

18.07.2012, 16:49:55

---

## С родного на чужой и обратно: тестируем 7 онлайн-переводчиков

*Выбор онлайн-переводчика – дело ответственное и неочевидное. Известно, что различные сервисы предоставляют неодинаковое качество перевода, более того – это качество у одного и того же сервиса различается в зависимости от выбранной языковой пары. Как не ошибиться в выборе и получить наилучший результат для нужного языка? Ответить на этот вопрос поможет сравнительное тестирование, проведенное профессиональным переводчиком.*

Онлайн-переводчики для сравнительного тестирования выбирались из списка наиболее известных сервисов на основе личных предпочтений автора. Приоритет был отдан тем, которые работают с русским языком и известны в России. Также были найдены онлайн-переводчики, для которых тот или иной язык (русский, немецкий, украинский, английский) является "родным", чтобы выяснить, насколько их уровень перевода на этот язык и обратно отличается от общепринятых мировых и российских лидеров.

Было рассмотрено 6 известных автору языков (русский, английский, немецкий, французский, украинский, польский), из них 4 синтетических, 2 аналитических, 3 славянских, 2 германских, 1 романский. Особое внимание уделялось сопоставлению уровня перевода между относительно близкими по строю языками (например, между славянскими или между западноевропейскими) с уровнем перевода между далекими по родству языками.

Обычно сервисы хорошо переводят с "родного" языка и обратно, при этом лучше всего работают с языковыми парами, в которых второй язык является близким "родному". Однако всегда есть место исключениям, как показывают оценки.

Оригиналы были взяты на всех шести языках из выборки по темам: новости науки, спорта и культуры, популярная психология, популярная лингвистика, здоровье, информационные технологии, авторские права в сети. Тексты были подобраны с минимумом сленга, узкоспециальной терминологии и запутанных грамматических конструкций – образцы литературного языка, доступного для понимания среднестатистическим пользователем, то есть средней сложности для переводчика. Все материалы были переведены в испытываемых сервисах на пять других языков из выборки, что составляет 30 направлений перевода.

### Допустимая погрешность

В нашем исследовании достоверность результата зависит от добросовестности одного специалиста – автора статьи, причем польским и французским она владеет весьма несвободно. Также обзор мог бы стать более репрезентативным и выверенным за счет включения в выборку большего количества языков, представляющих разные языковые группы, и скрупулезного отбора текстов для тестирования.

## Критерии выставления оценок (в баллах)

- 6 – "Работа специалиста" - ориентир, он же недостижимый идеал, на настоящий момент для всех технологий машинного перевода.
- 5 – "Готовый продукт" – перевод только с парой грамматических помарок и, возможно, не всегда самым-самым точным выбором лексики (встречается в некоторых сервисах при переводе между родственными языками).
- 4 – "Полуфабрикат" – ровный текст лишь с небольшими искажениями и редкими "темными местами", ясным общим смыслом, в целом хорошей лексикой и грамматикой, приятными находками в сложных случаях. Редко требуется обращение к оригиналу, но доработка специалистом необходима.
- 4,5 – "Полуфабрикат" плюс очень хорошая лексика и/или грамматика.
- 3,5 – "Полуфабрикат", но в тексте имеется повышенное содержание "темных мест", неправильных переводов слов, форм слов и связей между словами. N.B.  $(4,5 + 3,5)/2 = 4$
- 3 – "Штамповка" – пословный перевод, более-менее понятный, но нередко с неточной лексикой, искаженным смыслом и грамматическими структурами, а также малым количеством адекватной фразеологии. Без оригинала смысл местами неясен.
- 2 – "Суррогат" – пословный перевод, часто с языка-посредника, смысл и грамматические связи еле угадываются, искажены.
- 1 – "Яд" – "Суррогат" плюс много непереведенных слов.

Оригиналы на разных языках получились неодинаковой сложности, точнее, в каждом тексте были свои сложности. Тем не менее, такие неоднородные материалы позволили оценить точность работы сервисов в целом, их сильные и слабые стороны. Вне зависимости от особенностей оригинала онлайн-переводчик вполне может показать свой уровень перевода, "технику" – будь то "механическая" подстановка словарных значений из скудного словарика или грамотная дешифровка и передача структуры предложения в сочетании с хорошими словарными базами и алгоритмами поиска, или даже (высший пилотаж) анализ контекста и в случае синонимии подбор оптимальных слов, фраз, грамматических форм.

Уровень работы с текстом вполне поддается оценке независимо от числа и многообразия сложных мест в исходных текстах. Кроме того, недостаточно развитое в сервисе направление перевода или просто плохой переводчик "напортачат" и в элементарном, тогда как качественные машинные переводчики справляются с большинством сложностей и выдают ровный, внятный текст (хоть он и уступает человеческому и полон неточностей).

По этим критериям (понятность и не "дословность") качество могут оценить и сами пользователи, вероятно, именно так и выбирают онлайн-помощников. Преимуществом

данного обзора является лишь большее (по сравнению с реальными рабочими условиями) количество языков при единой базе "проверочных материалов".

## Параметры оценки

Оцениваются адекватность выбора лексики, фразеологизмов, терминов, грамматических структур, связность и понятность перевода в сравнении с переводами того же текста на другие языки у того же сервиса и соответствующими показателями других приложений.

