

ПРЕИМУЩЕСТВА И НЕДОСТАТКИ МАШИННОГО ПЕРЕВОДА ТЕХНИЧЕСКИХ ТЕКСТОВ

Борешка Даниил – студент

Научный руководитель – **Вержбовская М.В.**, преподаватель
УО «Гомельский государственный технический университет
имени П.О. Сухого», г. Гомель

В условиях современной жизни необходимо выполнять перевод больших объемов научной и технической литературы в короткие сроки. Человеческий перевод является трудоемким и дорогостоящим, а также требует больших временных затрат. В таких ситуациях переводчику на помощь приходят электронные словари и компьютерные программы перевода.

Интерес к автоматическому переводу текстов существует уже довольно давно. Он возник в середине прошлого века после проведения в США Джорджтаунского эксперимента, направленного на разработку системы компьютерного перевода текстов. В настоящее время в мире реализуются несколько проектов по созданию современных систем машинного перевода.

Машинный перевод – это процесс использования программного обеспечения для осуществления перевода без участия человека. Он заключается в передаче информации с исходного языка на переводящий [1]. Одним из способов описания работы ЭВМ является выполнение перевода на основе предварительно составленного алгоритма. Согласно Ю.В. Рождественскому, машинный перевод объединяет «теорию перевода как особую филологическую теорию и основанную на лингвистике теорию соответствий между двумя и более языками» [2].

Достоинства автоматического перевода очевидны: быстрота перевода, большой объем переводимого текста, консистентность (способность сохранять единообразие перевода в пределах одного документа), доступность; кроме того, он экономически выгоден.

Статистический машинный перевод (SMT) основан на использовании статистической модели, которая анализирует корпус текстов для создания переводов на целевой язык. При этом перевод «по правилам» (RBMT) использует лингвистические правила для создания перевода на целевом языке [3]. Примером SMT является программное обеспечение Google Translate от американской компании Google. Оно использует алгоритмы, основанные на статистике и теории вероятности. Система Google Translate имеет огромный и постоянно растущий корпус, содержащий на данный момент более триллиона слов [4]. В дополнение к этому, Google применяет так называемую переводческую память (translation memory),

что повышает качество текста перевода путем использования ранее переведенных фраз, содержащихся в корпусе.

В переводной модели программы Prompt.One, онлайн-переводчика российской компании PROMPT, используется лингвистическая информация для математического анализа целевого и исходного языков. Эта система перевода может быть дополнена функцией управления терминологией.

В данной статье проведен анализ перевода текстов, сделанный с использованием двух самых популярных и доступных систем машинного перевода: Google Translate и Prompt.One. Сопоставительный анализ переводов, выполненных программами на материале английских текстов технической и научно-популярной литературы, позволяет выявить достоинства и недостатки каждой из этих систем.

Сравним перевод абзаца научно-технического текста.

Исходный текст: «How often does this happen to you? You complete an important document just under the wire and send it off to a colleague as an e-mail attachment, only to hear that she couldn't open the file – she works on a different platform. Or maybe she doesn't have the right application loaded».

Перевод, выполненный посредством Google Translate: «Как часто это происходит с вами? Вы заполняете важный документ под проводкой и отправляете его коллеге в качестве приложения для электронной почты, только чтобы узнать, что она не может открыть файл – она работает на другой платформе. Или, может быть, у нее нет подходящего приложения».

Перевод, выполненный при помощи Prompt.One: «Как часто это происходит с Вами? Вы завершаете важный документ только под проводом и отправляете его коллеге как почтовое присоединение, только чтобы услышать, что она не могла открыть file – она работает над различной платформой. Или возможно ей не загрузили правильное приложение».

Проанализировав оба варианта перевода, приходим к следующим выводам:

1. Ни один из машинных переводчиков не перевел идиому «under the wire» – «в спешке».

2. Prompt.One не перевел слово «file», оставив его на латинице.

3. Prompt.One неверно перевел слово «right» как определение к существительному «application» («правильное приложение» вместо «подходящее приложение»).

4. Prompt.One перевел словосочетание «e-mail attachment» как «почтовое присоединение» вместо «почтовое вложение».

5. Prompt.One неверно выбрал перевод предлога «on» (works on a different platform / работает над различной платформой).

Таким образом, при переводе этого абзаца преимущество продемонстрировала система Google Translate.

