

А.А. Баркович, В.В. Колонтаева (Минск, БГУ)

## ПРОБЛЕМА АДЕКВАТНОСТИ МАШИННОГО ПЕРЕВОДА С КИТАЙСКОГО ЯЗЫКА НА РУССКИЙ

В последние годы стало популярно использовать «онлайн переводчики» с целью быстрого получения перевода («онлайн переводчик» — компьютерная программа для перевода текстов с одного естественного языка на другой в Интернете). Можно ли говорить о высоком качестве такого перевода? Конечно, «онлайн переводчики» обладают своими положительными качествами, такими как высокая скорость перевода, универсальность, конфиденциальность и доступность. Однако насколько такие переводчики учитывают особенности языка-оригинала, передают лексическое значение, учитывают омонимию и соблюдают прочие условия, необходимые для точности перевода?

Рассмотрим работу такой программы при переводе с китайского языка на русский. Статьи научно-популярного и публицистического стиля, отобранные на основе наличия в них сложных для перевода специфических языковых структур, дают достаточно объективную картину. Один из наиболее популярных «онлайн переводчиков» — «Переводчик Google» (англ. Google Translate) — веб-сервис компании «Google», предназначенный для перевода части текста или веб-страницы на другой язык. Для некоторых языков пользователям предлагаются варианты переводов, например, технических терминов. Выбор пользователей учитывается при совершенствовании и обновлении программы.

«Переводчик Google» — это бесплатная служба практически мгновенного перевода, которая поддерживает 58 языков. Она предоставляет возможность переводить слова, предложения и целые веб-страницы с одного на любой другой из представленных языков. Основная цель программы — сделать информацию повсеместно доступной и полезной независимо от того, на каком языке она создана<sup>1</sup>. «Переводчик Google» использует собственное программное обеспечение, называемое «статистической технологией перевода» (англ. Statistical-based machine translation). В отличие от традиционных программ, она не использует лингвистические алгоритмы перевода, а основана на приоритете статистически вычисленных наибольших вероятностей совпадений.

Для корректной работы программы необходимы огромные базы параллельных текстов, где попарно хранятся словосочетания (фразы из 2–3 слов) и их переводы, так называемые «N-граммы» (N-граммы — это n-буквенные последовательности, извлеченные из текста. Например, слово «констебль», разложенное в «триграммы» (трехбуквенные последовательности) будет выглядеть так: {«кон», «онс», «нст», «сте», «теб», «ебл», «бль»}. Существуют различные способы извлечения таких

---

<sup>1</sup> [http://translate.google.ru/about/intl/ru\\_ALL/](http://translate.google.ru/about/intl/ru_ALL/)

последовательностей<sup>2</sup>. В соответствии с принципами компьютерной лингвистики, основанное на статистических методах моделирование языка, в том числе с использованием N-грамм, вполне может использоваться для анализа текста и прогнозирования дальнейших последовательностей символов. Программа «выбирает» вариант перевода, основываясь на частоте совпадений, т.е. в конечном итоге будет представлен вариант, имеющий наиболее высокий процент совпадений. Очевидно, что для достоверности результата необходимо иметь в базе не просто большое, а максимально большое количество параллельных N-грамм. И, чем больше базы параллельных текстов, тем выше качество перевода.

Кроме того, обработка значительного объема информации требует наличия достаточно мощного программного обеспечения, что по силам только мощному корпоративному или интернет-серверу. Когда «Переводчик Google» «генерирует» перевод, он ищет соответствия в сотнях миллионов текстов, с целью выявить и предложить пользователю, возможно, наилучший перевод.

Сам процесс поиска соответствий в больших объемах текста называется статистическим машинным переводом. Статистический машинный перевод — это разновидность машинного перевода текста, основанная на сравнении большого количества «языковых пар». «Языковые пары» — тексты, содержащие предложения на одном языке и соответствующие им предложения на втором, уже переведенные человеком, владеющим обоими языками. Таким образом, статистический машинный перевод обладает свойством «самообучения» (чем больше в распоряжении имеется языковых пар и чем точнее они соответствуют друг другу, тем лучше результат статистического машинного перевода). Поскольку перевод выполняется программами, не учитывающими интуитивные способности «естественного» интеллекта, не все результаты идеальны. Но чем больше переведённых текстов есть в распоряжении «Переводчика Google» для анализа тем, наверняка, выше качество перевода.

Степень точности перевода на разных языках различна, что, безусловно, зависит и от «степени» родства взаимодействующих языков.

