяют очертить границы эффективности метода и перечислить явления, для анализа которых требуется использование иных или более точных подходов. Однако обеспечиваемое этими теориями качество перевода делает его готовым к редактированию.

Модель МПС или эквивалентная ей модель являются необходимым этапом работы всякий раз, как только будет ставиться задача на перевод, а не на другие преобразования текста (автоматическое реферирование, индексирование и т. п.). Смысл текста в разной детализации может быть, очевидно, передан разными путями, но для осуществления перевода в собственном смысле установление переводных соответствий представляется неизбежным.

Можно утверждать, что МП есть часть искусственного интеллекта. Как известно, довольно широко используемое понятие «искусственный интеллект» охватывает разнообразные виды че-

ловеческой умственной деятельности, однако до настоящего времени они достаточно просты и ни один из них не может сравниться по сложности и учету разнообразных факторов с переводом между двумя естественными языками. Между тем проблематика такой рекомбинации знаков, что между высказываниями на двух разных естественных языках при этом сохраняется инвариантность смысла, без сомнения относится к сфере исследований по искусственному интеллекту.

По каким путям может развиваться модель МПС и машинный перевод в целом?

Представляется, что путь повышения эффективности МП на базе переводных соответствий заключается в специализации алгоритмов и систем МП по видам текстов. Классификация информатических текстов и текстов массовой коммуникации может отражать важные для МП признаки и определенным образом строиться на этих признаках. Очевидно, что тексты разного характера будут отличаться друг от друга: так, реферат определенно имеет другой стиль и другой набор грамматических конструкций, чем патентные тексты, полные тексты научно-технических журналов явно отличаются от стиля монографий и т. п. Классификация текстов по «видам словесности» будет отражаться в соответствующей классификации систем МП, каждая из которых может быть ориентирована на ту или иную разновидность текста. Задав жанр перевода, можно ограничить исходный набор конструкций и тем самым повысить разрешающую способность алгоритмов. Специализация словарей позволит давать лучшее покрытие текстов. При такой классификации можно будет успешнее преодолевать рефлексии индивидуального стиля.

Второй параметр, классификация по которому позволит увеличить эффективность системы МП,— предметная классификация. Ясно, что четкое деление на тематические области даст возможность существенно снизить неоднозначность лексики, уменьшить число вариантов перевода грамматических конструкций, ограничить вариативность синтаксиса, дифференцировать семантические признаки. От крупных делений, таких как общая лексика и грамматика, научная лексика и пр., можно будет перейти к специализированному лингвистическому обеспечению конкретных языковых подсистем.

МП может действовать как в сфере массовой коммуникации, так и в области научно-технической информации (информатики). В последнем случае на вход систем МП поступают разнообразные информатические тексты, рассчитанные на разных потребителей. При переводе рефератов, аннотаций и пр. система МП будет все точнее настраиваться на перевод ключевых слов. В этой ситуации можно усмотреть естественный переход от словарей систем МП к словарям слов-дескрипторов и к системам, построенным по дескрипторному признаку. В двуязычной и многоязычной ситуации вполне возможно сочетание действующих систем МП с информационными системами, использующими дву- или

многоязычные тезаурусы. Для обеспечения такого соединения необходима предварительная работа в указанных выше направлениях: предметной специализации и дифференциации систем МП по видам текстов.

Оба эти направления важны не только для МП. Нетрудно видеть, что они обеспечивают как бы две оси координат, в системе которых циркулируют информационные потоки. Так, информационные потребности, например, военного дела, могут быть четко описаны определенными предметными областями и типами документов, которые циркулируют в каждой из них. Могут быть также и области знания, в которых такой дифференциации достаточно строго провести не удастся. К таким областям, например, можно отнести сельское хозяйство или другие политематические области. Таким образом, здесь мы выходим к одной из главных проблем информатики: где, в каких видах текстов может циркулировать научно-техническая информация, и как тип этой информации соотносится с типом текста. В частном аспекте ответы на соответствующие вопросы помогают составлять машинные словари; в общем речь может идти о процедуре общего информационного обслуживания, о научном планировании информационных процессов.

Планирование информационных процессов давно известно в человеческом обществе применительно к некоторым областям деятельности. Так, процедура суда включает совершенно определенные информационные процессы, совершаемые в строгом порядке в соответствии с принятой формой. Всемирные правила почтовой связи вносят порядок в такой, казалось бы, стихийный и неуправляемый процесс, как связь каждого отдельного человека с любым другим человеком на земном шаре.

МП может участвовать в информационной процедуре, в свою очередь влияя на нее и определяя планирование информационного процесса. В качестве примера можно привести ситуацию с переводом в странах с узаконенным двуязычием (многоязычием) или в распространяющихся межнациональных объединениях, например, типа Совета Экономической взаимопомощи или Европейского экономического сообщества, где действуют определенные строгие правила изготовления, перевода и распространения официальных и научно-технических документов. В этой стратегии направления информационного потока должно быть найдено реальное место машинному переводу. Практически он уже начинает занимать такое место в регламентации и упорядочении межнационального общения.

С развитием межнационального языкового общения роль МП будет неуклонно повышаться, поскольку другое средство преодоления языковых барьеров, а именно создание международного языка, специалисты по информатике продолжают считать утопией. Перевод с помощью ЭВМ, который также получает все большее распространение, как можно себе представить, является лишь упрощенной разновидностью МП, и как таковой будет

ПРИЛОЖЕНИЕ 1

В приложении 1 дается описание нескольких действующих систем машинного перевода, используемых в информационном обслуживании специалистов переводами с иностранных языков.

1. Система машинного перевода «СИСТРАН»

26 мая 1978 г. в Нью-Йорке в ООН состоялась встреча специалистов по вопросам машинного перевода. Была представлена система МП с редактированным СИСТРАН, являющаяся собственностью компании Уордл транслейш оф Канада. Научным руководителем системы является д-р Петер Тома. Большим стимулом к развитию системы послужил полет «Аполлон—Союз», в ходе которого была реализована система англо-русского и русско-английского МП. В течение полета было переведено большое количество текстов общим объемом в десятки миллионов слов в обоих направлениях. Коммерческая эксплуатация системы началась с 1975 г., когда начала действовать система англо-французского и французско-английского перевода. Ее активно использует и финансирует компания Дженерал Моторс для перевода технических описаний автотракторной техники, в том числе и особенно военной.

Достоинства системы:

- получаемый текст может подвергаться редактированию легко и быстро;
- имеется возможность переводить большой объем научно-технических материалов;
- система «многоязычна»;
- инженерные решения достаточно простые (стандартные ЭВМ).

Одно из главных достоинств заключается в том, что при МП обеспечивается соблюдение последовательной, непротиворечивой терминологии, в то время как в человеческом переводе терминологической лексики часто возникает разноречивость. Это достоинство является весьма существенным для научно-технического перевода. Компания Дженерал Моторс считает его основным в выборе именно МП для перевода спецификаций (Sereda, Levesque, 1977).

Скорость перевода составляет 1 млн. слов в час за время работы центрального процессора. Скорость редактирования — 10 тыс. слов в час, так что удобно иметь пять параллельно работающих редакторов. Для превращения в информационную систему реального времени нужно просто увеличить количество редакторов. Стоимость по машинному времени составляет

развиваться в сторону соединения с МП. Потребности машинного перевода будут определенным образом влиять на правила письменной корреспонденции участников общения: достаточно вспомнить, например, что потребности ввода документов в дескрипторные информационные системы заставили ряд научных журналов требовать от авторов сопровождения своих статей набором ключевых слов. Эта тенденция будет сочетаться с практикой регламентации средств языкового выражения в разрабатываемых технику учреждениях. Оценка систем МП будет, таким образом, строиться в зависимости от того, с какими другими информационными системами будут сопрягаться МП и каковы будут требования этих информационных систем.

Оптимизация правил движения текстов в интегрированной с МП информационной системе неминуемо будет касаться правил составления словарей и грамматик. Будет продолжаться тенденция стандартизации деловой корреспонденции, в качестве некоторого внешнего нормирования общения. Однако расширение языкового общения в рамках информационных систем неминуемо приведет также к разработке и становлению новых или скорректированных языковых норм. В отсутствие информационной сети нормирование в значительной мере произвольно; при действующей сети оно естественным образом оптимизируется. Здесь мы подходим к проблеме языковой нормы в широком понимании.

Традиционная грамматика, как отражение языковой нормы, интуитивно учитывала все сферы речевого общения. В течение долгого времени она играла роль регулятора информационного потока: для того, чтобы общаться, нужно было соблюдать правила грамматики. С распространением автоматизированных систем обработки информации грамматика продолжает играть свою регулирующую роль, но сама грамматика при этом становится объектом механических межъязыковых преобразований. Здесь можно говорить о проблеме искусственного интеллекта как о проблеме изучения закономерностей и всего механизма, управляющего формализацией языка, и его основы — грамматики.

Наиболее сильное нормирующее воздействие на язык в век научно-технической революции может оказать именно автоматическая обработка языковых данных и ее потребности, в частности и особенно машинный перевод.

Машинный словарь в принципе может быть максимально полным, чего не может достигнуть никакой другой словарь. Поэтому машинные словари как полные, так и отдельные по отраслям знаний могут оказать на речевое общение и на лексику такое влияние, которого не могли бы оказать любые другие средства, находящиеся в распоряжении социальных институтов, связанных с информатикой и с проблемой речевого общения в целом. Развитие вычислительной лексикографии, как и МП в целом, невозможно без сочетания теоретических разработок с уже возможным сейчас удовлетворением информационных потребностей.

Таковы, по нашим представлениям, современные проблемы машинного перевода.

1,4 цента за слово при массиве 40—50 тыс. слов, что является весьма хорошим показателем.

Система обеспечивает правильный перевод диаграмм, графиков и прочих нетекстовых включений. В ЭВМ могут вводиться изменения, например, при изменениях стандарта на терминологию. Поправки редакторов также постепенно накапливаются и вводятся в словарь. Перевод осуществляется для следующих пар языков: англо-французский, русско-английский, англо-русский; в течение 1978 г. введены в действие системы англо-испанская, француско-английская, англо-арабская. Создается система перевода с китайского языка. Для этого языка ввод текста не представляет трудностей, так как осуществляется на китайском телеграфном коде, что заодно обеспечивает легкое распознавание омографов.

Отредактированный готовый к размножению текст перевода производится со скоростью свыше 10 000 слов в день. Это в пять раз быстрее, чем обычные нормы перевода научно-технического материала. Один кодировщик может в среднем внести в лексическую базу системы 150 новых единиц. Если перевод предназначен для печати, то достигается значительная экономия благодаря использованию устройств фотовывода. Обработка 1 000 000 слов в месяц требует менее 5% мощности машины типа ИБМ 370/168.

Наиболее разработана русско-английская система. 300 000 слов и оборотов находятся «он-лайн» с компьютером. Европейское экономическое сообщество использует ряд подсистем СИСТРАНа в Люксембурге на текстах сельскохозяйственной тематики. В Сан-Диего два сотрудника работают над англо-итальянской системой. Англо-французский банк лексических данных к концу текущего года достигнет 200 000 слов; в основном это технический словарь. В зоне, прилегающей к территории фирмы Джeneral Моторс, обеспечивается выдача словарных справок переводчикам в масштабе реального времени через компьютерно-информационную сеть с видеотерминалами.

В начале 80-х годов компания предполагает осуществить перевод в пределах всех пар языков стран Европейского экономического сообщества.

Области науки и техники, в которых делаются переводы:

— автомобильная и пищевая промышленность;

— вычислительная техника;

— связь;

— авионавтика, электроника, инженерное дело, административные вопросы и пр. Система контроля за качеством перевода позволяет вести тщательный учет вносимых изменений.

Экономичность системы обеспечивается сочетанием машинного и человеческого труда, уменьшением числа ошибок.

Представляет интерес экономическая эффективность СИСТРАНа. В 1978 г. Европейское экономическое сообщество

провело оценку экономической эффективности по материалам выполненных переводов. СИСТРАН оценивался в двух аспектах: а) скорость и стоимость перевода; б) качество перевода. Проверка осуществлялась на переводах с английского на французский язык текстов по пищевой промышленности и технологии. Массив текстов включал: 20 рефератов текстов по пищевой промышленности и технологии из реферативного журнала, объемом приблизительно 3000 слов; две статьи из научных журналов, объемом приблизительно 4000 слов; 4 документа сообщества, объемом около 4200 слов. Использованный словарь содержал примерно 25 тыс. слов (словоформ) и 4500 оборотов. Проверка производилась на машине ИБМ с пополнением словаря после первого прогона. Одновременно отдел переводов сообщества перевел эти же тексты вручную с редактированием и отредактировал полученные на машине тексты после первого прогона (Chaumier, 1977).

