

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ЯЗЫК И КУЛЬТУРА

**Сборник статей
XXIV Международной научной конференции,
посвященной 135-летию
Томского государственного университета
(7–9 октября 2013 г.)**

Ответственный редактор профессор С.К. Гураль

Томск
Издательский Дом Томского государственного университета
2014

Л.В. Панасенко

Томский государственный университет, Томск, Россия

ПРОБЛЕМА ЭКВИВАЛЕНТНОСТИ МАШИННОГО ПЕРЕВОДА: YANDEX, GOOGLE, PROMT

Данная работа посвящена исследованию эквивалентности машинного перевода (МП) на материале онлайн-переводчиков Yandex, Google и Prompt.

Актуальность нашего исследования обусловлена интересом современной лингвистики к решению прикладных задач в сфере МП. Эти исследования были начаты еще в 50-е гг. XX в., получены определенные результаты, но качество и экономичность систем МП, разработанные в тот период, не позволяли использовать их повсеместно.

В настоящее время проблемы, связанные с разработкой систем МП, актуализируются снова в связи с появлением новых телекоммуникационных технологий. С увеличением потока информации и усилением процессов коммуникации между представителями разных культур намечается острая потребность в быстром ситуативном переводе. Для достижения этой цели создаются системы, работающие в онлайн-режиме. Каждый человек может использовать такую систему для быстрого извлечения информации из источников на иностранном языке и реализации коммуникационной потребности. Вместе с тем существующие в настоящее время системы онлайн-перевода по-прежнему не отвечают требованию качества. Вновь актуализируется проблема лингвистических исследований в этой области. Таким образом, можно говорить о том, что машинный перевод нуждается в последующем совершенствовании и доработке.

Одним из этапов совершенствования работы систем МП является диагностика степени эквивалентности перевода на различных уровнях языка. Данная работа, как уже было отмечено выше, посвящена выявлению закономерных системных несоответствий текстов перевода, произведенных различными системами МП, чтобы оценить качество их работы и разработать коррекцию алгоритмов перевода.

Материалом послужили обзорные комментарии футбольных матчей на английском языке и МП тех же самых комментариев на русский язык, выполненный системами Yandex, Google и Prompt.

Рассмотрим вариант перевода фразы *Daily Mirror match report by Simon Bird* (текст репортажа, из которого было взято это выражение – Newcastle 2-1 Sheff Utd, 08:42 06/04/10). Фраза наглядно демонстрирует несоответствия МП:

Google	Yandex	Prompt	Перевод, сделанный переводчиком
Ежедневный отчет о матче Mirror Саймон-птицы	Daily Mirror об отчете матча Саймон Птица	Матч Daily Mirror сообщает Саймоном Бирдом	Протокол матча в ежедневной газете Daily Mirror от Саймона Бёрда

Мы наблюдаем несоответствия на лексическом и грамматико-синтаксическом уровнях.

Рассмотрим первое несоответствие переводов выражения на лексическом уровне.

1. Перевод имени собственного. Google переводит имя собственное *Simon Bird* как *Саймон-птицы*, что неверно, потому что иностранные фамилии принято транслитерировать. Несмотря на это, Google переводит фамилию репортера *Bird* как *птицы*. Отсюда следует, что Google не умеет различать имя собственное и имя нарицательное.

В отличие от Google, Yandex различает имя собственное и имя нарицательное, о чем свидетельствует графический облик слова (оно начинается с заглавной буквы). Несмотря на это, Yandex как и Google, переводит фамилию репортера футбольного матча *Bird* как *Птица*.

В отличие от предыдущих примеров перевода, Prompt транслитерирует данное имя собственное (*Simon Bird* – *Саймоном Бирдом*) и, соответственно, различает имя собственное и имя нарицательное.

Перевод названия газеты *Daily Mirror* также представляет проблему. В оригинале данное имя собственное обозначает ежедневную газету, которая выпускается в Англии. Google переводит *Daily Mirror* как *Ежедневный Mirror*. Слово *Daily* он перевел как *Ежедневный*, а *Mirror* оставил без перевода. Таким образом, Google последовательно не различает имя собственное и имя нарицательное.

Yandex и Promt не переводят название газеты *Daily Mirror* и оставляют его без перевода – *Daily Mirror*. Так как мы уже говорили о том, что имена собственные не переводятся, а транслитерируются, то, следовательно, эти системы совершили ошибку.

Таким образом, можно говорить о том, что они также не способны различать имя собственное и имя нарицательное.