В настоящее время «машинный перевод» художественных текстов практически всегда оказывается неудовлетворительного качества<sup>3</sup>. Тем не менее, при наличии у пользователя специализированных словарей и некоторой настройке системы на особенности того или иного типа текстов возможно получение перевода приемлемого качества, который нуждается лишь в небольшой редакторской корректировке. Чем более формализован стиль исходного документа, тем более высокого качества перевода можно ожидать. Самых лучших результатов при использовании машинного перевода можно достичь для текстов, написанных в техническом (различные

---

<sup>2</sup> <http://habrahabr.ru/post/75509/#habracut>

<sup>3</sup> Машинный перевод означает анализ текста на основе лингвистических правил и перевод на другой язык с помощью комплексной компьютерной программы.

описания и руководства) и официально-деловом стиле. С нашей точки зрения, это происходит по причине наличия в таких стилях большого количества языковых клише и терминологии, что делает возможным предоставление более верных и точных вариантов перевода.

Простейшей формой машинного перевода является «дословный перевод», при котором с помощью «морфологического анализа»<sup>4</sup> вычисляется начальная форма каждого слова (лемматизация<sup>5</sup>). Далее лемма отыскивается в одном из машинных словарей и в текст вставляется соответствующая единица на языке-адресате. Полученный таким образом текст чаще всего трудно использовать, поскольку он далек от литературной речи.

На следующем после лемматизации этапе анализируется синтаксическая структура исходного текста, а выявленные особенности предложения сохраняются в символической промежуточной форме, из которой впоследствии генерируется предложение на языке-адресате. Результат перевода здесь может быть вполне удовлетворительным для текстов с простой синтаксической структурой и репрезентативным<sup>6</sup> машинным словарем.

Рассмотрим статью научно-популярного стиля «Летучие рыбы» — «飞鱼»<sup>7</sup>. При переводе научной статьи с китайского языка переводчик почти не учитывает грамматические категории (род: *обтекаемый и красивое тело*, число: *длинные грудные плавники был*, падежная несогласованность: *неявные реализации синее море, красивый живописных*), пропускает необходимые в русском языке глаголы-связки, не соблюдает корректный порядок слов в предложениях (*Обратно цвет летучей рыбы и морской воды рядом*).

Перевод в данной ситуации производится пословно, без учета особенностей как китайского, так и русского языков. Например, предложение 飞鱼不是飞翔, 感觉上好像是在拍打翼状鳍, 其实只是滑翔 (перевод «Переводчика Google»: *Летучей рыбы не летают, я чувствую, как хлопает крыльями как плавниками, по сути, просто скользят.*) было бы более адекватно перевести: *Летучая рыба не летает, чувствуется, будто она хлопает плавниками, как крыльями, действительно скользит*<sup>8</sup>. В некоторых случаях смысл оказывается искаженным. Например: 飞鱼的飞动

---

<sup>4</sup> Определение морфологических характеристик слова.

<sup>5</sup> Лемматизация — процесс привода словоформы к лемме, начальной словарной форме. В русском языке начальными формами считаются следующие формы: для существительных — именительный падеж, единственное число; для прилагательных — именительный падеж, единственное число, мужской род; для глаголов, причастий, деепричастий — форма инфинитива.

<sup>6</sup> Одно из основных свойств корпуса текстов наряду с полнотой и экономичностью. Подразумевает под собой наличие в корпусе большого числа текстов разных стилей (от литературы религиозной до технической, от публичной политической речи до неформальной беседы).

<sup>7</sup> Взято с <http://t.cnool.net/view/post:1220356>.

<sup>8</sup> Перевод Колонтаевой В.В.

русский английский **китайский**

飞鱼长相奇特，胸鳍特别发达，像鸟类的翅膀一样。长长的胸鳍一直延伸到尾部，整个身体像织布的“长梭”。它凭借自己流线型的优美体型，在海中以每秒10米的速度高速运动。它能够跃出水面十几米，空中停留的最长时间是40多秒，飞行的最远距离有400多米。飞鱼的背部颜色和海水接近，它经常在海水面活动。蓝色的海面上，飞鱼时隐时现，破浪前进的情景十分壮观，是南海一道亮丽的风景线。

Разрешить написание латиницей



английский **русский** украинский

Летучие рыбы своеобразный внешний вид разработан грудные плавники, как крыло птицы. Длинные грудные плавники был продлен до тыла, всем телом, как ткачество долго трансфер. Это само по себе обтекаемый и красивое тело, высокоскоростного движения в море скоростью 10 метров в секунду. Он может из воды 10 метров, оставаясь в воздухе в течение длительного периода более чем на 40 секунд, максимальное расстояние полета более 400 метров. Обратное цвет летучей рыбы и морской воды рядом, часто в деятельности морской поверхности. Неявные реализации синее море, летучие рыбы, рассекая сцена очень эффектный, красивый живописных маршрутов в Южно-Китайском море.