Рассмотрим кратко историю разработки и принципы лингвистического устройства СИСТРАНа. Судя по публикациям и библиографическим отсылкам, в начале разработки прототип системы СИСТРАН был связан с Джорджтаунским университетом (Тома, 1974). Один из первых вариантов системы назывался СЕРНА (система перевода с русского на английский) и 8 июня 1959 г. демонстрировался в Пентагоне (Тома, 1977). По крайней мере с 1970 г. система использовалась для технических переводов с русского языка на английский в отделе иностранной техники ВВС США на авиабазе Райт-Паттерсон в Сан-Диего, Калифорния. Нынешний вариант д-р П. Тома называет третьим поколением от системы СЕРНА.

Лингвистическое обеспечение системы описано в чрезвычайно общих выражениях. Следует отметить, что если д-р П. Тома в двух статьях делает почтительные ссылки на трансформационную грамматику и глубинные структуры Н. Хомского, представители СИСТРАНа в Нью-Йорке Д. Аллен и Д. Берден указали, что семантические структуры Н. Хомского и его методы преобразований в системе практически не использованы. Это подтверждается также обсуждением принципов работы системы в 1976 г. (Chapin, 1976).

П. Тома описывает лингвистические основы системы следующим образом. Наиболее важная группа программ анализирует предложение исходного языка и устанавливает в нем синтаксические и семантические отношения между его частями. Эти программы составляют распознающую грамматику системы, основные параметры которой определяются закономерностями входного языка. Аналитический компонент этой грамматики определяет функции каждого слова в предложении и записывает эту информацию в специальных кодах. Независимо от анализируемого языка, эти коды указывают на одни и те же синтаксические отношения: так, русский, китайский или немецкий глагол всегда будут указывать на свой объект в опреде-

ленных разрядах ячейки, и для всех языков эти разряды будут те же самые. Связь между подлежащим и сказуемым также записывается в определенных разрядах. В процессе разработки системы для перевода с русского, китайского и немецкого языков стало ясно, что теория лингвистических универсалий действительно представляет собой отражение объективной реальности и в этом отношении допускает практическое использование. Наиболее очевидным подтверждением этого является тот факт, что одинаковые типы синтаксических отношений (члены предложения) могут быть распознаны в каждом из названных языков. Алгоритмы анализа, которые определяют эти отношения, базируются на контекстуальной информации, заключенной в пределах границ предложения, простого или придаточного внутри сложного. Система синтаксического анализа производит классификацию предложений на главное, придаточные, типы придаточных и пр.

Для распознавания элементов предложения программы начального лингвистического анализа используют словарную информацию, которая приписана каждому слову в словаре. Эта информация используется после разрешения омографии. Каждая словарная статья имеет набор признаков, характеризующих часть речи слова, его особые характеристики управления, семантические свойства, омографичность и т. д.

Особое внимание уделялось составлению словаря. Составители словаря использовали самые разнообразные источники, однако главным критерием описания слова было его текстовое употребление. Словарь специально корректировался по результатам опытных переводов; составление словарной статьи производилось с использованием принципа обратной связи. В результате тщательного изучения текстового поведения слов представляется возможным приписать словам ряд разнообразных функциональных признаков (семантические категории, такие как «мера», «движение», «химический процесс»; признаки управления и согласования и т. п.).

Следующая подробная цитата дает представление о стиле описания лингвистического обеспечения и о принципах, лежащих в основе последнего. «При составлении программ для распознавания синтаксических отношений лингвист/программист скоро убеждается, что школьные грамматик совершенно недостаточны. Поскольку программа или процедура должны быть составлены человеком, т. е. говорящим—слушающим (или, более точно, пишущим—слушающим, программа неизбежно становится средством, дающим возможность машине в некоторой степени воспроизвести человеческое поведение в этом процессе. В этом отношении аргументы Хомского, использующие языковое поведение (performance) как некоторую исходную информацию (включая данные, предоставляемые интроспекцией) становятся весьма важными для лингвиста/программиста, разрабатывающего перевод с русского на английский, который

может, например, спросить себя: «Откуда я знаю, что это конструкция «чем... тем» и что «тем» в этом случае не означает падежной формы от «эти»? и т. п. Когда он начинает исследовать эту проблему, он видит, что этот начальный вопрос ведет к пятнадцати или двадцати другим вопросам, с помощью которых проблема раскладывается на некоторые логические шаги. Его программа будет состоять из нужных вопросов относительно данной конструкции; альтернативные ответы будут вести к отдельным подпрограммам внутри программы. Кроме определения того, что действительно встретилась конструкция «чем... тем»... программа должна установить определенные обозначения, указывающие на отношения между предложениями, начинающимися с «чем» и «тем», установить отношения внутри предложений и подготовить всю конструкцию к эквивалентному переводу с помощью соответствующей английской структуры. Таким образом, программа находит систему кодов, с помощью которой через несколько секунд будет работать общая программа перевода и английского синтеза.

Новые достижения лингвистической мысли и теории могут иногда оказаться чрезвычайно полезными в лингвистическом программировании (возможно, вследствие этого, становясь чрезмерно практическими и неинтересными в глазах лингвиста-теоретика)» (Томп, 1974, 162).

Из этого описания можно сделать следующие выводы:

- лингвистическое описание отделено от алгоритма только в самой необходимой, скорее всего в минимальной степени;
- составление лингвистического алгоритма тесно связано с составлением программы. Оптимальное кодирование и последовательность лингвистических операций ставятся в основу составления процедуры анализа;
- алгоритм бинарен по своей логической природе и ориентирован на выходной язык. Его «универсальность» носит чисто технический характер.

Лингвистический анализатор СИСТРАНа состоит из следующих четырех основных этапов. Первый этап разрешает омографию. Второй исследует предложение справа налево, запоминая при переходе от слова к слову, какие типы синтаксических отношений потенциально возможны внутри каждого самостоятельного предложения, если учесть типы встретившихся слов. Используя набор регистров, в которых содержатся данные о возможных типах синтаксических отношений, этот этап затем устанавливает основные структуры внутри предложения (глагол плюс объект, предлог плюс объект и т. д.). Третий этап, двигаясь слева направо, уточняет эти отношения, опознавая типы объектов, определительные структуры и пр.

Четвертый этап, используя данные, полученные в результате работы предшествующих этапов, включая информацию о границах предложений, типах главных и зависимых предложе-

ний, осуществляет анализ в границах этих предложений, определяя подлежащее и сказуемое в каждом из этих предложений.

В каждой системе СИСТРАНа имеются также сотни специальных лексических процедур, основанных на изучении особенностей лексического поведения единиц в языке-источнике. Эти лексические процедуры осуществляются после четырех основных этапов анализа, но перед синтезом и перед реализацией необходимых преобразований синтаксических структур предложений в выходном языке. Программы синтеза обеспечивают построение правильного лексического эквивалента в выходном языке (корень плюс необходимые суффиксы и окончания, а также предлоги, где это необходимо) для каждого переводимого слова. Специальные программы преобразований структур обеспечивают необходимые преобразования структур предложений с тем, чтобы предложение имело вид, естественный для выходного языка.

Следует отметить, что в некоторых случаях информация может передаваться от одного предложения к другому. В памяти ЭВМ отведено специальное место для такой информации. Она особенно необходима для перевода местонимений, антецеденты которых могут находиться в другом предложении.

В ходе перевода необходимые переводные трансформации могут осуществляться на различных уровнях. Каждая данная трансформация может затрагивать одну лексическую единицу, фразеологический оборот, придаточное предложение, или, в некоторых случаях, может выходить за границы одного предложения. СИСТРАН, по утверждению авторов публикаций, имеет практически неограниченные трансформационные возможности, что является его отличительной чертой как системы третьего поколения.

Особенностью системы является способность разрешать неоднозначность на различных уровнях, в том числе и на уровне словаря. Эта способность разрешать неоднозначность может быть значительно увеличена добавлением новых процедур без изменения основного алгоритма.

Для разрешения лексической многозначности предусмотрено несколько способов. Одним из них является введение и использование семантических признаков. Поскольку семантические коды приписываются словам входного языка, проблема неоднозначности фактически разрешается на уровне этого языка и таким образом синтез на любом другом языке может опираться на результаты разрешения неоднозначности во входном языке. Единственным исключением из этого правила является перевод предлогов.

Организация словарной статьи в СИСТРАНе позволяет неограниченно пополнять ее новой информацией.

Особо следует отметить способы использования семантической информации. Для СИСТРАНа было разработано триста семантических категорий (признаков). Эти категории были по-

лучены эмпирическим путем, т. е. практически во время работы системы. Эмпирическая выработка категорий по тексту имеет ряд преимуществ перед теоретическим составлением набора таких категорий, поскольку в первом случае гарантируется их широкое и универсальное применение. Важным моментом в этом отношении является также и то, что эти категории используются на уровне анализа входного языка. Это обуславливает их применимость к переводу с разных языков или с одного языка на многие. К лексическим единицам или к словосочетаниям может также быть приписана практически любая прагматическая информация.

Большое, если не решающее значение имеет создание программного обеспечения. Практически любая публикация по СИСТРАНу начинается с программного обеспечения. Утверждается, что разработка этого обеспечения была наиболее трудоемкой частью работы и что на выполнение ее были затрачены многие годы и сотни часов машинного времени. «Восемь лет научно-исследовательской работы и осуществление систем первого и второго поколения убедили меня в 1964 году в том, что основой для большой, универсально пригодной системы машинного перевода должно быть тщательно продуманное, универсальное программное обеспечение, которое могло бы воплотить все необходимые компоненты системы машинного перевода. В ходе разработки стало ясно, что основные программы должны быть написаны на языке, наиболее близком языку машины, т. е. на ассемблере» (Томп 1977, 571). Основные задачи, которые выполняются универсальным программным обеспечением СИСТРАНа, следующие:

- считывание текста;
- разбивка текста на слова;
- поиск слов в различных словарях;
- приписывание словарных кодов словам или словосочетаниям текста;
- распределение памяти под анализируемые предложения;
- контроль за выполнением программ перевода;
- подготовка перевода к печати, выдаче на микрофиши и пр.

Основное программное обеспечение СИСТРАНа дополняется различными отдельными программами и подпрограммами, реализующими частные задачи и задачи этапов. Так, специальные программы обеспечивают пополнение словарей, вспомогательные виды работ со словарями (составление конкордансов и пр., изучение диагностического материала и т. п.).

Ввод текста для перевода может быть осуществлен с помощью самых различных устройств (с перфокарт, перфоленты, с оптического считывающего устройства. Рассматривался также вопрос об устном вводе, однако здесь конкретных результатов еще не достигнуто). Текст проходит через ЭВМ, и результат выдается или на печать, или на устройство фотовывода для

вставки графиков, диаграмм и т. п. Переведенные предложения выводятся на экран дисплея, где постредактор редактирует его в случае необходимости, направляя затем на вывод (распечатку на бумаге, микрофлши и пр.). Исправления, внесенные редактором, автоматически накапливаются и затем выдаются лингвистам для внесения изменений в систему. СИСТРАН обеспечивает лингвистов большим объемом информации в виде конкордансов и специальных словарей, с тем, чтобы оценить распространенность и тип того или иного синтаксического и семантического явления самым тщательным образом, прежде чем принимать решение об учете его в системе перевода.

Программа поиска в словаре работает следующим образом. Как только текст введен в машину, каждое слово сравнивается с данными, содержащимися в специальной таблице наиболее частотных слов. После нахождения слова в таблице, слову текста приписывается номер, в котором содержится указание на адрес, по которому расположена информация к данному слову. В специальной таблице информации, которая вызывается в оперативную память, содержится вся необходимая информация грамматического и семантического характера к основным наиболее частотным словам. Слова низкой частоты, отсутствующие в этой таблице, сортируются в таком же порядке, в каком построен основной массив словаря, и затем отыскиваются в словаре. Основной словарь для поиска упорядочен по длинам корней слов, однако это важно только для флективных языков типа русского. Для других языков используется другая методика.

Каждая программа системы перевода вызывается управляющей программой и вводится в оперативную память по мере необходимости. Управление вызовом программ полностью автоматическое; другими словами, как только текст введен в ЭВМ и программа управления начала работать, СИСТРАН самостоятельно переходит от одного этапа процесса перевода к другому и последовательно осуществляет весь процесс, от поиска по словарю до анализа входного текста и синтеза выходного.

Для анализа в памяти отведена специальная зона. Результаты анализа записываются на специальном языке-посреднике (interlingua). Разработка этого языка-посредника дает возможность переводить сразу на несколько языков.

В зоне анализа каждому слову отводится 160 байтов. Из них часть содержит словарную информацию, в других записывается информация, полученная в ходе работы этапов анализа. Таким образом, информация при каждой словоформе содержит синтаксические и семантические данные, полученные как из содержащегося в ЭВМ словаря, так и в результате анализа входного текста программами этапов. Специальные коды СИСТРАНа, приписанные отдельным словам, а также сочетаниям слов, заключают в себе обширную семантическую и синтаксическую информацию о каждом слове и его взаимодействии

с другими словами предложения. В последнюю входят такие широкие традиционные концепции как прямые и косвенные дополнения, а также такие нюансы, как тип субъекта или объекта, предпочитаемый данным глаголом. Как указывалось выше, общие синтаксические характеристики для всех языков, переводимых СИСТРАНам, записываются в те же самые места ячеек и выражаются в одинаковых кодах. Это обеспечивает пригодность системы для разных языков.

