

Robert Szymula
Uniwersytet w Białymstoku

Машинный перевод – основные проблемы

Машинный перевод является интересной теоретической и практической задачей современной лингвистики. Нет сомнений в том, что эта проблема будет исследоваться также в будущем.

Актуальность автоматизации перевода продолжает оставаться высокой. Все время растут и расширяются возможности современной компьютерной информационной технологии, благодаря чему эффективный машинный перевод становится все более реальным. Спрос на перевод разного рода увеличивается соответственно тому, как все больше естественных языков вступает в коммуникационную информационную сферу, а ведь перевод с одного языка на другой – это единственный эффективный способ преодоления существующих языковых барьеров, поскольку другие способы неэффективны и никогда не заменят перевода. Стоит упомянуть также, что высока и научная привлекательность этой идеи.

Термин „перевод” можно понимать двояко: как сам процесс или его результат. Существует большое количество видов переводов и их классификаций (в зависимости от переводного материала, по форме презентации текста перевода и текста оригинала, по цели, с какой делается перевод и т. п.), но для нас самой важной является классификация по степени механизации процесса перевода. По этой классификации можно выделить традиционный перевод, а также три способа перевода с использованием компьютера: перевод текста человеком с использованием компьютера (при данном способе текст переводит человек, а за переводными эквивалентами слов, которые ему неи-

звестны, он обращается к электронным словарям), перевод текста компьютером с использованием человека (при этом способе человек на определенных этапах подключается к процессу перевода текста компьютером – может выступать предредактором, который осуществляет подготовительную обработку текста, интерредактором, к которому компьютер в процессе осуществления перевода может обращаться за справками, или постредактором, который осуществляет полную редакторскую правку переведенного автоматической системой текста), перевод текста компьютером без помощи человека (машинный или автоматический перевод). Машинный или автоматический перевод можно определить как выполняемое компьютером действие по преобразованию текста на одном естественном языке в текст на другом естественном языке при сохранении эквивалентности содержания, а также результат такого действия¹.

Говоря о системах машинного перевода, стоит обратить внимание на некоторые методы, используемые учеными при их составлении.

Несомненно следует упомянуть прямой способ перевода, имеющий место в системах, которые с самого начала ориентированы на перевод с одного определенного языка на другой определенный язык. Такие системы ограничиваются только тем минимумом операций над текстом, который необходим для получения требуемого перевода. Напротив, непрямой способ перевода имеет место в системах, в которых анализ текста на входном языке и синтез текста на выходном языке представляют собой принципиально независимые процедуры.

В свое время ученые часто использовали при разработке систем машинного перевода перевод через язык-посредник. О таком виде перевода говорят в связи с системами, в которых представление смысла входного текста предполагается независимым от какого-либо естественного языка и непосредственно используется для синтеза текста на выходном языке. Идея языка-посредника основана на понятии лингвистических универсалий.

Во многих системах используются также межъязыковые операции. О них принято говорить в тех случаях, когда глубинное представление значения структурно-грамматической единицы может быть различным в зависимости от языка, который был исходным при

¹ Зубова И. И., *Новые технологии в лингвистике*, Минск 2001, с. 87.

анализе рассматриваемой единицы или должен быть выходным при ее синтезе. В таком случае в процесс перевода должен включаться еще один этап (т. е. этап межъязыковых операций), на котором должен происходить переход от представлений значения, специфических для одного языка, к соответствующим им представлениям, специфическим для другого языка. В целом процесс перевода с использованием межъязыковых операций – это последовательное осуществление сначала анализа, затем межъязыковых операций, а после них – синтеза. Перевод через язык-посредник и перевод с использованием межъязыковых операций выступают в системах, для которых характерен непрямой способ перевода, так как в системах, реализующих прямой способ перевода, не предусматривается построение формального представления значения.

Для раннего этапа разработки систем машинного перевода характерен локальный подход к обработке текста. К таким системам относят те, в которых основными единицами анализа являются слова. Анализ фактически сводится в них к выполнению для каждого слова отдельной определяемой им процедуры, направленной на то, чтобы установить, какой частью речи является данное слово, входит ли оно в тот или иной идиоматический оборот, а также на то, чтобы определить его конкретный смысл. Для этого система рассматривает слова, расположенные слева и/или справа от обрабатываемого слова. Построение общей структуры предложения как такового в них не предусматривается.

Обнаружение недостатков алгоритмов локального анализа привело ученых к выводу, что надо заменить его анализом глобальным. К системам, опирающимся на такой подход, относят те системы, в которых значение слова определяется его контекстом в рамках общей структуры предложения. При таком анализе для каждой словоформы выписываются все ее гипотетические связи, а затем проверяется, какие из них реально выступают в данном предложении. Отсеивание несовместимых гипотез проводится при помощи комплекса правил, формулирующих требования к правильным структурам данного языка. Анализ проводится по инструкции, указывающей операции, которые следует проделать, чтобы проверить, соответствует ли анализируемое предложение в целом требованиям, предъявляемым к правильным структурам.