主要 是逃离捕食者 — *Летучие рыбы вылетел бежать хищников* (перевод «Переводчика Google»). Более правильным, конечно, будет такой перевод: *Движения, похожие на полет, у летучих рыб преимущественно являются способом бегства от хищников*<sup>9</sup>. Как видим, во втором варианте перевода используется совершенно другие конструкции, позволяющие наиболее гармонично соотнести особенности двух языков и правильнее передать смысл предложения.

Обратимся к тексту публицистического стиля «女人这些细节男人 难容忍!»<sup>10</sup>

Здесь перевод также не является приемлемым для прямого использования. Не учитываются грамматические категории, иногда общий смысл коверкается. В первом предложении допущена явная ошибка: «...就能

русский английский **китайский**

也许，一个小动作，就能让你们的恋爱宣告一个段落。所以，不管是在恋爱或者婚姻中，女性朋友没都要注意自己的言行举动，别以为结婚以后就可以百无禁忌，如果你确实让他受不了了，那么他也会选择离婚收场。

细节决定成败。很多人以为，男人都是大大咧咧的，不会在乎生活中的一些细节部分。其实并非所有男人都这样不在乎，男人要么毫不在乎，要么在乎起来比女人还要挑剔。

Разрешить написание латиницей



английский **русский** украинский

Может быть, маленькая хитрость, вы можете позволить вашей любви заявил пункта. Таким образом, любовь или брак, подруга не нужно обращать внимание на свои слова и поступки шаг, не думаю, что после свадьбы, вы можете позволить себе пойти, если вы делаете так, что он терпеть не могу, то он выберет для заканчивается разводом.

Подробнее определить успех или неудачу. Многие люди думают, что люди небрежно, не заботятся о жизни некоторых деталей. На самом деле, не все мужчины все равно, люди либо не заботятся, или ухода, чем женщины, но и придиричивым.



让你 们的恋 爱宣告一个段落» — это, видимо, не «...вы можете позволить вашей любви заявил пункта» как предлагает «Переводчик Google». «...может позволить вашей любви выдержать еще один срок»<sup>11</sup> — будет

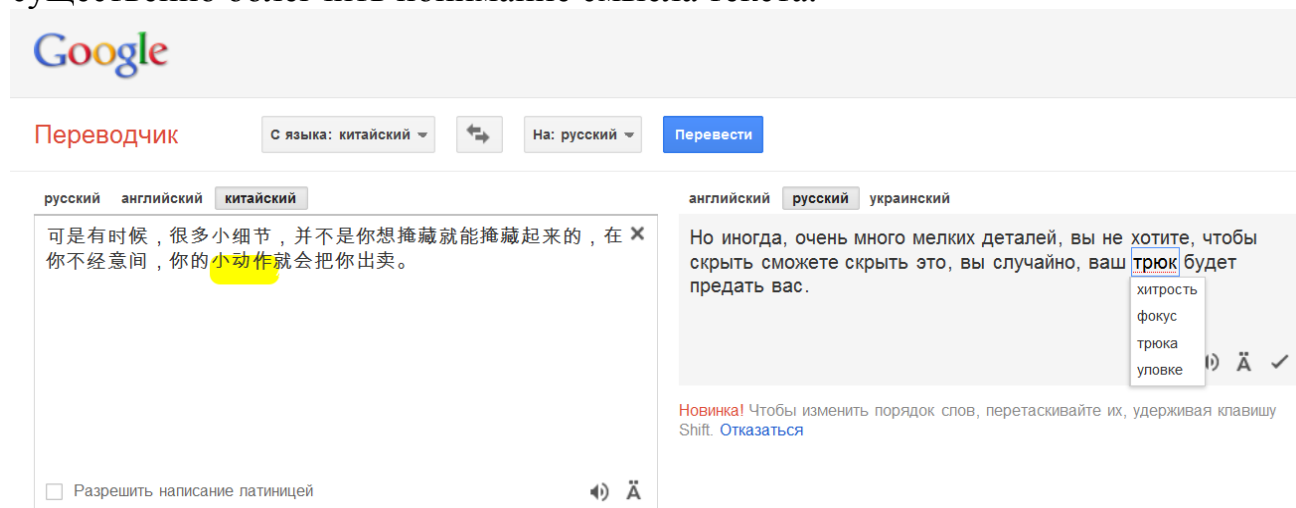
<sup>9</sup> Перевод Колонтаевой В.В.