Программное обеспечение системы позволяет обеспечить скорость перевода в 200 000—300 000 слов в час.

2. Джорджтаунская система машинного перевода

Второй достаточно известной системой МП практического действия является так называемая Джорджтаунская система.

С 1964 г. в рамках автоматизированного информационного обслуживания министерства энергетики США и других федеральных ведомств используется действующая система МП. Оук-Риджская национальная лаборатория фирмы Юнион Карбайд применяет эту систему для перевода с русского языка на английский. Впервые эта система была реализована в Джорджтаунском университете в 1960 г., затем в течение нескольких лет совершенствовалась по мере применения новых моделей ЭВМ. Хотя выполняемые переводы не отличаются высокими литературными достоинствами, система, благодаря своему особому словарю, используемому контекст, обеспечивает перевод технической литературы сравнительно быстро и недорого. В системе большую роль играет так называемый лингвистически-ориентированный язык программирования (Simulated Linguistic Computer — SLC) — ЛОЯП. С помощью этого языка система в целом может быть быстро перестроена на новый материал. Словарная статья любого слова может быть дополнена процедурами, которые позволяют изменять значение слова в переводящем языке в зависимости от контекста. Все предложение в целом анализируется более сложными лингвистическими алгоритмами.

Свыше 90% потребителей оценивают переводы как «полезные» или «хорошие». По мере того, как применение системы расширяется, она совершенствуется и более подробно документируется.

Система МП Джорджтаунского университета была начата разработкой в 1956 г. под руководством Л. Достерта (известный в истории МП так называемый Джорджтаунский эксперимент, первый в США эксперимент по переводу на ЭВМ специально подобранного текста с использованием небольшого словаря и фрагмента грамматики; имел цель обратить внимание научной общественности на идею машинного перевода). Наибольший вклад в развитие последующих идей был сделан М. Заречняком (метод «общего анализа») и Э. Брауном («анализ по предложениям»). При «общем анализе» предложение

анализируется на нескольких уровнях, включая: а) образование слов и идном; б) группы слов, образующие словосочетания существительного с прилагательным и глагола с существительным, с соответствующим управлением, и в) синтаксические отношения, такие как подлежащее и сказуемое. С помощью метода «анализ по предложениям» лингвист может, исходя от некоторой незначительной исходной информации, расширить систему все более сложными данными и процедурами, повышая ее «компетенцию» до тех пор, пока она не будет в состоянии анализировать любое предложение.

Русско-английский перевод осуществляется на ЭВМ ИБМ 360—370 на ОС 360. Все программы написаны на ассемблере. Их всего четыре: один реализует перевод, две используются для внесения изменений в словарь, одна — для распечатки словаря. Перевод 10 тыс. слов (50 страниц) занимает 41 секунду (время работы центрального процессора). Предполагается переводить с помощью этой системы также и нетехнические тексты, например, материалы по советско-американским соглашениям о культурном обмене (Jordan, Brown, Hutton 1977).

Возможности системы, ее основные лингвистические принципы могут быть обнаружены изучением используемого языка программирования, поэтому целесообразно рассмотреть его, опираясь на имеющиеся публикации.

Лингвистически-ориентированный язык программирования представляет собой набор мнемонических команд, в число которых входят команды просмотра предложения в определенном порядке для получения разнообразной лингвистической информации, команды добавления или уничтожения информации, а также изменения порядка следования любого элемента предложения.

ЛОЯП был разработан и применен впервые Э. Ф. Р. Брауном в период с 1952 по 1963 годы в рамках Джорджтаунской системы МП с русского языка на английский. Язык совершенствовался все последующие годы. Основные задачи, которые выполняет ЛОЯП, следующие:

- простая подготовка словаря;
- обновление старых элементов словаря;
- обработка введенного (русского) текста;
- процесс перевода;
- основные команды ЛОЯП: обработка языковых строк (последовательности информации в виде заполненных информационных ячеек слов); выполнение команд над окончаниями;
- проверка и изменение информации в элементах;
- кодировка простых идном.

Система перевода имеет два основных режима работы:

- 1) режим генерации,
- 2) режим перевода.

Под генерацией системы понимается составление специаль-

ного словаря и связанные с этим лингвистические операции. Режим перевода осуществляет собственно перевод.

Фаза перевода системы использует файлы словаря и операции над входным текстом для выработки информации, необходимой для синтеза переводящего текста на выходном языке.

Фаза генерации используется для составления нового словаря или для изменения существующего словаря. Заново откомпилированные элементы затем объединяются со старым словарным файлом для выработки измененной версии. Словарь при необходимости может быть распечатан весь или по частям.

Программа перевода состоит из файла стандартных лингвистических операций для анализа, которые пригодны для разбора практически любого предложения русского языка, и из словарного файла, в котором содержатся элементы, позволяющие осуществить весь процесс перевода, грамматическая информация, а также набор операций, изменяющих переводные эквиваленты в зависимости от контекста. Пользуясь этими данными, программа перевода воспринимает предложение русского языка и создает из него список «элементов» по одному для каждого слова и для знака препинания. «Элемент», таким образом, соответствует некоторому месту памяти в объеме одной ячейки (информационная ячейка — ИЯ) или нескольких ячеек. В таком элементе, однако, содержатся не только словарные данные и информация, полученная в результате анализа, но также все специальные операции и инструкции относительно того, что следует произвести с данным словом в дальнейшей обработке.

Вводится специальная метка для конца предложения — астериск. Она вручную приписывается каждому предложению. Таким образом, перевод осуществляется по предложениям.

На первом этапе анализа каждое слово предложения рассматривается независимо и в его ИЯ содержится грамматическая информация, а также информация о порядке обработки этого слова в любой ситуации. На втором этапе ИЯ с их раздельными наборами инструкций анализируются на взаимодействие так, чтобы можно было выработать правильный идноматичный перевод с русского на английский. В системе это осуществляется объединением всех словарных инструкций в отдельный набор, который упорядочен по приоритетам. Затем над предложением работает каждая инструкция, начиная с инструкции высшего приоритета, или — в случае одинакового приоритета — слева направо.

Каждая из инструкций содержит несколько проверок, которые нужно осуществить, сканируя элементы предложения и отмечая «характеристики» лингвистической информации, такие как род, число, падеж и пр., либо другим способом конкретизируя единицы, например, с помощью номера. Если проверка производит отрицательный результат, то выполнение данной инструкции прекращается. Если условие проверки выполняет-

ся, то осуществляется изменение одного или более элементов слов. После завершения выполнения одной инструкции система выбирает для исполнения следующую по величине приоритета инструкцию. Когда инструкции кончаются, предложение может быть выведено, начиная с первого элемента, причем выводится результирующий английский эквивалент каждого элемента, кончая точкой. Завершив вывод предложения, система считывает новое и осуществляет весь процесс сначала, начиная с поиска по словарю.

Кроме инструкций, связанных с отдельными словами, существуют операции, касающиеся каждого отдельного предложения в целом. Например, сюда относятся инструкции, касающиеся определения подлежащего предложения. Всем таким инструкциям также приспан определенный приоритет, в соответствии с которым они выполняются.

Из вышеприведенных описаний видно, что система осуществляет перевод по предложениям. Предложение считается предельным контекстом, который может в данной системе анализировать ЭВМ. Поэтому такие действия, как например, нахождение antecedентов, которые требуют выхода за пределы одного предложения, в системе не реализуются.

Оценивая в общем указанные принципы грамматического анализа, можно утверждать, что оба этапа, т. е. этап составления «элементов» (ИЯ) и этап сравнения (сопоставления) информации к элементам достаточно громоздки. Первый этап загромождается неоднократными обращениями к словарю входного языка за различными данными. Второй — за счет соединения описания, или «грамматки», с механизмом. Система приоритетов есть достаточно сложная система, она значительно сложнее, чем, например, система жестких алгоритмов работы над текстом, порядок которых задан. Тогда можно отказаться, например, от некоторых неудобных допущений, которые существуют в системе: так, когда при равенстве приоритетов инструкции выполняются слева направо, это ведет к ошибкам в анализе. Многократное же обращение к словарю заведомо увеличивает время перевода (Jordan, Brown, Hutton 1977).

Рассмотрим основные принципы организации словаря. В системе используется комбинированный словарь: основ слов и словоформ. Каждая основа вводится в словарь лишь один раз; каждое окончание, которое может быть использовано при разных основах, также вводится один раз. В словарной статье указано, какие окончания могут находиться при данной основе. Такая организация упрощает поиск по словарю и сокращает время поиска.

Русско-английский словарь, используемый системой, состоит из 31 000 основ слов, что, если учесть падежные, временные и пр. окончания, эквивалентно объему примерно в 300 000 словоформ. Словарь подготовлен для перевода технической литературы; при составлении словаря использовались не существ-

ующие словари, а тексты. Тем не менее словарь содержит значительное количество общеупотребительных русских слов: опытный перевод русского текста про балет дал достаточно приемлемые результаты.

Основная форма словарной статьи следующая:

Русская основа	— указание на грамматический класс	— лингвистический номер	— английский перевод
----------------	------------------------------------	-------------------------	----------------------

Русская основа транслитерируется по стандартной схеме преобразования кириллицы. До грамматического класса могут находиться одно или несколько указаний на словоизменительные элементы или на парадигму данной основы. Если словоизменительных данных нет, дефис означает нулевое окончание. Указание на грамматический класс факультативно: если слово неизменяемое, такого указания нет. Лексический номер представляет собой индекс, который указывает на какие-либо словарные или грамматические особенности данного слова, связь его с другими словами и т. п. В случае отсутствия таких особенностей номер нулевой. Английский эквивалент данному русскому слову приписывается заранее и всегда.

Несколько слов об истории и современном состоянии системы. В 1964 г. копия системы была передана Комиссии по атомной энергии в Оук-Ридже, штат Тенесси, для проведения широких испытаний и всесторонней оценки системы. Одновременно такая же договоренность была достигнута с информационным Центром Евратома в г. Испра, Италия. Известный доклад наблюдательной комиссии академии наук по машинному переводу в 1966 г. отрицательно повлиял на развитие МП в США, но Джорджтаунская система оказалась затронутой этим докладом в наименьшей степени, поскольку результаты ее работы устраивали заказчика. «Только потребитель, которому нужна информация, может оценить, удовлетворяет ли его материал, который он получает, и готов ли он тратить дополнительное время на чтение стилистически негладкого текста. Последнее может оказаться небольшой ценой за быстрое получение перевода, вместо того, чтобы ожидать его в течение многих дней или месяцев» (Jordan, Brown, Hutton 1977, с. 27).

Следует обратить внимание, что создатели Джорджтаунской системы Л. Достерт и другие исключали пре- и постредактирование, стремясь получить максимально пригодный продукт на ЭВМ.

В 1973 г. Боженной Достерт было проведено исследование по оценке качества МП и изучению реакции потребителей. Это исследование базировалось на опыте десятилетнего обслуживания потребителей. Основные выводы:

— качество машинного перевода оценивается как «хорошее» или «приемлемое» 92,4% из 58 ответивших на запрос;

— знакомство с предметом имеет решающее значение для понимания МП;

— в случае, если время ожидания «человеческого» перевода в три раза больше, чем время машинного перевода, 87% потребителей предпочитают МП;

— 96% ответивших рекомендуют МП своим коллегам (Dostert 1973).

МП представляет собой весьма полезное информационное обслуживание. Пользователи выразили желание и в дальнейшем прибегать к МП и способствовать развитию и совершенствованию системы. Выяснилось, что словарь имеет гораздо большее значение для понимания и для качества перевода, чем синтаксические особенности текста.

В настоящее время система состоит из четырех частей: лингвистическое обеспечение, используемое для перевода, словарь, которым программы пользуются в переводе, сами программы, и ЛОЯП. Лингвистический подход остается в принципе неизменным в том виде, в каком он был разработан М. Заречником в Джорджтаунском университете. Некоторые усовершенствования подготовлены, но будут внесены в дальнейшем, когда это позволят финансовые соображения. Словарь в значительной мере тот же самый, который был разработан ранее, однако существенные изменения были внесены в него в ходе эксплуатации системы. Значительно усовершенствовались программы и все программное обеспечение в целом, благодаря новым возможностям вычислительной техники.

ЛОЯП не ограничен только русским языком; фактически он был впервые разработан и использован для перевода с французского на английский. Представляется, что ЛОЯП будет полезен и для перевода с других языков.