Рассмотрим отрывок из научно-популярного текста.

Исходный текст: «Perhaps you've just distributed an eye-catching new brochure or ad for review. But when your clients open the electronic file, text and columns are askew, the logo has mysteriously vanished, and computers have substituted plain vanilla Courier for the elegant font you selected. Then, after you've taken the extra time to distribute paper copies, your clients return a hodgepodge of papers cluttered with scrawled comments, sticky notes, and stapled attachments».

Перевод Google Translate: «Возможно, у вас есть привлекательная новая брошюра или реклама для обзора. Но когда ваши клиенты открывают электронный файл, текст и столбцы искажены, логотип загадочно исчез, а компьютеры заменили обычный ванильный курьер на тот элегантный шрифт, который вы выбрали. Затем, после того, как вы потратили дополнительное время на распространение бумажных копий, ваши клиенты возвращают мешанину бумаг, заполненную нацарапанными комментариями, липкими заметками и скрепленными вложениями».

Перевод Prompt.One: «Возможно, Вы только что распределили привлекательную новую брошюру или объявление для анализа. Но когда Ваши клиенты открывают электронный файл, текст и колонки искоса, эмблема загадочно исчезла, и компьютеры заменили простым Курьером ванили изящный шрифт, который Вы выбрали. Затем после того, как Вы заняли дополнительное время, чтобы распределить бумажные копии, Ваши клиенты возвращают мешанину бумаг, загроможденных небрежно писавшими комментариями, липкими *Примечаниями* и сшитыми приложениями».

Анализируя перевод, можно выявить следующие неточности:

1. Prompt.One перевел фразу «text and columns are askew» как «текст и столбцы искоса» вместо «текст и столбцы смещены».

2. Google Translate наделил фразу «computers have substituted plain vanilla Courier for the elegant font you selected» противоположным смыслом (в отрывке компьютеры заменили выбранный шрифт простым шрифтом Courier, а не наоборот).

3. Prompt.One перевел фразу «you've taken the extra time» как «Вы заняли дополнительное время», что не является наилучшим вариантом. Перевод Google Translate «вы потратили дополнительное время» более точен.

4. Google Translate переводит слово «attachments» как «скрепки», хотя в отрывке речь идет о вложениях. Перевод Prompt.One («сшитыми присоединениями») лучше передает смысл.

По результатам анализа можно заключить, что как Google Translate, так и Prompt.One могут допускать контекстуальные неточности в разных частях текста.

На данный момент, когда речь идет о выборе лучшего онлайн-переводчика для языковой пары «английский-русский», преимущество оста-

ется за системой Google Translate. Она применяет самообучающийся алгоритм машинного перевода. Поскольку система может искажать контекст и ухудшать качество перевода, то редактирование вручную остается необходимым условием при работе с машинными переводчиками.

В перспективе при развитии самообучающихся нейронных сетей и систем искусственного интеллекта возможно появление более совершенных автоматизированных систем перевода, способных решать проблему перевода многозначных слов или выражений [5]. На данный момент только человек способен точно передать смысл иноязычного текста с его стилистическими особенностями и нюансами перевода.

Список цитированных источников

1. Нелюбин, Л. Л. Компьютерная лингвистика и машинный перевод / Л. Л. Нелюбин. – М. : ВЦП, 1991. – 207 с.

2. Рождественский, Ю. В. Философия языка. Культуроведение и дидактика. Современные проблемы науки о языке / Ю. В. Рождественский. – М. : Грантъ, 2003. – 239 с.

3. Исламов, Р. С. Анализ современных систем машинного перевода типа SMT и RBMT / Р. С. Исламов, А. Г. Фомин. // Филологические науки. Вопросы теории и практики : в 2 ч. – 2013. – № 3 (21), ч. I. – С. 69–73.

4. Зубов, А. В. Информационные технологии в лингвистике : учеб. пособие для студ. лингв. фак-тов высш. учеб. заведений / А. В. Зубов, И. И. Зубова. – М. : Изд. центр «Академия», 2004. – 208 с.

5. Исламов, Р. С. Лексическая амбивалентность естественного языка в среде систем машинного перевода : на материале английской официально-деловой документации : автореф. дис. ... канд. филолог. наук : 10.02.19 / Р. С. Исламов. – Кемерово, 2015. – 23 с.