Операции перевода с одного языка на другой являются сложнейшей задачей интеллекта, так как требуют не только замены слов одного языка словами другого, но и передачи мысли автора со всеми

нюансами и тонкостями. Главными элементами, которые надо учитывать при переводе, являются: входной язык (язык, с которого осуществляется перевод), выходной язык (язык, на который переводится текст) и предметное содержание переводимого текста. Ученые пытаются построить процесс машинного перевода на основе моделирования деятельности переводчика-человека. Процесс перевода текста человеком состоит из трех основных этапов: постижения текста на входном языке, его интерпретации и создание текста на выходном языке. Суть первого этапа заключается в дословном понимании текста, его стилистическом понимании и понимании идейного замысла автора. Главной проблемой, связанной с интерпретацией текста на входном языке, является факт, что в двух языках никогда нет полного семантического тождества. Лингвистически точный перевод текста в принципе невозможен, поэтому так важна его правильная интерпретация при помощи средств выходного языка, благодаря чему переводчик в состоянии передать объективный смысл всего произведения. Самым трудным является этап создания текста на выходном языке путем перевыражения исходного текста. Текст перевода должен сохранить предметное содержание текста на входном языке и его идейный замысел при помощи стилистически верной замены предложений исходного языка предложениями выходного языка.

Из всего этого видно, что очень трудно подражать такому во многом неясному процессу. Его сложность создает большие трудности в ходе построения алгоритмов машинного перевода, поэтому эта задача до сих пор не имеет однозначного решения. В ходе исследований ученые пришли к выводу, что процесс машинного перевода должен базироваться на процессах анализа и синтеза применительно к входному и выходному языкам, т. е. программа, осуществляющая перевод текстов, должна выполнять следовательно морфологический анализ каждого слова предложения входного языка (в ходе которого каждое слово получает наборы лексико-грамматических признаков, сформированные либо по формальным признакам слов, либо с опорой на специальный автоматический словарь), затем синтаксический анализ каждого предложения исходного текста (который сводится к поиску основных членов предложения), синтаксический синтез каждого предложения перевода (задачей которого является не только создание предложения перевода согласно имеющимся в памяти компьютера синтаксическим правилам выходного языка и структурой текста оригинала, но и замена слов входного языка их переводными эквивалентами) и в конце морфологи-

ческий синтез каждого слова предложения перевода (т.е. постановка слов перевода в нужной форме – числе, роде, падеже и т. д., согласно их лексико-грамматическим признакам).

Очень важное место в эффективном выполнении этого процесса занимает автоматический словарь, который используется почти на всех вышеуказанных этапах перевода. Его задачами является определение способа представления лексических единиц (в виде словоформы или в виде основы слова, в зависимости от типа языка, объема проектируемого словаря и назначения системы перевода), отбор лексики для входного словаря (обычно за основу берется опубликованный словарь соответствующего подъязыка, который пополняется лексикой, полученной путем статистического анализа журнальных статей, монографий и других материалов, изданных в последнее время, или частотно-алфавитный словарь, составленный путем анализа большого объема текстов, если нет опубликованного словаря по рассматриваемой предметной области), отбор соответствующих переводных эквивалентов (которые извлекаются из опубликованного двуязычного словаря по подъязыку с опорой на русский частотный словарь соответствующего подъязыка или путем консультаций со специалистами по данной области) и определение способов кодирования лексико-морфологической, синтаксической, стилистической и семантической информации о словах обоих языков (например, признак части речи, к которой относится лексическая единица, ее морфологические и семантические признаки, сведения об ее стилистическом использовании и использовании ее как части фразеологизма или устойчивого словосочетания и другие). В промышленных системах машинного перевода автоматические словари занимают в памяти компьютера много места, и поэтому они обычно хранятся во внешней памяти.

Важным моментом в ходе процесса автоматического перевода является также выбор определенной синтаксической структуры предложения на выходном языке. Структура этого предложения не только должна быть правильной с точки зрения грамматики языка перевода, но и должна употребляться в выходном языке чаще, чем другие структуры, передающие смысл оригинала. Полученное предложение должно также передавать максимум информации, помещенной в тексте на исходном языке, и отражать его стилистические особенности.

Машинный перевод является одной из самых интересных теоретических и практических задач современной лингвистики. Деся-

тилетия развития научных исследований и практических разработок внесли свой существенный вклад в теорию и в приложения. Нет сомнений в том, что эта проблема будет исследоваться также в будущем. Несмотря на значительное число работ по данной области многие вопросы все еще остаются актуальными.

Использованная литература

1. Гончаренко В. В. и др., *Системы машинного перевода, аннотирования и индексирования и обучения языку*. // Новые компьютерные технологии в обучении языком, Мн., 1997, с. 74-75.
2. Зубова И. И., *Информационные технологии в лингвистике*, Минск 2001.
3. Пиотровский Р. Г., *Машинный перевод (Некоторые итоги и перспективы)*. // Проблемы структурной лингвистики 1971, М., 1972, с. 521-533.
4. Порозинская Г. В., *Семантико-когнитивные аспекты машинного перевода*. // Когнитивные аспекты языкового значения. Иркутск, 1997, с. 39-43.
5. Розенцвейг В. Ю., *Машинный перевод*. // Теоретические проблемы советского языкознания, М., 1968, с. 171-192.
6. Слокум Дж., *Обзор разработок по машинному переводу: история вопроса, современное состояние и перспективы развития*. // Новое в зарубежной лингвистике, М., 1989, Вып. XXIV, с. 357-407.
7. Шаляпина З. М., *Текст как объект автоматического перевода*. // Текст и перевод, М., 1988, с. 113-129.