Скорость перевода: недавний перевод 2100 слов английского текста занял 8,76 сек. времени работы центрального процессора на машине ИБМ 360—195. Другой прогон в объеме 8—10 000 слов (50 страниц) занял 41 секунду времени работы процессора. Все программы написаны на ассемблере, ни одна из них не требует объема памяти большего чем 190 К.

3. Система машинного перевода «КУЛЬТ»

В течение последних шести лет в Китайском университете Гонконга под руководством профессора Ш. С. Ло была разработана система машинного перевода с китайского на английский язык. Система переводит математические и физические тексты. С 1975 г. эта система, называемая КУЛЬТ CULT (Chinese University Language Translator) — регулярно используется для перевода китайского математического журнала АСТА МАТЕМАТИКА SINICA, публикуемого Китайской Академией наук в Пекине. Система реализована на американском варианте стандартного языка программирования ФОРТРАН. Про-

грамма может работать на любой ЭВМ, имеющей оперативную память в пределах 32 К слов и внешнюю память объемом около 1 мегабайта.

Работы по МП были начаты в Китайском университете Гонконга в конце 1969 г. Целью было изучить возможности автоматизации перевода с китайского языка на английский. После предварительного изучения китайских научных текстов с точки зрения их лингвистической структуры был сделан вывод, что автоматизация их перевода возможна, хотя полностью автоматизированный высококачественный машинный перевод в духе Бар-Хиллела осуществить трудно. Первый вариант системы КУЛЬТ был разработан и опробован в октябре 1972 г. В последующие годы система подверглась значительным усовершенствованиям, в результате чего эффективность ее работы существенно повысилась.

Помимо упомянутого математического журнала, выходящего ежеквартально, система переводит также журнал АСТА РНИ-SIKA SINICA.

Система КУЛЬТ состоит из основной программы и подпрограмм.

Китайские предложения сначала кодируются телеграфным кодом и вводятся в ЭВМ, вместе с формулами и цитатами. Специальная подпрограмма кодирует формулы с помощью четырехразрядного вспомогательного кода. Числовые коды хранятся на диске отдельно от кодов формул. Одна запись в файле содержит одно предложение (одно предложение отделяется от другого пятиразрядным кодом от 10100 до 10800).

Предложения анализируются по одному. Специальная подпрограмма осуществляет сравнение телеграфных кодов с записями в машинном словаре. По принципу наибольшего возможного совпадения из отдельных букв (нероглифов) складываются слова. Другая подпрограмма приписывает каждому слову особый функциональный код. На этой стадии слова подвергаются предварительному анализу, в результате чего разрешается омография «глагол — существительное» и каждому из слов приписывается одна синтаксическая функция.

На втором этапе обработки проверяются знаки препинания. Вопросительное предложение переводится иначе, чем повествовательное. Информация о типе предложения накапливается в специально отведенных для этого ячейках. Третий этап обработки анализирует предложные обороты. Каждое слово во фразе переводится соответствующей подпрограммой (по частям речи) и накапливается в памяти. На четвертом этапе обработки в предложении отыскивается глагол. В настоящее время система может переводить только предложения, в которых есть по крайней мере один глагол.

На пятом этапе проверяется первое слово предложения. Если это союз, то он переводится и запоминается. Если это глагол, то анализируется сказуемое. Если в предложении есть под-

лежащее, то часть фразы перед глаголом извлекается из текста и переводится. Каждое слово анализируется соответствующей подпрограммой, в зависимости от функциональных кодов. Выбирается наиболее подходящая форма для слова в выходном языке и определяются порядок слов, артикли и вспомогательные слова. Однако наречия, непосредственно стоящие перед глаголом, на этом этапе не анализируются. Эти наречия передаются для обработки в ту часть программы, которая анализирует сказуемые, и только потом эти наречия переводятся.

При анализе с целью нахождения сказуемого сначала разбираются наречия. Эти наречия после соответствующего анализа переводятся, и определяется их место в предложении — должны ли они размещаться до или после глагола. Переводятся и размещаются в предложении вспомогательные глаголы. Затем программа МП переходит к следующему этапу для определения главного глагола. На этом этапе определяются залог и время сказуемого. Из словаря выбирается соответствующая глагольная форма.

Затем программа второй раз проверяет пунктуацию. Если переводится вопросительное предложение, то вставляются вспомогательные глаголы и производятся соответствующие изменения в порядке следования слов. Затем глагол проверяется на переходность/непереходность. Переходный глагол направляет программу на анализ дополнений: проверяется количество дополнений и каждое из них переводится отдельно. Неperеходный глагол проверяется на наличие при нем обстоятельств: если они имеются, то переводятся, в противном случае вновь проверяется пунктуация.

Проверка пунктуации нужна для определения того, имеет ли далее место конец предложения или закончена лишь часть его. На этом этапе рассматриваются три возможных случая. Если знак препинания отделяет часть предложения, то программа возвращается к моменту перевода дополнений. Если знак препинания означает конец предложения, программа проверяет, не является ли предложение вопросительным, и если да, произведены ли необходимые изменения порядка слов. Если знак препинания есть ни то, ни другое, программа проверяет, является ли основное слово глаголом. Если да, то глагол переводится (инфинитивом). Если нет, то произошла ошибка, программа распечатывает место с ошибкой и переходит к анализу следующего предложения. Если ошибок нет, то программа вновь переходит к задаче анализа и перевода дополнений.

Наконец, анализируется и переводится все предложение. На различных этапах процесса переводы накапливались в разных местах, теперь они все сводятся в одно место памяти. Упорядочивается пунктуация. Специальная подпрограмма объединяет переведенные предложения и их части, а также формирует абзацы. При этом производится вставка формул, уравнений и прочих нетекстовых включений. Программа проверяет, все ли

предложения получили перевод. На заключительном этапе производится распечатка результатов перевода и работа программы закончена.

В анализе используются следующие части речи: существительное, числительное, специальное слово, глагол, предлог, прилагательное, мера, наречие, вспомогательный глагол, союз, знак препинания. В синтаксическом анализе используются следующие категории (типы предложений и члены предложения): группа существительного, группа прилагательного, основное существительное, предложный оборот (предложная группа), подлежащее, сказуемое, дополнение (объект), основной глагол, глагольная группа, главное предложение, подчиненное предложение.

Особо отмечается, что трудности перевода были во многом связаны не столько с машинным переводом в собственном смысле, сколько с отсутствием правильной переводной терминологии научных и технических терминов, встречающихся в литературе. Это особенно заметно в области физики.

В настоящее время изучается вопрос о применении системы КУЛЬТ к переводу как других типов текстов, так и для других языков (Loh, Kong L. 1977).

4. Другие системы машинного перевода

Г. Брудерер (Bruderer 1977, с. 538) дает следующую схему, согласно которой можно квалифицировать системы МП, в том числе действующие и разрабатываемые (см. рис. 17).

Метод 1: пословный перевод с морфологическим анализом.

Метод 2: перевод с морфологическим и синтаксическим анализом.

Метод 3: перевод с морфологическим, синтаксическим и семантическим анализом (кроме того, возможен также прагматический анализ с помощью искусственного интеллекта).

Метод 4: перевод с помощью универсального языка-посредника (интерлингвы).

Методы 1 и 2 возможны сегодня; метод 3 возможен частично; метод 4 все еще недоступен.

На конференции «Преодоление языковых барьеров» в Люксембурге в мае 1977 г. большое внимание было уделено разработке систем МП в русле исследований, проводимых в течение многих лет группой под руководством проф. Б. Вокуа. Исследования эти были начаты в Гренобльском университете, продолжались затем в группе, объединяющей также и другие университеты Франции; с некоторого времени работы по данной методике начались также в Монреальском университете в Канаде и в других научных центрах (Voitet 1977). В 1974 г. на базе исследований этой группы был создан коллектив «Лейбниц», занявшийся разработкой практически пригодных систем МП. В задачку нашего обзора не входит детальное описание разно-

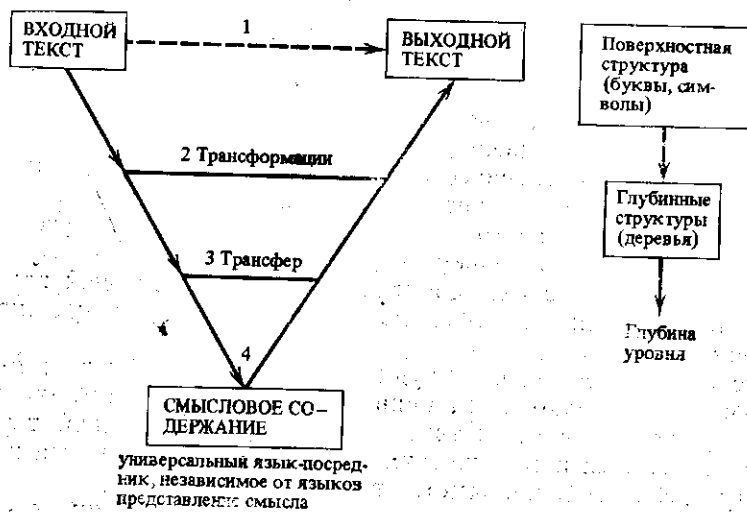


Рис. 17. Схема классификации системы МП

образных научных направлений в МП и теоретических принципов, которых придерживается тот или иной исследовательский коллектив, поэтому мы ограничимся описанием системы МП Саарландского университета в Саарбрюккене, ФРГ. Это описание содержится в докладе Г. Брудерера на упомянутой конференции (Bruderer 1977). Сохраним анкетную форму представления данных как весьма наглядную.

Университет Саарланда, Саарбрюккен, ФРГ

1. Общие положения
 - 1.1. Название системы
 - 1.2. Цель
 - 1.3. Руководители
 - 1.4. Цель исследований
 - 1.5. Дата начала работ
 - 1.6. Когда начались эксперименты
 - 1.7. Нынешние сведения по состоянию на

—	автоматический перевод
	Ганс Эггерс, Хайнс Дитер Маас
	выработка методики
	1967
	последний вариант в 1974 г.,
	ранние — в 1969—1973 гг.
	сентябрь 1976 г.
2. Лингвистические данные
 - 2.1. Общие характеристики
 - 2.1.1. Какие пары языков

—	русско-немецкий, англо-немецкий, эсперанто-немецкий
	в настоящее время необратим (зависит от словаря)
	анализ может быть применен для любых (подобных) языков; в настоящее время он используется как для русского, так и для немецкого языка
 - 2.1.2. Направление перевода (в какую сторону производится)
 - 2.1.3. Возможности расширения

- 2.2. Какие тексты
- 2.3. Обработка алфавита
- 2.4. Словарь
 - 2.4.1. Основная структура
 - 2.4.2. Способ записи
 - 2.4.3. Объем
 - 2.4.4. Области языка
 - 2.4.5. Поиск по словарю
- 2.5. Грамматика
 - 2.5.1. Модели языка
 - 2.5.2. Перевод

математические и лингвистические публикации; научно-популярные тексты (несколько млн словоформ) обращение кириллического алфавита в отношении 1:1

- а. отдельные словари для анализа и трансфера;
- б. частотный словарь, общий словарь;
- в. словарь основ
 - а. переменная длина словарной статьи;
 - б. морфологические, синтаксические и частично семантические данные
- русский словарь: приблизительно 13 500 слов (8500 основ)
- русский частотный словарь: 255 словоформ
- русско-немецкий словарь: 8500 словарных статей
- эсперанто: 4000 слов
- английский: несколько сот слов
- общеупотребительная лексика
 - а. каждое слово текста разделяется (сегментируется) справа налево до тех пор, пока не найдено его окончание в таблице окончаний. Левая часть — возможная основа — последовательно в соответствии с индексами отыскивается в словаре. Разделение слова продолжается и в том случае, если уже найдено решение
 - б. принцип наибольшего возможного совпадения фиксированных синтагм (идиом)
- а. система разрабатывается в двух фазах:
 - первая фаза: поверхностная грамматика (с частичным разрешением омографий на основе дистрибуции)
 - вторая фаза: рассмотрение трансформационных структур (с разрешением синтаксической неоднозначности)
- б. трансформационные правила
- в. опора на синтаксический анализ
- г. грамматика введена частично как грамматика непосредственно составляющих, независимая от алгоритма, частично включена в алгоритм
 - а. три стадии: анализ, трансфер, перевод
 - б. анализ ориентирован на входной язык, а синтез на выходной
 - в. в результате анализа определяется роль слов в предложении. Для

именных групп частично используется предсказуемый анализ г. осуществляются по крайней мере три просмотра предложения слева направо

никакого пре- или постредактирования не предусматривается перевод возможен только для относительно простых структур предложений (например, только для придаточных относительных)
индекс — KWIC

Telefunken TR 440
MV17 (Maintenance Version)
52 К слов (программа перевода делится на пять основных программ, которые работают последовательно; иначе потребуется примерно 200 К слов)
магнитные диски, магнитная лента
Фортран, Телефункен ассемблер
перфокарты, дисплей
печатающее устройство, дисплей
заглавные буквы

нельзя сказать (стадия отладки)
Deutsche Forschungsgemeinschaft
нет
15 000—20 000 слов в час
нельзя сказать (нет данных)

Университет Саарланда является членом международной исследовательской группы по машинному переводу «Лейбниц», основанной в 1974 г. Саарбрюккенская система МП была успешно продемонстрирована в Саарбрюккене 24 сентября 1976 г. (Языки: русский—английский—эсперанто—немецкий).