<sup>10</sup> Взято с <http://lady.qq.com/a/20120226/000029.htm>.

<sup>11</sup> Перевод Колонтаевой В.В.

более корректным. В варианте перевода «на самом деле, не все мужчины все равно, люди либо не заботятся, или ухода, чем женщины, но и придирчивым» (более корректный перевод, возможно: *На самом деле отнюдь не всем мужчинам безразличны недостатки в женщине*<sup>12</sup>). Общий смысл нетрудно разобрать, однако грамматические ошибки все же нарушают целостность и правильность восприятия информации.

Как видно, многозначность и вариативность перевода с китайского на русский язык пока не учитывается «онлайн переводчиком», т.к. грамматические особенности и смысловая связь в предложении зачастую могут утрачиваться при таком переводе. Таким образом, использование «онлайн переводчика» актуально для поверхностного ознакомления с предоставленной информацией. Дополнительно в «Переводчике Google» есть дополнительная функция, позволяющая пользователю самостоятельно подбирать вариант перевода (с учётом значения и формы слова), что может существенно облегчить понимание смысла текста.



«Если вы обнаружите перевод, который, по-вашему, неверен, в большинстве случаев Переводчик Google предложит другие варианты. Чтобы просмотреть их, нажмите на переведенный текст. Если указать альтернативный перевод, который вы считаете более подходящим, Переводчик Google запомнит ваш выбор и вы улучшите его работу» — сообщает «Переводчик Google».

Такой «переводчик», безусловно, — хорошее подспорье для пользователей, не имеющих специальной подготовки. Для детального, тщательного и более верного перевода обязательно редактирование текста человеком, с использованием специальных словарей. Человек осмысливает любую информацию как часть более широкого контекста: соседние предложения определяют и объясняют скрытые или многозначные реалии. «Онлайн переводчики», даже учитывающие синтаксические особенности, часто выдают гладкие, но совершенно невразумительные переводы.

<sup>12</sup> Перевод Колонтаевой В.В.

*«Хотя мы работаем над этой проблемой, не исключено, что система, обеспечивающая быстрый и качественный перевод, появится еще не так скоро. Однако мы надеемся, что на данном этапе предоставляемые нами услуги в большинстве случаев окажутся полезными».* Появление идеальной программы для перевода будет не скоро. На данном этапе развития информационных технологий возможно ограниченное использование таких программ. Часто машинный перевод подвергается критике (иногда справедливой), но программы созданы и работают, пусть и с ошибками. Реалии компьютерной лингвистики сегодня — в преодолении трудностей, в том числе, отладки программ-менеджеров, обнаружения и исправления ошибок.

#### ЛИТЕРАТУРА

1. Машинный перевод [Электронный ресурс]. — [http://ru.wikipedia.org/wiki/машинный\\_перевод](http://ru.wikipedia.org/wiki/машинный_перевод). — Дата доступа: 20.02.2012.
2. Переводчик Google [Электронный ресурс]. — [http://ru.wikipedia.org/wiki/google\\_translate](http://ru.wikipedia.org/wiki/google_translate). — Дата доступа: 20.02.2012.
3. История машинного перевода [Электронный ресурс]. — [http://www.prima-vista.ru/rus/articles/mashinniy\\_perevod](http://www.prima-vista.ru/rus/articles/mashinniy_perevod). — Дата доступа: 20.02.2012.
4. Переводчик Google: Часто задаваемые вопросы [Электронный ресурс]. — [http://www.google.ru/intl/ru/help/faq\\_translation.html#quality](http://www.google.ru/intl/ru/help/faq_translation.html#quality). — Дата доступа: 20.02.2012.
5. Переводчик Google: О переводчике [Электронный ресурс]. — [http://translate.google.ru/about/intl/ru\\_all](http://translate.google.ru/about/intl/ru_all). — Дата доступа: 20.02.2012..
6. Анализ машинного перевода [Электронный ресурс]. — [http://comp.potrebitel.ru/?action=model\\_list&num\\_id=71&cat\\_id=669](http://comp.potrebitel.ru/?action=model_list&num_id=71&cat_id=669). — Дата доступа: 20.02.2012.
7. Машинный перевод: правила против статистики [Электронный ресурс]. — <http://filolingvia.com/publ/25-1-0-898>. Время доступа: 20.02.2012.
8. Новиков В.А. Трансфер в современных системах машинного перевода [Электронный ресурс]. — Режим доступа: <http://www.dissercat.com/content/transfer-v-sovremennykh-sistemakh-mashinnogo-perevoda>. — Дата доступа: 20.02.2012.