

УДК 82.035

Т. М. Королева

ОСНОВНЫЕ ПРОБЛЕМЫ МАШИННОГО ПЕРЕВОДА

В статье рассмотрены основные лингвистические проблемы, усложняющие работу систем машинного перевода (МП). Описаны перспективы развития различных видов МП и главные направления совершенствования знаний в теории и практике перевода.

Ключевые слова: машинный перевод, адекватность, полисемантическая, алгоритмы систем перевода.

У статті розглядаються головні лінгвістичні проблеми, що ускладнюють роботу систем машинного перекладу (МП). Описані перспективи розвитку різних типів МП та головні напрями вдосконалення знань щодо теорії та практики перекладу.

Ключові слова: машинний переклад, адекватність, полісемантичність, алгоритми систем перекладу.

Some principle linguistic problems that complicate the automatic translation systems processing are discussed in the article. The perspectives of automatic translation systems development and main trends of translation research both -theory and practice — are read.

Key words: automatic translation, adequacy, polysemantics, algorithms of translation systems.

Задача машинного перевода ставит перед разработчиком множество различных проблем. Не вдаваясь в технические проблемы, решение которых целиком зависит от профессионализма разработчика, остановимся на тех, которые делают задачу МП особенно сложной и нетривиальной. Все они лежат в трех областях:

- проблемы неоднозначности;
- проблемы структурного и лексического различия между языками;

- проблемы грамматических конструкций, состоящих из нескольких слов, таких как идиомы и словосочетания.

Рассмотрим поочередно особенности каждой из вышеуказанных проблем в письменной форме МП.

Проблема неоднозначности высказывания или предложения — одна из наиболее сложных для программистов и филологов, занимающихся вопросами машинного перевода. Очевидным наилучшим вариантом для перевода является такое положение, при котором каждое слово языка имело бы одно значение. Но это далеко не так. Наличие полисемантической (лексической, синтаксической (структурной) и фонетической неоднозначности) значительно усложняет процесс машинного перевода и дешифровки устного речевого сигнала.

Неоднозначность достаточно часто встречается во всех естественных языках. Случай, когда предложение может иметь несколько значений, — это скорее правило, чем исключение. Усложняет работу и то, что неоднозначности могут накапливаться. Например, если предложение состоит из двух слов и каждое может иметь два значения, в худшем случае его можно перевести четырьмя разными способами. Естественно, может понадобиться рассмотреть все возможные случаи перевода, чтобы выбрать только один из них. Таким примером является следующий, где полисеманτικότητα слова “get” при переводе на украинский язык является той ключевой единицей, которая влияет на адекватность передачи смысла в переводном варианте.

When you get it. You get it. Коли ти це отримувси. Ти розумієш.

Как правило, в частных случаях использование некоторых правил [1] сильно упрощает работу МП. Например, сочетаемость и последовательность употребления частей речи в определенном языке,

Структурные и лексические различия языков при переводе часто приводят к проблеме “языковых пробелов”. Это означает, что на одном языке приходится использовать целое предложение, чтобы выразить то, на что в другом языке используется всего одно слово. К этому же случаю можно отнести и слова, которые не имеют полностью адекватного перевода на другие языки: английское “cottage”, русское “дача”, или “загородный дом” и т. п. Приведем еще один пример:

The better stain gette в переводе звучит: *Найкращий уловлювач плям.*

В этом случае переводчику приходится пользоваться неологизмом или приводить необходимое пояснение. Следует отметить, что не

всегда может увенчаться успехом кропотливая работа разработчиков по внесению подобных типов трансформаций в лексикон машинного словаря. Что вполне понятно при быстрых темпах появления неологизмов и заимствований во всех современных языках. В таком случае наиболее эффективным способом коррекции переводного варианта является постредактирование.

К структурным различиям относятся не только различия в грамматическом построении самих языковых единиц, но и различное семантическое толкование одной структуры разными языками. Не следует упускать из виду и те случаи, когда два языка имеют две одинаковые конструкции, но с разными ограничениями для них. Из-за этого перевод нельзя произвести наиболее тривиальным способом. Например:

Everything we do is driven by you. Все, що ми робимо, керовано тобою.