постредактирования) противопоставлена практическому информационному обслуживанию. Можно, однако, предположить, что такая ситуация в какой-то мере вызвана спецификой социального статуса исполнителя. Университетский исследовательский центр, по-видимому, призван обеспечить в первую очередь высокое качество и эффективность учебного процесса, и лишь затем информационное обслуживание сторонних потребителей.

В некоторых публикациях упоминаются также и другие действующие системы МП. Джон Шандье в докладе о разработке системы МП второго поколения для перевода технических описаний по авиации в Канаде (Chandioux 1977) в числе действующих систем второго поколения называет Логос III (США), которую он призывает не считать происходящей от Джорджтаунской системы, и МЕТЕО (Канада), для перевода сводок погоды. Более подробных сведений об указанных системах составитель настоящего обзора не имеет.

Требуется некоторого уточнения положение о «поколениях» систем машинного перевода. Видимо, впервые представление о поколениях систем МП было введено некоторыми советскими учеными (см., например Автоматический перевод 1971). В их представлении в МП следует различать три поколения. Первое характеризуется преобладанием словаря и морфологического анализа. Синтаксический и семантический анализ в этих системах не применяется, перевод в большинстве случаев пословный с незначительным грамматическим согласованием. Алгоритм не отделен от лингвистического описания. Второе поколение характеризуется усиленным вниманием к синтаксическому анализу и некоторым отходом от словаря. Реализуется принцип отделения описания от алгоритма, грамматики от механизма. Введено понятие многовариантности на всех уровнях. Третье поколение базируется на формализации семантики. Детально разработаны алгоритмы синтаксического анализа, используются алгоритмы семантического анализа и синтеза, вновь используется словарь, однако особого типа, с обширной синтаксической и семантической информацией. Многовариантность обеспечена на всех уровнях анализа и синтеза. Главное внимание в системах третьего поколения обращено на разработку универсальных языков смысла, в области программного обеспечения созданы высокоформализованные языки, базирующиеся на формальных грамматиках.

Критика концепции поколений основана главным образом на том, что все действующие системы относятся по существу либо к первому, либо неизвестно к какому поколению, поскольку принципы, положенные в основу классификации поколений, на самом деле препятствуют построению действующих систем МП и практическому осуществлению МП. Так, полная и абсолютная независимость описания от алгоритма перевода практически не дает возможности построить алгоритм, многовариантность на всех уровнях не позволяет вообще получать ког-

2.6. Вмешательство человека

2.7. Качество перевода

2.8. Другие применения

3. Технические характеристики

3.1. Вычислительная машина

3.2. Операционная система

3.3. Основные требования к памяти

3.4. Вид памяти

3.5. Языки программирования

3.6. На чем вводятся данные

3.7. Вывод

3.8. Представление букв

4. Экономические характеристики

4.1. Объем переводов

4.2. Кто финансирует

4.3. Потребители

4.4. Скорость перевода

4.5. Стоимость

5. Примечания

Судя по данным таблицы, можно сказать, что система обладает рядом недостатков, препятствующих на данном этапе ее практическому использованию. Так, анализу могут подвергаться только простые предложения или такие сложные, придаточные в которых только определенного типа. В результате подобных ограничений система не имеет потребителей, несмотря на то, что разрабатывается уже в течение сравнительно долгого времени. В свою очередь, отсутствие потребителей не дает возможности определить не только экономические параметры (например, стоимость), но и лингвистические возможности системы применительно к реальным текстам. Таким образом, принципиальная целевая установка на полностью автоматизированный высококачественный машинный перевод (отказ от пре- и

да-либо текст перевода без вмешательства человека, перевод всех явлений языка на глубинный уровень чрезвычайно неэкономичен, нетехнологичен и отрывает от нужной поверхностной структуры для выходного языка и т. п. Тем не менее некоторые зарубежные ученые, в частности исследователи группы Б. Вокуа, придерживаются этой концепции. Любопытно в связи с этим оговорки, которые они вынуждены делать, пытаясь призвать разрабатываемые системы к тому или иному поколению.

«Осознавая ограничения второго поколения с семантической точки зрения, исследователи проекта ТАУМ¹ тем не менее придерживаются достижений этого поколения на сегодняшний день:

- полный синтаксический анализ английского предложения;
- англо-французский трансфер;
- независимое порождение выходного языка;
- написание грамматик на языках высокого уровня.

В конце концов система предназначена для технического перевода и это определяет как терминологию, так и синтаксис с семантикой. Разработка системы, которая была бы более универсальной, потребовала бы гораздо больших усилий, и пока еще нет доказательств, что такая система возможна» (Chandioux 1977, с. 615). Применительно к словарю: «Во-первых, следует отметить, что в ходе трансфера не производится никакого обращения к словарю: это не означает пересмотра основных принципов второго поколения, а просто стремление к экономии, которое вполне логично для строго бинарной системы» (Chandioux 1977, с. 617). В ходе этих оговорок практически снимаются основные классификационные принципы в таком понимании поколений МП.

Один из ведущих исследователей группы Б. Вокуа К. Буате следующим образом описывает эволюцию исследований: «Эксперименты, проведенные по анализу русского, немецкого и японского языков и по синтезу французского, достигли своей кульминации в период между 1967 и 1970 годами, в течение которого грамматика и словарь русского и французского языков были завершены составлением и использованы для перевода убедительного объема текстов русских научных статей, содержащего более чем 300 000 слов, с использованием компьютера ИБМ 7044 Гренобльского университета.

Тем не менее эти эксперименты показали определенные недостатки системы. С одной стороны, названная выше методология была выработана в течение многих лет и не могла полностью соблюдаться во всех компонентах системы. С другой стороны, формальные модели, лежащие в основе двух принципиальных фаз синтаксического анализа («расширенная внекон-

текстная грамматика» в нормальной форме) и «накленвания ярлыков» (изменения структуры, приводящие к глубинному представлению, или к «языку-посреднику») оказались слишком трудными, чтобы с ними могли работать лингвисты, начиная с некоторого объема грамматики. В конце концов формализм «языка-посредника», может быть, был слишком претенциозным: поскольку все следы поверхностных феноменов терялись в конце анализа, часто оказывалось необходимым вновь синтезировать информацию, которую просто можно было бы не терять; к тому же, что более серьезно, почти всегда оказывалось невозможным переводить по «кускам» предложение, для которого имелись только частные результаты анализа» (Boitet 1977, с. 90). Относительно применения языков высокого уровня К. Буате говорит: «Может оказаться сообразительным использовать простой формализм (LISP, Q — systems) для написания и реализации всей системы машинного перевода, поскольку такой формализм может быть предоставлен лингвисту. Но помимо того факта, что эти системы всегда имеют разрешающую способность машин Тьюринга и поэтому невычислимы, это приводит к тому, что расходуются большие объемы памяти и времени для простых задач обработки (таких, как морфологический анализ и синтез), в такой степени, что ни одна из попыток в этом направлении не привела к созданию систем, способных переводить большие объемы текстов» (Boitet 1977, с. 93).

В связи с этим, если несколько утрировать, можно сказать, что на самом деле в построении практических систем развитие идет в направлении, противоположном развитию теоретической мысли в описываемой концепции поколений, а именно к более тщательному учету специфики научно-технических подязыков, собственно переводной функции языка, технико-экономическим и прагматическим соображениям, использованию взаимодействий с человеком и пр. Из этого, конечно, не следует отрицание важности разработки фундаментальных теоретических положений как машинного перевода в собственном смысле, так и лежащих в его основании наук, в первую очередь вычислительной лингвистики. Представляется, что классификация систем МП должна основываться, в первую очередь, на сравнении и сопоставлении полученных на этих системах практических результатов в смысле качества собственно машинного перевода, скорости перевода, стоимости и пр. Видимо, основой классификации с идейной точки зрения должны стать прикладные модели (теории), оптимальным для вычислений образом описывающие собственно переводческий процесс.

П. Тома, описывая СИСТРАН, употребляет термин «поколение» уже в другом смысле, не касаясь отделения анализа от синтеза, многовариантности и т. п. Скорее всего, употребление им этого термина носит как бы генеалогический характер, определяя в первую очередь эволюцию его собственной системы.

¹ ТАУМ — Traduction Automatique de l'Université Montreal — группа машинного перевода Монреальского университета.

Учитывая все вышесказанное, читателю рекомендуется не придавать термину «поколение» применительно к системам МП какого-либо абсолютного или тем более эталонного значения.

Общее состояние МП оригинально определено в выступлении председателя секции МП на упомянутой конференции в Люксембурге. «Поскольку теоретические основы в достаточной степени созданы, перед нами теперь стоят проблемы совершенствования систем и включения человека-переводчика в автоматические процессы, особенно в диалоговом режиме. Становится очевидным слияние автоматического перевода и устройств помощи переводчику. Безусловно, возникнут новые проблемы: проблемы конкуренции, даже соперничества, плюрализма и направленной эволюции. Конференция в Люксембурге может стать важной вехой в развитии, подобно тому, какой стал знаменитый доклад десять лет тому назад, однако в противоположном направлении» (Zemf 1977)².

ПРИЛОЖЕНИЕ 2

В настоящем приложении содержится восемь отредактированных высококвалифицированными редакторами Всесоюзного центра переводов экземпляров одного текста, перевод которого был выполнен на ЭВМ системой АМПАР. Текст взят из бюллетеня ЮНИСИСТ, в составлении схем анализа и перевода не использовался, пре-редактированию не подвергался. Каждый из девяти редакторов редактировал свой текст независимо, какого-либо специального инструктажа для редактирования произведено не было, редакторы действовали по своей обычной процедуре. Один из редакторов вместо отредактированного текста представил замечания на отдельном листе.

Приложение демонстрирует:

— большой разброс требований человека к качеству машинного перевода;

— пригодность перевода, выдаваемого системой АМПАР, для последующего редактирования.

Как можно заметить, предложенный для постредактирования отрывок содержит сравнительно небольшое число ошибок принципиального, «логического», характера, искажающих смысл сообщения, но зато довольно много ошибок в виде неправильных окончаний, нарушения согласования по роду, числу, падежу и пр. Практика использования редактирования машинного перевода в промышленном режиме, подобно тому как это имеет место во Всесоюзном Центре переводов или в организациях

Европейского экономического сообщества, использующих систему СИСТРАН, показала, что редактирование «глупых» ошибок машины в течение достаточно долгого рабочего времени представляет собой большую нагрузку на редактора. Эта работа чрезвычайно утомительна и через некоторое время вызывает большое раздражение, такое, что в целом ряде случаев редактор предпочитает просто-напросто заново перевести текст, обратившись к языку оригинала и не читая машинного перевода. Существует, таким образом, определенный психологический барьер, достаточно серьезный, чтобы его игнорировать. Выходом из положения, безусловно, может явиться существенное повышение качества машинного перевода. Однако, поскольку такое повышение пока что не может быть достигнуто в сколь-нибудь значительных пределах, реальным способом применения МП является обход этого психологического барьера, а именно — применение машинного перевода в тех системах, где он достаточен как неотредактированный продукт в качестве лишь сигнальной информации. Автоматический морфологический редактор также может явиться достаточно эффективным выходом из положения в тех случаях, когда «человеческое» редактирование неизбежно.

² Имеется в виду доклад АЛЬПАК (американской комиссии по вопросам машинного перевода), который отрицательно повлиял на дальнейшее развитие машинного перевода. (См. Язык и машины 1967).

НОМЕР ТЕКСТА 5315

ИМЯ : :3

ДАТА ВВОДА ТЕКСТА 07/09/78

AT THE INITIATIVE OF THE EUROPEAN CO-ORDINATION CENTRE FOR RESEARCH AND DOCUMENTATION IN THE SOCIAL SCIENCES (VIENNA CENTRE), THE CONFERENCE WAS CONVENED FROM 22 TO 25 JUNE 1977 IN MOSCOW UNDER THE AUSPICES OF UNESCO AND HOSTED BY THE INSTITUTE OF SCIENTIFIC INFORMATION ON SOCIAL SCIENCES (UNION) OF THE ACADEMY OF SCIENCES OF THE USSR, FORTY-FIVE DELEGATES REPRESENTING NINETEEN EUROPEAN COUNTRIES AND CANADA, AS WELL AS NINE INTERNATIONAL ORGANIZATIONS ATTENDED THE MEETING.

