

СИСТЕМЫ МАШИННОГО ПЕРЕВОДА

Миргаязова Л.Ф.,

научный руководитель канд. техн. наук Вдовенко В. В.
*Сибирский государственный аэрокосмический университет
им. академика Ф.М. Решетнева*

Прежде всего, следует подчеркнуть разницу между двумя терминами: машинный перевод и автоматизированный перевод. Как известно, "машинный перевод" – технология перевода текстов с одного естественного языка на другой с помощью специальной компьютерной программы. Иными словами текст полностью переводится системой машинного перевода. Тогда как при автоматизированном переводе весь процесс перевода осуществляется человеком-переводчиком, программа только способствует ему в достижении поставленной цели.

Кроме того нет никаких сомнений в том, что машинный и автоматизированный перевод, с одной стороны, и традиционный перевод человеком-профессионалом, с другой, дополняют друг друга. Рассмотрим простое объяснение. В условиях глобального информационного пространства в выполнении перевода множества и множества информации, представленной на самых разных языках, заинтересованы две категории лиц: распространяющие информацию и потребляющие последнюю.

Таким образом, группы людей выделяются по их цели:

- синтез текста;
- анализ текстовой информации.

Синтез текстов, где особо важна когнитивная сторона коммуникации: технической документации, договоров, нот МИДа, научных трудов – традиционно осуществляется с помощью деятельности человека-переводчика. В то время как производить переводы без задействования систем машинного перевода для таких целей, как поиск информации для реферата или аналитической записки, переписка техподдержки компании с заграничным пользователем, оказывается экономически невыгодным и требует излишних временных затрат. Такая информация нужна здесь и сейчас для быстрого анализа и применения. Именно в этом случае незаменимы системы машинного перевода.

В ходе исследования будет рассмотрен ряд различных программных средств, с целью выявления возможностей и особенностей функционирования современных систем автоматического перевода и приложений, основанных на машинном переводе. В частности мы ознакомимся с тремя разными видами программного инструментария переводчика: MCHsystem - приложение, основанное на машинном переводе; лингвистический процессор ЕТАР-3; система автоматизированного перевода OmegaT. А также на наглядных примерах будет рассмотрена проблема языковой неоднозначности.

OmegaT – это инструмент автоматизированного перевода, основанный на технологии TranslationMemory, предназначенный для профессиональных переводчиков. Согласно лицензии пользователь может свободно изменять и распространять это приложение. Как и основная масса систем автоматизированного перевода OmegaT не

осуществляет полный перевод вместо человека-переводчика. Указанная программа обладает следующими особенностями:

- Память перевода (TranslationMemory): OmegaT хранит ваши предыдущие переводы и, затем, предлагает перевод в разумных пределах похожих фрагментов текста, над которым вы работаете в настоящее время.
- Управление терминологией (Terminologymanagement): OmegaT использует упрощенные двуязычные словари для специфической области, то есть глоссарии, содержащие переводы небольших выражений или отдельных слов.
- Работа со сложной файловой структурой: OmegaT предоставляет возможность переводить материалы от одного файла до папки, содержащей подпапки, каждая из которых включает файлы разных форматов.

МСНІ – это многоязычная система чата, спроектированная и разработанная учеными из TokyoUniversityofAgricultureandTechnology. Указанная система основана на машинном переводе и снабжена функцией презентации изображений, встроенных в содержание сообщений коммуникатора для того, чтобы реципиент мог отследить ошибки перевода. Система МСНІ обладает такими преимуществами, как поддержка многоязычного общения, довольно малое время отклика, небольшая когнитивная нагрузка для пользователя. Текущая версия МСНІ поддерживает шесть языков: английский, французский, японский, корейский, вьетнамский и китайский. Вторая особенность данной системы обеспечивается за счет комбинации языка программирования PHP с технологией AJAX. Наконец, использование изображений в качестве подсказок позволяет достичь третьего преимущества МСНІ, так как пользователь может без затруднений понять их.

ЭТАП-3 является многоязыковым лингвистическим процессором, разработанным лабораторией компьютерной лингвистики Института проблем передачи информации РАН. Он основан на теории И. Мельчука "Смысл ↔ Текст". ЭТАП-3 имеет следующие основные возможности:

1. система машинного перевода на основе правил;
2. система синонимического перефразирования предложений;
3. инструментальное средство синтаксической разметки корпуса текстов;
4. модуль перевода на основе универсального сетевого языка (UNL);
5. естественно-языковой интерфейс для баз данных;
6. корректор грамматики.

Текущие составляющие машинного перевода ЕТАР-3 содержат русско-английскую и англо-русскую пары; а также имеется несколько прототипов: русско-французский, русско-немецкий, русско-испанский.

Большим препятствием на пути ученых, занимающихся разработкой систем машинного перевода, является проблема разрешения языковой неоднозначности. Рассмотрим следующий англоязычный пример лексической и синтаксической неоднозначности: *He made a general remark that everything was normal*. Это предложение можно перевести на русский язык двумя способами:

- ‘Он сделал общее замечание, что все нормально’
- и ‘Он заставил генерала заметить, что все нормально’.

Ещё один нетривиальный пример неоднозначности на русском языке выглядит следующим образом: Три типа стали есть на складе. "Три типа металла имеется на складе" или "Три мужчины начали употреблять пищу в складском помещении"? Очевидно, что в таких случаях в работу системы машинного перевода должен

вмешиваться переводчик и делать выбор между возможными вариантами перевода того или иного предложения.

Программные средства, основанные на машинном переводе, созданы для облегчения повседневной обработки текстов на естественном языке, и они с этим справляются на удовлетворительном уровне. Конечно, и в компьютерной лингвистике имеются проблемы, которые ожидают своего разрешения. Но не возможно не согласиться с тем, что перспектива создания совершенной системы машинного перевода выглядит очень и очень заманчиво.