2. Перевод имени нарицательного. Семантическое несоответствие. Google и Yandex переводят слово *report* как *отчет*. Это является правильным вариантом перевода. Promt переводит *report* как *сообщает*. В «Толковом словаре русского языка» Т.Ф. Ефремовой слово *отчет* имеет следующие значения:

отчет – м.

1) а) Письменное или устное сообщение о своих действиях или о выполнении возложенного поручения, представляемое лицу или учреждению.

б) Документ об израсходованных средствах.

в) Публикуемые данные о работе какого-л. учреждения, о положении дел в нем.

2) Объяснение или изложение происшедшего.

сообщать – несов. перех. и неперех.

1) а) Доводить до чьего-л. сведения; уведомлять, извещать.

б) Рассказывать.

2) а) перен. перех. Передавать кому-л. свое состояние, настроение и т.п.

б) Вызывать в другом такое же состояние, настроение и т.п.

3) перен. перех. Придавать кому-, чему-л. какое-л. свойство, качество.

Таким образом, из приведенных значений видно, что *отчет* и *сообщать* схожи по первому значению. Promt перевел *report* как *сообщает* по первому словарному значению – доводить до чьего-л. сведения; уведомлять, извещать. Но данный переводчик допускает грамматическое несоответствие при переводе: в оригинале слово *report* является существительным, а в переводе становится глаголом в наст. вр. – *сообщает*.

Рассмотрим несоответствия на грамматико-синтаксическом уровне.

Обратимся к последнему примеру перевода слова *report*.

Google и Yandex переводят существительное *report* как *отчет*, Promt переводит существительное *report* как глагол в наст. вр. *сообщает*. Таким образом, Promt не сумел правильно подобрать нужную часть речи в данном предложении.

Перевод части речи и субъектно-объектной организации высказывания:

Google		Yandex		Promt	
субъект	объект	субъект	объект	субъект	объект
<i>Daily Mirror</i>	<i>match</i>	<i>Daily Mirror</i>	<i>match</i>	<i>report</i>	<i>match</i>

В качестве субъекта предложения Google предоставляет имя собственное *Daily report* и переводит его как *Ежедневный отчет*, разбивая имя собственное *Daily Mirror* на две части – *Ежедневный* и *Mirror*.

В качестве объекта Google выделяет слово *match* и переводит его как *о матче*.

Представленные субъектно-объектные отношения соответствуют норме литературного языка.

Yandex в качестве субъекта предлагает имя собственное *Daily Mirror*, которое оставляет без перевода. Yandex неправильно выбрал субъект выражения потому, что в данном случае он неправильно идентифицировал логическую структуру исходного текста. В качестве объекта данный переводчик выбирает словосочетание *match report* – *отчете матча*. Такой вариант мог бы иметь право на существование, если бы данная система добавила дополнительный элемент служебной или главной частей речи.

Promt в качестве субъекта высказывания выбирает слово *match* – *матч*, а имя собственное *Daily Mirror* выступает в роли определения. Такое представление субъекта здесь является неверным, так как далее система совершает ошибку при переводе слова *report*: глагол *сообщает*, что делает лексему *матч* субъектом предложения. Последнее является невозможным, потому что *матч* не может сообщать сам себя.

Анализ переводов одного из выражений ясно показывает проблемные зоны, значимые для систем МП:

- непоследовательное различие имени собственного и имени нарицательного;
- неточный подбор лексико-семантического эквивалента;
- неточный подбор грамматического эквивалента;

– неспособность идентифицировать логическую структуру высказывания.

В разных системах конфигурация этих проблем проявляется по-разному.

Таким образом, теоретическая и практическая ценность научной работы заключается в подробном исследовании и выявлении переводческих несоответствий в машинном переводе текстов футбольной лексики с английского языка на русский. В будущем полученные данные могут быть использованы для создания методологического пособия для студентов и преподавателей вузов, чтобы облегчить применение этого вида перевода, а также для специалистов, которые смогут использовать полученные данные при доработке систем МП.

THE PROBLEM OF EQUIVALENCE IN MACHINE TRANSLATION

Panassenko L.V.

Keywords: machine translation; current linguistics; equivalence degree; MT system; handing of applied problems.

Abstract. This article is devoted to the equivalence study of machine translation as exemplified in online translators: Yandex, Google, Promt. Timeliness of this research is determined by concern of current linguistics to handing of applied problems in the matter of machine translation. Nowadays, those problems, connected with MT system study are actualizing once again with the advent of new telecommunication technologies. But, such as online translation systems still miss mark of quality, so one of the steps in development of MT work is diagnostics of the equivalence degree at the different language levels.