THIS FIRST EUROPEAN CONFERENCE ON INFORMATION AND DOCUMENTATION IN SOCIAL SCIENCES WAS CONVENED AT A MOST PROPITIOUS TIME WHEN THE HELSINKI CONFERENCE ON SECURITY AND CO-OPERATION IN EUROPE HAS OPENED THE WAY FOR COMMUNICATION AND CO-OPERATION AMONG THE COUNTRIES OF THE REGION, THE DELEGATES FELT THAT THERE WAS NO BETTER WAY TO IMPLEMENT THE PRINCIPLES OF THE HELSINKI FINAL ACT THAN BY ENSURING THE EXCHANGE OF INFORMATION AND CO-OPERATION IN MATTERS OF DOCUMENTATION, PARTICULARLY IN THE FIELDS OF THE SOCIAL SCIENCES, FOR MUTUAL UNDERSTANDING AND PEACE. PARTICIPANTS IN THE CONFERENCE NOTED WITH SATISFACTION THE EXISTENCE OF THE NEED, AS WELL AS GOODWILL TO CONTINUE SEARCHING FOR PRACTICAL WAYS TO CO-OPERATE IN THE FIELD OF INFORMATION. THEY ALSO RECOGNIZED THE VALUABLE WORK OF THE MANY INTERNATIONAL BODIES, GOVERNMENTAL AND NON-GOVERNMENTAL, ALREADY STRIVING FOR IMPROVEMENTS IN THIS FIELD AND RESOLVED TO MAINTAIN CLOSE LINKS WITH THESE BODIES. MUTUAL ACKNOWLEDGEMENT OF THE SITUATION IN EACH COUNTRY IS A USEFUL AND NECESSARY STEP IN THE RIGHT DIRECTION, AND IT MAY SERVE AS A POSITIVE EXAMPLE DURING THE BELGRADE MEETING.

НОМЕР ТЕКСТА 5315

ИМЯ : :3

ДАТА ВВОДА ТЕКСТА 07/09/78

В ИНИЦИАТИВЕ ЕВРОПЕЙСКОГО ЦЕНТРА ДЛЯ ИССЛЕДОВАНИЙ И ДОКУМЕНТАЦИИ В ОБЩЕСТВЕННЫХ НАУКАХ

(ВЕННА // СОСРЕДОТОЧИВАЮТ), СОВЕЩАНИЕ БЫЛО СОБРАНО ОТ 22 ДО 25 ИЮНЯ 1977 В МОСКОВЕ // ПОД ПОКРОВИТЕЛЬСТВОМ ЮНЕСКО // И ИНСТИТУТОМ НАУЧНОЙ ИНФОРМАЦИИ ОТНОСИТЕЛЬНО ОБЩЕСТВЕННЫХ НАУКАХ (ИЮНИОН //) АКАДЕМИИ НАУК СССР // СОРОКА - ПЯТИ ДЕЛЕГАТОВ ПРЕДСТАВЛЯЮЩИХ ДЕВЯТНАДЦАТЬ ЕВРОПЕЙСКИХ СТРАН И (КАНАДА /), КАК ХОРОШО КОГДА ДЕВЯТЬ МЕЖДУНАРОДНЫХ ОРГАНИЗАЦИЙ ПРИСУТСТВУЮЩИХ ВСТРЕЧАЛИСЬ С .

ЭТОГО ПЕРВОГО ЕВРОПЕЙСКОГО СОВЕЩАНИЯ ПО ИНФОРМАЦИИ И ДОКУМЕНТАЦИИ В ОБЩЕСТВЕННЫХ НАУКАХ БЫЛИ СОБРАНЫ В НАИБОЛЕЕ БЛАГОПРИЯТНОЕ ВРЕМЯ КОГДА СОВЕЩАНИЕ (HELSINKI // ПО БЕЗОПАСНОСТИ И СОТРУДНИЧЕСТВУ //) В ЕВРОПЕ // ОТКРЫЛ ПУТЬ ДЛЯ СВЯЗИ И СОТРУДНИЧЕСТВА - СРЕДИ СТРАН РАЙОНА. ДЕЛЕГАТОВ ОШУТИЛИ ЧТО ТАМ НЕ БОЛЕЕ ХОРОШИ ПУТЬ ОСУЩЕСТВИТЬ ПРИНЦИПЫ (HELSINKI // ЗАКЛЮЧИТЕЛЬНЫЙ АКТ ЧЕМ ОБЕСПЕЧИВАЯ ОБМЕН ИНФОРМАЦИЕЙ И СОТРУДНИЧЕСТВОМ (ИЮНИОН //) ВОПРОСАМ ДОКУМЕНТАЦИИ, ОСОБЕННО В ОБЛАСТЯХ ОБЩЕСТВЕННЫХ НАУК, ИБО ВЗАИМНОГО ПОНИМАНИЯ И МИРА. УЧАСТНИКОВ НА СОВЕЩАНИИ ОТМЕЧАЛОСЬ С УДОВЛЕТВОРЕНИЕМ СУЩЕСТВОВАНИЕ НЕОБХОДИМОСТИ ДОБРОЖЕЛАТЕЛЬНОСТИ ПРОДОЛЖАТЬ ИСКАТЬ ПРАКТИЧЕСКИЕ ПУТИ СОТРУДНИЧЕСТВА В ОБЛАСТИ ИНФОРМАЦИИ. ОНИ ТАКЖЕ ПРИЗНАЛИ ЦЕННУЮ РАБОТУ МНОГИХ МЕЖДУНАРОДНЫХ ОРГАНИЗАЦИЙ, ПРАВИТЕЛЬСТВЕННЫХ И НЕПРАВИТЕЛЬСТВЕННЫХ, УЖЕ БОЮЩИХСЯ ДЛЯ УЛУЧШЕНИЯ В ЭТОЙ ОБЛАСТИ И РАЗРЕШИЛИ ПОДДЕРЖАТЬ ТЕСНЫЕ СВЯЗИ С ЭТИМИ ТЕЛАМИ. ВЗАИМНОЕ ПОДТВЕРЖДЕНИЕ ОБСТАНОВКИ В КАЖДОЙ СТРАНЕ ПОЛЕЗНЫЙ И НЕОБХОДИМЫЙ ШАГ В ЭТОМ НАПРАВЛЕНИИ, И МОЖЕТ ПОСЛУЖИТЬ В КАЧЕСТВЕ ПОЗИТИВНОГО ПРИМЕРА ВО ВРЕМЯ ЗАСЕДАНИЯ (BELGRADE //).

НОМЕР ТЕКСТА 5315 ИМЯ : :3

ДАТА ВВОДА ТЕКСТА 07/09/78

По ^{координационного} по

В ИНИЦИАТИВЕ ЕВРОПЕЙСКОГО ЦЕНТРА ДЛН ИССЛЕДОВАНИИ И ДОКУМЕНТАЦИИ В ОБЩЕСТВЕННЫХ НАУКАХ

ВЕНА, ЦЕНТР ^{по} ВЕНА // СОСРЕДОТОЧИВАЮТ, БОРЯЩИЕ БЫЛ СОБРАН ОН 22/25 ИЮНЯ 1977 В ((MOSCOW // ПОД ПОЖЕРТВЕНЬ-
 СОСТОЯЛАСЬ КОНФЕРЕНЦИЯ КОДЕСКО ДЛН ШТАТЛИТЕ ПО, ПР МЕЖДУ НАРОДАМИ
 ТВАМИ ((СНЕСОЕ // И ИНИЦИАТИВУ НАУЧНОЙ ИНФОРМАЦИИ ФУНДОМЕНТНО-ОБЩЕСТВЕННЫХ НАУКАХ ((((INION //)) АКАД
 СОВР 45

ЕМЛЮ НАУК ((URSS //, ВОРОКА - ПЯТИ ДЕЛЕГАТОВ, ПРЕДСТАВЛЯЮЩИХ ДЕВЯТНАЦАТЬ ЕВРОПЕЙСКИХ СТРАН И ((CANADA
 КРОМЕ ТОГО, БЫЛО ПРЕДСТАВЛЕНА У МЕЖДУНАРОДНЫХ ДОГАВОРОВ
 , КАК ХОРОШО КОГДА ДЕЙСТВИИ МЕЖДУНАРОДНЫХ ОРГАНИЗАЦИЙ ПРИСУТСТВУЮЩИХ ВОТРЕГНАЦИОНАХ С-1.

ЭТОГО ПЕРВОГО ЕВРОПЕЙСКОГО СОВЕЩАНИЯ ПО ИНФОРМАЦИИ И ДОКУМЕНТАЦИИ В ОБЩЕСТВЕННЫХ НАУКАХ БЫЛ
 СОЗДАНО ^{в Кельсик} РЕШАНЫ В НАИБОЛЕЕ БЛАГОПРИЯТНОЕ ВРЕМЯ КОГДА СОВЕЩАНИЕ ((HELSINKI // ПО БЕЗОПАСНОСТИ И СОТРУДНИЧЕСТВА
 В ЕВРОПЕ // ОТКРЫЛ ПУТЬ ДЛЯ СВЯЗИ И СОТРУДНИЧЕСТВА - БРЕНДИ-СТРАН РАЙОНА. ДЕЛЕГАТОВ ОУЩАРИЛИ ЧТО
 ТАМ НЕ БОЛЕЕ ХОРОШИИ ПУТЬ ОСУЩЕСТВИТЬ ПРИНЦИПЫ ((HELSINKI // ЗАКЛЮЧИТЕЛЬНИИ АКТУ ДЕМ ОБЕСПЕЧИВАЮТ ОБМЕН
 ОБМЕН ^{по} ИНФОРМАЦИИ И СОТРУДНИЧЕСТВА ((-НА) ВОПРОСАМ ДОКУМЕНТАЦИИ, ОСОБЕННО В ОБЛАСТЯХ ОБЩЕСТВЕННЫХ НАУК,
 В ЦЕЛЯХ ^{и отмечая} К НЕО ВЗАИМНОГО ПОНИМАНИИ И МИРА. УЧАСТНИКОВ НА СОВЕЩАНИИ ОМЕРЧАЕМ С УЛОВЛЕТВОРЕНИЕМ СУЩЕСТВУЮЩИМ
 К ПРОДВИЖЕНИЮ ПОСКОЕ

НАУКАМ ХОРОШО КОГДА ДОБРОСОВЕЩАТЕЛЬНОСТИ ПРОДОЛЖАТЬ ИСКАТИ ПРАКТИЧЕСКИИ ПУТИ СОТРУДНИЧАТИ В ОБЛАС -
 И ПРОДВИЖЕНИЮ ДЛН ЭТОМ ДОБРО О ВОНА

ТИ ИНФОРМАЦИИ ОНИ ТАКЖЕ ПРЕЖДАЛИ ПРНУЮ РАБОТУ МНОГИХ МЕЖДУНАРОДНЫХ ОРГАНИЗАЦИЙ, ПРАВИТЕЛЬСТВЕННЫХ И
 НЕ - ПРАВИТЕЛЬСТВЕННЫХ, УЖЕ БОРЯЩИИ ДЛН УЛУЧШЕНИЯ В ЭТОЙ ОБЛАСТИ И РАЗРЕШИТИ ПОДДЕРЖАТЬ ТЕСНЫЕ СВЯЗИ С
 ДОГАВОРОВЩИИМИ, ОДИКОМЕРНИИ ОУ
 ЭТИМИ ЦЕЛЯМИ, ВЗАИМНОЕ ПОДДЕРЖАНИЕ ОБСТАНОВКИ В КАЖДОЙ СТРАНЕ-ПОЛЕЗНЫИ И НЕОБХОДИМЫИ ШАГ В ЭТОМ НА -
 ПРАВЛЕНИИ, И МОЖЕТ ПОСЛУЖИТЬ В КАЧЕСТВЕ ПОЗИТИВНОГО ПРИМЕРА ВО ВРЕМЯ ЗАСЕДАНИИ ((ВЕЛЕНАДИИ).

НОМЕР ТЕКСТА 5315

ИМЯ

: :3

ДАТА ВВОДА ТЕКСТА 07/09/78

По ^{координационного} по

В ИНИЦИАТИВЕ ЕВРОПЕЙСКОГО ЦЕНТРА ДЛН ИССЛЕДОВАНИИ И ДОКУМЕНТАЦИИ В ОБЩЕСТВЕННЫХ НАУКАХ (ВЕНО)

((-ЧЕВНА // СОСРЕДОТОЧИВАЮТ), СОВЕЩАНИЕ БЫЛ СОБРАН ОН 22/25 ИЮНЯ 1977 В ((MOSCOW // ПОД ПОЖЕРТВЕНЬ-
 ЕВРАМИ ((СНЕСОЕ // И ИНСТИТУТОМ НАУЧНОЙ ИНФОРМАЦИИ ФУНДОМЕНТНО-ОБЩЕСТВЕННЫХ НАУКАХ ((((INION //)) АКАД
 ЕМЛЮ НАУК ((СНЕСОЕ //, СОРОКА - ПЯТЬ ДЕЛЕГАТОВ, ПРЕДСТАВЛЯЮЩИХ ДЕВЯТНАЦАТЬ ЕВРОПЕЙСКИХ СТРАН И ((CANADA
 , КАК ХОРОШО КОГДА ДЕЙСТВИИ МЕЖДУНАРОДНЫХ ОРГАНИЗАЦИЙ ПРИСУТСТВУЮЩИХ ВОТРЕГНАЦИОНАХ С-1.