Видно, что разница между английским языком и украинским в том, что в английском дополнительное придаточное предложение вводится беспредложным соединением с главным, а в украинском — предложным. Это делает грамматический перевод почти невозможным для систем, которые не производят абстрактных синтаксических построений (то есть для систем прямого перевода).

Сложные конструкции — идиомы и словосочетания, это еще одна из насущных проблем, требующих однозначного решения при МП. Одним из таких примеров может служить нижеприведенная реклама.

Buy a ticket TO THE CONCERT. Купи билет НА КОНЦЕРТ.

Фонетические особенности речи, зафиксированные в орфографии, также представляют собой определенные загадки для письменной формы МП.

They're g-r-r-r-eat! Вони ч-у-у-у-дові!

Следует отметить, что фактически всеми системами осуществляется перевод только на уровне поверхностного синтаксиса, поскольку еще не разработаны эффективные модели формального представления смысла, носителем которого может выступать язык-посредник — интерлингва, хотя для отдельных узких отраслей такие модели строятся (например, МЕТЕО и LingoWare). Специалисты связывают построение адекватных систем МП с развитием искусственного интеллекта: машина сможет переводить с одного языка на

другой, когда научится думать, как человек [2]. Это — одно из перспективных направлений повышения эффективности МП. На современном этапе создание квантовых компьютерных систем делает это направление одним из наиболее перспективных для использования в МП.

Другой путь совершенствования МП, более доступный на современном этапе, — составить корпус переводных соответствий на интересующих языках. Можно предположить, что такие работы ведутся многими разными командами, но их действия не скоординированы, и потому результат — мал.

Критики современных систем МП полагают, что установка на жанровую ограниченность (научить машину сначала понимать совсем простые, специально отобранные тексты) на практике привела к тому, что задача моделирования естественного языка фактически уступила место задаче моделирования ограниченных (и крайне примитивных) подъязыков отдельных отраслей знания. При этом наилучшего результата на этом пути, как было сказано выше, достигла канадская система TAUM-МЕТЕО, отлично выполняющая задачу англо-французского перевода сводок погоды. Простейшим видом систем такого рода являются автоматические разговорники для туристов, предлагающие пользователю более или менее разнообразные “меню” стандартных вопросов и ответов на двух или нескольких языках [3].

Нельзя не упомянуть и о развитии онлайн-овых технологий перевода, которые имеют большое значение и все больше привлекают внимание пользователей. Этот вид МП незаменим в случаях, когда человек нуждается в срочной информации или не знает языка, на котором ему приходится общаться в настоящий момент.

Следует отметить, что качество перевода, предлагаемого компьютерными системами, в последнее время значительно улучшилось. Это касается главным образом научно-технических текстов. При этом важно сказать и о том, что квалифицированные переводчики редко используют МП и предпочитают полагаться на свой опыт и знания.

В заключение следует сказать несколько слов о перспективах развития машинного перевода. Данный вопрос связан с дальнейшей разработкой и углублением теории и практики перевода, как машинного, так и “человеческого”. Для развития теории важны результаты

сопоставительного языкознания, общей теории перевода, теории закономерных соответствий, способов представления знаний, оптимизации и совершенствования лингвистических алгоритмов. Новые и более эффективные словари с необходимой словарной информацией, строгая терминологизация лексики. Теория и практика работы с подязыками помогут повысить качество перевода лексических единиц. Формальные грамматики, ориентированные на перевод, дадут возможность оптимизировать алгоритмы нахождения переводных соответствий в данной коммуникативной ситуации, которая может быть описана в рамках соответствующих прикладных теорий представления знаний. Наконец, новые возможности программирования и вычислительной техники также будут вносить свой вклад в совершенствование и дальнейшее развитие теории и практики машинного перевода.

ССЫЛКИ И ПРИМЕЧАНИЯ

1. Кво Ч. К. Технологии перевода. — М.: Академия, 2008. — 252 с.
2. White J. S. How to Evaluate Machine Translation // Computers and Translation / H. L. Somers (ed.). — Amsterdam: John Benjamins, 2003. — 211 p.
3. Austermuhl F. Electronic Tools for Translators. — Manchester: St. Jerom Publishing, 2001. — 134 p.

Стаття надійшла до редакції 23.09.2017