

Д. Шульжик

ОСОБЕННОСТИ МАШИННОГО ПЕРЕВОДА ИДИОМ С АРАБСКОГО ЯЗЫКА НА РУССКИЙ

В настоящее время разработчики систем машинного перевода ставят своей задачей создание алгоритма для передачи смысла всего текста, не ограничиваясь прямым переводом отдельных слов. Поэтому проблема правильного перевода идиом является одной из самых актуальных в когнитивных исследованиях и исследованиях в области интеллектуальных систем.

Целью нашей работы стало изучение особенностей перевода идиоматических выражений с глаголами движения и позиции с арабского языка на русский. Материалом исследования выступил «Арабско-русский словарь идиоматических выражений. Справочник переводчика» (А. В. Капшук, А. Джабер, 2018). Всего из словаря было отобрано и проанализировано 110 идиом с глаголами движения и около 90 идиом с глаголами позиции. Перевод каждой идиомы проверялся трижды: с помощью словаря, а также в системах машинного перевода Google Translate и Яндекс.

Проведенное исследование показало, что оба переводчика в основном переводят идиомы дословно. Обе машины умеют подбирать эквиваленты идиом: около 14 % идиом с глаголами движения и 19 % выражений с глаголами позиции. Обе машины выдали полный эквивалент идиомы *لم يترك فيه حجرا على* ‘камня на камне не оставить’. Некорректный или перевод с искажением смысла чаще выдавал Google Translate – 38 % выражений с глаголами движения и 31 % с глаголами позиции были переведены неверно. Переводчик от Яндекса допустил ошибки при переводе 27 % идиом с глаголами движения и 28 % выражений с глаголами позиции.

Установлены следующие проблемы при машинном переводе идиом с глаголами движения и позиции с арабского языка на русский.

1. Отсутствие у машины от Яндекса системы определения арабских огласовок, которая есть у переводчика Google. В идиоме *دَبَّ في احشائها* ‘ползать внутри’ переводчик определил слово *دَبَّ* ‘ползать’ как ‘медведь’.

2. Неправильное определение значения многозначных слов в некоторых выражениях у обоих переводчиков. В идиоме *ترك اثرا* ‘оставлять след’ Google Translate неправильно выбрал значение многозначного слова *اثر* ‘влияние’.

3. Машины не определяют некоторые слова в определенных выражениях и передают их путем транслитерации либо просто выбрасывают их. Например, в выражении *ذهب برشدة* ‘пойти умом’ машина перевела слово *رشدة* ‘ум’ транслитерацией ‘рушада’.

4. Обе машины (чаще Google Translate) выдают неправильные определения слов или устойчивых выражений. В идиоме *رك الامر في عنقه* ‘возложить дело на него’ машина от Google неправильно перевела глагол *رك* ‘возложить’ как ‘надеть’.