ЭТОГО ПЕРВОГО ЕВРОПЕЙСКОГО СОВЕЩАНИЯ ПО ИНФОРМАЦИИ И ДОКУМЕНТАЦИИ В ОБЩЕСТВЕННЫХ НАУКАХ БЫЛ
 СОЗДАНО ^{в Кельсик} РЕШАНЫ В НАИБОЛЕЕ БЛАГОПРИЯТНОЕ ВРЕМЯ КОГДА СОВЕЩАНИЕ ((HELSINKI // ПО БЕЗОПАСНОСТИ И СОТРУДНИЧЕСТВА
 В ЕВРОПЕ // ОТКРЫЛ ПУТЬ ДЛЯ СВЯЗИ И СОТРУДНИЧЕСТВА - БРЕНДИ-СТРАН РАЙОНА. ДЕЛЕГАТОВ ОУЩАРИЛИ ЧТО
 ТАМ НЕ БОЛЕЕ ХОРОШИИ ПУТЬ ОСУЩЕСТВИТЬ ПРИНЦИПЫ ((HELSINKI // ЗАКЛЮЧИТЕЛЬНИИ АКТУ ДЕМ ОБЕСПЕЧИВАЮТ ОБМЕН
 ОБМЕН ^{по} ИНФОРМАЦИИ И СОТРУДНИЧЕСТВА ((-НА) ВОПРОСАМ ДОКУМЕНТАЦИИ, ОСОБЕННО В ОБЛАСТЯХ ОБЩЕСТВЕННЫХ НАУК
 В ЦЕЛЯХ ^{и отмечая} К НЕО ВЗАИМНОГО ПОНИМАНИИ И МИРА. УЧАСТНИКОВ НА СОВЕЩАНИИ ОМЕРЧАЕМ С УЛОВЛЕТВОРЕНИЕМ СУЩЕСТВУЮЩИМ
 К ПРОДВИЖЕНИЮ ПОСКОЕ

НАУКАМ ХОРОШО КОГДА ДОБРОСОВЕЩАТЕЛЬНОСТИ ПРОДОЛЖАТЬ ИСКАТИ ПРАКТИЧЕСКИИ ПУТИ СОТРУДНИЧАТИ В ОБЛАС -
 И ПРОДВИЖЕНИЮ ДЛН ЭТОМ ДОБРО О ВОНА

ТИ ИНФОРМАЦИИ. ОНИ ТАКЖЕ ПРЕЖДАЛИ ПРНУЮ РАБОТУ МНОГИХ МЕЖДУНАРОДНЫХ ОРГАНИЗАЦИЙ, ПРАВИТЕЛЬСТВЕННЫХ И
 НЕ - ПРАВИТЕЛЬСТВЕННЫХ, УЖЕ БОРЯЩИИ ДЛН УЛУЧШЕНИЯ В ЭТОЙ ОБЛАСТИ И РАЗРЕШИТИ ПОДДЕРЖАТЬ ТЕСНЫЕ СВЯЗИ С
 ДОГАВОРОВЩИИМИ, ОДИКОМЕРНИИ ОУ
 ЭТИМИ ЦЕЛЯМИ, ВЗАИМНОЕ ПОДДЕРЖАНИЕ ОБСТАНОВКИ В КАЖДОЙ СТРАНЕ-ПОЛЕЗНЫИ И НЕОБХОДИМЫИ ШАГ В ЭТОМ НА -
 ПРАВЛЕНИИ, И МОЖЕТ ПОСЛУЖИТЬ В КАЧЕСТВЕ ПОЗИТИВНОГО ПРИМЕРА ВО ВРЕМЯ ЗАСЕДАНИИ ((ВЕЛЕНАДИИ).

- Автоматический перевод. М.: Прогресс, 1971.— 368 с.
- Автоматический перевод 1949—1963. Критико-библиографический справочник. М.: ВИНТИ. 1967.— 517 с.
- Алексеев П. М. Частотный словарь английского подязыка электроники: Дис. на соиск. ученой степени канд. филолог. наук.— Л., Ленингр. ун-т, 1964.— 156 с.
- Апресян Ю. Д., Богуславский И. М., Номдин Л. Л. и др. Лингвистическое обеспечение в системе автоматического перевода третьего поколения. АН СССР, препринт научного совета по проблеме «Кибернетика». М., 1978.— 47 с.
- Большой англо-русский словарь. Под общ. руков. И. Р. Гальперина. М.: Сов. энциклопедия, 1972, ч. 1—822 с., ч. 2—863 с.
- Бар-Хиллел И. Будущее машинного перевода.— Филол. науки, 1962, № 4, с. 205—211.
- Бархударов Л. С. Язык и перевод.— М.: Международные отношения, 1975.— 239 с.
- Бельская И. К. Язык человека и машина.— М.: Изд. Моск. ун-та, 1969.— 409 с.
- Берков В. П. Вопросы двуязычной лексикографии.— Л.: Изд. Ленингр. ун-та, 1973.— 190 с.
- Бут А. Д., Локк У. Н. Историческое введение.— В кн.: Машинный перевод. М., 1957.— с. 15—32.
- Вардуль И. Ф. Об изучении семантического аспекта языка.— Вопросы языкознания, 1973, № 6, с. 9—21.
- Василевский А. Л., Марчук Ю. Н. Вычислительная лингвистика.— М., МГПИИЯ им. М. Тореца, 1970, ч. 1—2—147 с., ч. 3—4—256 с.
- Вишнякова С. М. Многозначность слов естественного языка и методы ее устранения при автоматическом индексировании: Автореф. дисс. на соиск. ученой степени канд. филолог. наук АН БССР, Ин-т языкознания. Минск, 1973.— 26 с.
- Гарвин П. Алгоритм синтаксического анализа «Фулькрум».— В кн.: Автоматический перевод. М., 1971, с. 26—40.
- Глазерфельд, фон Э. «Мультистор» — система корреляционного анализа для английского языка.— В кн.: Автоматический перевод. М., 1971, с. 281—318.
- Денисов П. Н. Очерки по русской лексикологии и учебной лексикографии. М.: Изд. Моск. ун-та, 1974.— 254 с.
- Денисов П. Н. Принципы моделирования языка.— М.: Изд. Моск. ун-та, 1965.— 204 с.
- Доклад временной научно-технической комиссии для разработки предложений по дальнейшему развитию работ в области машинного перевода. ГКНТ СМ СССР. М., 1973.— 33 с.
- Драмбян Л. В., Истомина И. М., Семенова Е. К. и др. Особенности программной реализации некоторых этапов процесса перевода в системе АМПАР.— В кн.: Современное состояние машинного перевода и автоматизации информационных процессов. М., 1977, с. 59—65.
- Звезинцев В. А. Предложение и его отношение к языку и речи. М.: Изд. Моск. ун-та, 1976.— 307 с.
- Зубов А. В. Переработка текста естественного языка в системе «человек—машина».— В кн.: Статистика речи и автоматический анализ текста. Л., 1971, с. 286—434.
- Иванов В. В., Топоров В. Н. Постановка задачи реконструкции текста и реконструкции знаковой системы.— В кн.: Структурная типология языков. М., 1966, с. 3—25.
- Ингве В. Значение исследований в области машинного перевода.— НТИ, 1965, № 7, с. 42—47.
- Ингве В. Г. Синтаксис и проблема многозначности.— В кн.: Машинный перевод. М., 1957, с. 281—304.
- Комиссаров В. Н. Слово о переводе. М.: ИМО, 1973.— 237 с.
- Контекстологический словарь для машинного перевода многозначных слов английского языка на русский. Сост. Ю. Н. Марчук. М.: ВЦП, 1976, ч. 1—264 с., ч. 2—256 с.
- Котелова Н. З. Значение слова и его сочетаемость. Л., Наука, 1975.— 162 с.
- Котов Р. Г. Лингвистика и современное состояние машинного перевода в стране.— Вопросы языкознания, 1976, № 5, с. 37, 49.
- Кулагина О. С. Исследования по машинному переводу. М.: Наука, 1979.— 319 с.
- Куно С., Эттингер А. Многовариантный синтаксический анализатор.— В кн.: Автоматический перевод. М., 1971, с. 102—180.
- Лейкина Б. М., Никитина Т. Н., Откупщикова М. И. и др. Система автоматического перевода, разрабатываемая в группе математической лингвистики ВЦ ЛГУ.— НТИ, 1966, № 1, с. 40—50.
- Лемб С. Стратификационная лингвистика как основа машинного перевода.— НТИ, 1964, № 10, с. 33—40.
- Леонтьева Н. Н., Никодосов С. Л. Контуры анализа в системе французско-русского машинного перевода.— В кн.: Современное состояние машинного перевода и автоматизации информационных процессов. М., 1977, с. 49—58.
- Мак Дэниел Дж., Прайс У. Л., Зансер Дж. М., Нейтс Д. М. Оценка возможности использования переводов, полученных на машине в Национальной физической лаборатории (Теддингтон) и краткое описание алгоритма перевода.— В кн.: Автоматический перевод. М., 1971, с. 234—249.
- Марчук Ю. Н. Опыт машинной реализации дистрибутивной методики определения лексических значений.— В кн.: Статистика речи и автоматический анализ текста. 1972. Л., 1973, с. 181—232.
- Марчук Ю. Н. О последовательности омографов. Деп. рукопись.— Сб. рефератов деп. рукоп. ВИМИ, 1973а, № 14, 11 с.
- Марчук Ю. Н. Характеристики частотного словаря английских газетных текстов. Деп. рукопись.— Сб. рефератов деп. рукоп. ВИМИ, вып. II, 1973б, 21 с.
- Марчук Ю. Н. Вычислительная лексикография.— М.: ВЦП, 1976.— 183 с.
- Марчук Ю. Н. Синтактико-семантический анализ в системе машинного перевода АМПАР.— Международный форум по информатике и документации. М.: ВИНТИ. 1980, том 5, № 2, с. 15—17.
- Марчук Ю. Н. О моделировании перевода.— Вопросы информационной теории и практики. М.: ВИНТИ, 1980, № 43, с. 78—86.
- Марчук Ю. Н., Тихонова Н. Г., Убин И. И. Методика составления семантических частотных словарей в системе СЛОК.— Тезисы докладов Международного семинара по машинному переводу. М., 1975, с. 86—89.
- Марчук Ю. Н., Тихомиров Б. Д., Щербинин В. И. Система машинного перевода с английского языка на русский.— В кн.: Машинный перевод и автоматизация информационных процессов. М., 1975, с. 18—33.
- Марчук Ю. Н., Власов А. Н. Некоторые принципы автоматизации перевода с немецкого языка на русский.— Zeitschrift «Fremdsprachen», 1980, № 2, с. 91—99.
- Мастерман М. Тезаурус в синтаксисе и семантике.— В кн.: Математическая лингвистика. М., 1964, с. 160—176.
- Мастерман М. Полуавтоматический перевод с английского языка на французский. Система «человек—машинный тезаурус».— В кн.: Автоматический перевод. М., 1971, с. 264—280.
- Молошная Т. Н. Алгоритм перевода с английского языка на русский.— Проблемы кибернетики, вып. 3. М.: Наука, 1960, с. 209—272.
- Москович В. А., Вишнякова С. М. Оценка качества автоматического перевода.— НТИ, сер. 2, 1970, № 6, с. 32—36.
- Моторин Ю. А., Марчук Ю. Н. Реализация автоматического перевода на современных серийных ЭВМ общего назначения.— Вопросы радиоэлектроники, сер. ЭВТ, 1970, вып. 7, с. 20—29.
- Общее и прикладное языкознание. Библиографический указатель. 1963—1967. М.: Наука, 1972.— 295 с.

- Овчинников В. Г.* Автоматизированные системы информационного обеспечения. М.: Энергия, 1977.—256 с.
- Паков Д. Ю., Ляпунов А. А., Мухин И. С.* Автоматизация перевода с одного языка на другой. М.: Изд. АН СССР, 1956.—22 с.
- Паркер-Родс А. Ф.* О последних работах в области машинного перевода методом тезауруса и языка-посредника.— В кн.: Математическая лингвистика. М., 1964, с. 229—245.
- Певзнер Б. Р.* Влияние неразличения омонимии на уровень поискового шума.— НТИ, сер. 2, 1973, № 10, с. 15—20.
- Пешковский А. М.* Избранные труды. М., 1959.—248 с.
- Пиотровский Р. Г.* Текст, машина, человек.— Л.: Наука, 1975.—327 с.
- Пиотровский Р. Г.* Инженерная лингвистика и теория языка.— Л.: Наука, 1979.—111 с.
- Пушкин В. Н.* Принципы математического моделирования познания и пути построения систем искусственного интеллекта.— В кн.: Вопросы кибернетики. Ситуационное управление. Теория и практика. М.: Изд. научного совета по проблеме «Кибернетика» АН СССР, 1980, с. 5—33.
- Резин Н. И., Розенцвейг В. Ю.* Основы общего и машинного перевода.— М.: Высшая школа, 1964.—242 с.
- Рецкер Я. И.* Теория перевода и переводческая практика.— М.: Международные отношения, 1974.—215 с.
- Рождественский Ю. В.* Типология слова.— М.: Высшая школа, 1969.—321 с.
- Рождественский Ю. В.* Введение в общую филологию.— М.: Высшая школа, 1979.—223 с.
- Таббе М.* Вычислительные машины и здравый смысл.— М.: Прогресс, 1964.—182 с.
- Трибис Л. И.* Об одной модели распознавания лексических значений неоднозначных слов.— В кн.: Статистика речи и автоматический анализ текста 1972. Л., 1973, с. 131—180.
- Уивер У.* Перевод.— В кн.: Машинный перевод. М., 1957, с. 33—44.
- Фреге Г.* Смысл и денотат.— Семантика и информатика, вып. 8. М., 1977, ВИНТИ, с. 181—210.
- Чернышев А. Г.* Омонимия лексико-грамматических подклассов «процессных» и «предметных» существительных и методы ее формального разрешения. Автореф. дис. на соиск. учен. степени канд. филолог. наук. М., МГПИИЯ, 1970.—23 с.
- Шалапина З. М.* Семантико-синтаксический анализ в системе англо-русского автоматического перевода (АРАП). АН СССР, Ин-т русского языка, предв. публ. 1974, вып. 47, 37 с.
- Швейцер А. Д.* Перевод и лингвистика.— М.: Воен. изд-во, 1973.—279 с.
- Шрейдер Ю. А.* Равенство, сходство, порядок.— М.: Наука, 1971.—255 с.
- Шерба Л. В.* Избранные работы по языкознанию и фонетике, т. 1. Л., 1958.—431 с.
- Язык и машины (Применение электронных вычислительных машин в переводе и в исследовании языка. Отчет Наблюдательного комитета по автоматической обработке текстов. Национальная Академия наук США. Национальный научно-исследовательский совет).— НТИ. Информационные процессы и системы, сер. 2, 1968, № 8, с. 25—37.

OLB—Third European Congress of Information Systems and Networks. Overcoming the Language Barriers. Luxembourg, 3—6 May, 1977, v. 1—2. K. G. Saur. London—Paris—New York, 1978.

- Bloncourt-Herselin J.* Dictionnaire Italien-Français, Français-Italien. Paris. Garnier-Flammariion, 1964. 413 p.
- Boilet C.* Where does GETA stand at the beginning of 1977?—OLB, v. 2, p. 88—120.
- Boitinger D.* The automatization of Meaning.—Language, 1965, v. 41, N 4, p. 555—573.

- Bruderer H.* Handbuch der maschinellen und maschinunterstützten Sprachübersetzung. Verlag Dokumentation Saur KG München—New York, 1978.—864 s.
- Bruderer H.* The Present State of Machine and Machine-Assisted Translation.—OLB, V. 1, p. 529—556.
- Catford J.* A linguistic theory of Translation. Oxford, 1965.—216 p.
- Chapin P. G.* Report on Seminar on Machine Translation. Rosslyn, Virginia, 1976.—3 p.
- Chandioux J.* Creation of a Second Generation System for Machine Translation of Technical Manuals.—OLB, V. 1, p. 613—621.
- Chaumier J.* Assessment of SYSTRAN.—OLB, V. 2, p. 74—82.
- Dagobert S.* A universal Source Thesaurus as a Classification Generator. J. Amer. Soc. Inform. Sci. 1972, 23, № 5, p. 299—305.
- Dostert B. H.* User's Evaluation of Machine Translation. Calif. In-t of Technology, 1973.—106 p.
- Euvard A., Lecomte J.* Elaboration d'une chaine de Traduction automatique d'Anglais en Français. Cahiers du C. R. A. L. Num. 36. Nancy, 1979.—157 p.
- Gumb R. D.* The Inapplicability of the Dictionary—Encyclopedia Distinction to the Semantics of Natural Languages. Information Storage and Retrieval, 1971, V. 6, N 6, p. 445—452.
- Harris Z. S.* La structure distributionelle. Langages, Paris, 1970, p. 14—34.
- Jäger G.* Translation and Translationlinguistik. Leipzig, Max Niemeyer Verlag Halle (Salle), 1975.—214 S.
- Jäger G.* Zu Gegenstand und Zielen der Übersetzungswissenschaft. Übersetzungswissenschaftliche Beiträge. I. Vermittelte Kommunikation. Leipzig, VEB Verlag Enzyklopädie, 1977, S. 14—26.
- Jordan S. R., Brown A. F., Hutton F. C.* SLC Primer for Russian Translation Users. 1977, ORNL/CSD-24. Nuclear Div., Union Carbide Corp.—73 p.
- Jordan S. R., Brown A. F. R., Hutton F. C.* Computerized Russian Translation at ORNL.—Journal of the American Society for Information Science, 1977, Jan., p. 26—33.
- Kade O.* Kommunikationswissenschaftliche Probleme der Translation.— In: Grundlagen der Übersetzungswissenschaft (Beihefte zur Zeitschrift Fremdsprachen). Leipzig, 1968, S. 148—231.
- Kahn D.* The Codebreakers. New York, MacMillan, 1968.—1164 p.
- Kaplan A.* An experimental Study of Ambiguity and Context. Mechanical Translation 1955, V. 2, N 2, p. 39—46.
- Ljudskanov A.* Mensch+Maschine als Übersetzer. Halle (Salle), VEB Max Hueber Verlag, 1972.—260 S.
- Loh S. C., Kong L.* Computer Translation of Chinese Scientific Journals.—OLB, V. 1, p. 631—645.
- Madhu S., Lytle D. W.* A figure of Merit Technique for the Resolution of Non-Grammatical Ambiguity. Mechanical Translation, 1965, V. 8, N 2, p. 9—13.
- Marchuk Ju. N.* Machine Translation in the All-Union Translation Center.—OLB, V. 1, p. 623—624.
- Marchuk Ju. N.* Some General Features of Machine Dictionaries. Preprints of Coling 76, Ottawa, 28 June—2 July 1976, Vortrag Nr 37, 2S.
- Marchuk Ju. N.* The Contextological Dictionary: Use in Programmed Language Teaching. Computers and Humanities. New York, 1979, V. 13, N 4, p. 277—281.
- Menzerat P., Meyer-Eppler W.* Sprachuntersuchungen. Studia Linguistica, 1950, Kopenhagen, IV, N 1—2, p. 54—93.
- Nicholson M.* A Dictionary of American-English Usage. New York, 1958.—671 p.
- Nida E.* Science of Translation. Language, 1969, V. 45, N 3, p. 451—521.
- Oettinger A. G.* Automatic Language Translation. Cambr., Mass., 1960.—380 p.
- Panzer B.* Kontextdetermination und Morphembedeutung. Beitr. Sprachkunde und Informationsverarbeitung. 1965, № 7, p. 7—28.
- Pfafflin S. M.* Evaluation of Machine Translations by Reading comprehension tests and subjective Judgements.—Mechanical Translation 1965, V. 8, № 2, p. 2—8.

Предисловие	3
Глава 1. О современном состоянии машинного перевода	9
1.1. Два основных подхода к решению проблемы	9
1.2. Возникновение идеи и периодизация истории машинного перевода	17
1.3. Действующие системы машинного перевода	22
1.4. Требования технологии	24
Глава 2. Лингвистические основания моделирования перевода	30
2.1. Представление о переводе	30
2.2. Объект и цели моделирования	32
2.3. Переводные соответствия и «единица перевода»	35
2.4. Типология переводных соответствий	38
2.5. Переводные соответствия и передача смысла	42
2.6. Языковой знак и перевод	46
2.7. Технология и методика построения модели	52
Глава 3. Модель машинного перевода на основе переводных соответствий	56
3.1. Общая структура модели	56
3.2. Общие принципы алгоритма анализа и синтеза	60
3.3. Принцип системной организации динамического компонента	68
3.4. Лексика. Словарь	74
3.5. Грамматика	81
3.6. Семантика	86
3.7. Переводной элемент предметного компонента	88
3.8. Динамический компонент модели	91
3.9. Программный элемент динамического компонента	104
3.10. Эффективность работы системы и качество перевода без редактирования	115
Глава 4. Проблемы словаря: вычислительная лексикография	126
4.1. Основные отличия машинных словарей от обычных	126
4.2. Предмет, методы и теория вычислительной лексикографии	133
4.3. К проблеме лексической неоднозначности	138
Глава 5. Контекстологический словарь	146
5.1. Принципы построения	146
5.2. Состав и структура словаря	152
5.3. Комплементарные методы	161
5.4. Типизация детерминант и общие свойства КС	164
5.5. Эффективность перевода слов контекстологическим словарем	166
Глава 6. Перспективы машинного перевода	168
6.1. Качество машинного перевода	168
6.2. Роль и место редактора в редактировании МП	177
6.3. Техничко-экономическая эффективность МП	180
6.4. Машинный перевод в системе научно-технической информации	183
6.5. Задачи и перспективы машинного перевода	186
Приложения	193
Литература	228
Принятые сокращения	232

Rado G. Outline of a systematic Translatology. Babel, 1979, V. XXV, № 4, p. 187—196.

Reißler E. Linguistic and Engineering Studies in Automatic Language Translation. Report No RADC-TR-60-11. Univ. of Washington, 1960.— 492 p.

Reimold P. A formal psycholinguistic model of sentence comprehension.— Amer. Journal of Computational Linguistics. New York, 1975, p. 3—46.

RETX — Russian-English Translation Exchange. Translation Research Institute. 5914. Pulask. Av. Philadelphia, USA, 1980.

Robinson H. R. Computer-detectable Semantic Structures. Information Storage and Retrieval. 1970, V. 6, № 3, p. 273—288.

Rolling L. The facts about Automatic Translation.— 6th Annual Canadian Association for Information Science (CAIS) Conference. Montreal (QUEBEC) 1978.— 9 p.

Sankoff D. Dictionary Structure and Probability Measures.— Information and Control, 1971, V. 19, № 2, p. 104—113.

Sereda S., Levesque A. Technical Bulletin, Dep. of Nat. Defence, Canada, 1977.— 4 p.

Tesnière L. Eléments de syntaxe structurale. Paris, 1959.— 212 p.

Toma P. Computer Translation: in its own Right.— Kommunikationsforschung und Phonetik-JPK-Forschungsberichte, 1974, Bd. 50, S. 158—164.

Toma P. SYSTRAN as a Multilingual Machine Translation System — OLB, V. 1, p. 569—581.

Trotke R. The Future of MT.— American Journal of Computational Linguistics. Microfische 51. The Finite String, 1976, V. 13, № 6, p. 47—49.

Vanquois B. Structures profondes et Traduction automatique. Le système du C. E. T. A.— Revue roumaine de Linguistique (Bucaresti), 1968, № 2, XIII, p. 105—129.

Vincent J. Dictionnaire Anglais-Français, Français-Anglais. Paris, Garnier-Flammarion, 1964.— 393 p.

Winograd T. Five lectures on Artificial Intelligence. Linguistic Structure Processing. North-Holland Publ. Co. Amsterdam—New York—Oxford, 1977, V. 5, p. 491—572.

Zemb J.-M. Introduction by the Chairman of session 5.— OLB, V. 2, p. 73.

ПРИНЯТЫЕ СОКРАЩЕНИЯ

АМПАР — автоматизированный машинный перевод с английского языка на русский	НТИ — Научно-техническая информация. Ежемесячный научно-технический сборник. М., ВИНТИ
ВЦП — Всесоюзный центр переводов научно-технической литературы и документации ГКНТ и АН СССР	ОСТ — общий структурный тип
ДК — диагностирующая конструкция	ПМС — перевод многозначных слов
ИЯ — информационная ячейка	ППП (п.п.п.) — подпрограмма пропуска
КМП — контекстуальная модель перевода	ПТ — промежуточный текст
КС — контекстологический словарь	САП — словарь автоматического перевода
МП — машинный перевод	САР — словарь автоматического редактирования
МПС — машинный перевод по переводным соответствиям	СТ — структурный тип
НП — номер перевода	СФ — синтаксическая функция
	ЦЭ — цифровой эквивалент
	ЧСТ — частный структурный тип
	ЭВМ — электронная вычислительная машина