## Список участников сравнения онлайн-переводчиков

Google Translate  
 PROMT  
 SDL FreeTranslation.com  
 Reverso  
 WordLingo  
 Bing Translator  
 Pragma

Далее онлайн-переводчики ранжированы не на основе результатов из таблицы, а по субъективно оцененному признаку "филигранности работы" с особенностями русского, английского и немецкого языков. Первые два наиболее актуальны для российского пользователя, а немецкий показатель при тестировании тем, что, как и русский, он весьма сложен в переводе, в том числе машинном, поэтому хорошее "владение" им придает приложению, так сказать, респектабельности.

При характеристике переводчиков также принималось во внимание (малое) число "халтурных" направлений перевода и дополнительная польза специалисту. К примеру, Google Translate "подсказывает" имена, термины, живые обороты речи, а онлайн-переводчики PROMT и SDL – грамматику.

## Оценка онлайн-переводчиков в различных направлениях перевода\* \*\*

Онлайн-переводчик	Google Translate	PROMT	Bing Translator	Pragma	SDL FreeTranslations.com	WordLingo	Reverso
Направление перевода	Google, США	PROMT, Россия	Microsoft, США	Trident, Украина	SDL, Англия	WordLingo США-Великобритания (?)	Reverso-Softissimo, Германия (?)
РУ – АН	4,5	4	3,5	4	4,5	4,2	4
АН – РУ	4,2	4	4	3,8	4,2	2,5	4
РУ – УКР	5	-	3,5	4,8	-	-	-
УКР – РУ	5	-	3	5	-	-	-
РУ – ПОЛ	4	-	3,5	4,2	-	3,8	-
ПОЛ – РУ	4	-	3,5	3,5	-	2	-
РУ – НЕМ	4	3,5	3,5	3	3,5	2,5	3,5
НЕМ – РУ	3	4,5	2,5	1	4,5	1	3,5
РУ – ФР	4	4,2	3,5	2	4,5	3,5	3,5
ФР – РУ	3,5	4	3	3	4,5	1	3,5

\*Диапазон баллов: от 1 б. (бред) до 6 б. (работа специалиста).  
 Цветовые обозначения:

Сиреневый – очень хорошо

Серый – посредственно

Болотный – очень плохо

\*\*Всего 30 направлений. Для просмотра полной версии нажмите на таблицу

Источник: CNews Analytics, 2012

Сервисы из конца списка тоже предлагают хорошие переводы, но только в некоторых направлениях.

Google Translate: равный среди первых

Наиболее известный сервис – Google Translate – получает стабильно высокие баллы по многим рассмотренным направлениям перевода, а в парах с английским и в русско-украинском направлении максимально высокие.

Google Translate делает упор на лексику и фразеологию, статистически наиболее вероятное и точное соответствие фрагментов текста. Нередок замечательный перевод имен собственных, терминов, выбор подходящей лексики и живых оборотов речи, что выгодно выделяет продукцию Google среди конкурентов. При этом тексты перевода достаточно внятные, количество ляпов, непереуведенных слов не больше, чем у других качественных машинных переводчиков.

Грамматическим связям уделяется меньшее внимание, и они передаются не всегда на должном уровне, грамматика не столько дешифруется сама по себе, сколько "автоматически" наследуется при нахождении соответствия. Это похоже на обучение языку погружением в среду: человек запоминает готовые фразы, а не правила построения. Действует!

Главный минус сервиса в том, что в большинстве направлений текст переводится сначала на английский, а потом на язык перевода. Результат, таким образом, несколько искажается, добавляется "шум", к тому же преломляется и по минимуму используется информация, заложенная в грамматических формах. В частности, перевод на аналитический английский "обнуляет" многие грамматические значения (например, в окончаниях).

Оказалось, что этот принцип языка-посредника использует не только Google Translate, но и другие "мультиязычные" сервисы. Однако даже с таким эффектом "сломанного телефона" Google Translate мало уступает другим, более "дотошным" программам, то есть эти "непрямые" переводы проигрывают лишь переводам того же Google с английского и на английский, но не результатам других сервисов.

Говоря о вреде языка-посредника в многоязычных машинных переводчиках, хочется отметить высокую точность переводов Google с русского на украинский и наоборот. В качестве исключения Google переводит в этом направлении без посредничества английского, что позволяет избежать "шума" в переводе, ненужного искажения, какое мы можем наблюдать, если в любом машинном переводчике переведем какой-нибудь текст "туда и обратно". Эти языки очень похожи, перевод между ними – "легкая добыча" для машинного перевода; разработчикам лучше вовсе не предлагать это направление, чем "перегонять" тексты через английский, то есть чем использовать английский в качестве посредника между близкими по строю языками.

Есть предположение, что русско-украинские переводы Google, организованные столь исключительным для сервиса способом, блещут за счет того, что компания выкупила некую "базу лексики и грамматики" для этой языковой пары у какого-нибудь небольшого местного игрока, специализирующегося в данном направлении, и

отшлифовала. Жаль, что другие славянские языки не подключены по той же схеме: перевод на польский с русского и украинского и обратно производится через промежуточный английский.

Юзабилити: 4,8. Интерфейс самый современный, много дополнительных опций в виде озвучки, моментального перевода, синхронного с вводом текста, и проч. – но копировать переведенный текст неудобно.

PROMT: перевод через "метаязык"

Продукция известного российского машинного переводчика PROMT выделяется отличной дешифровкой и передачей грамматических связей, что позволяет закрыть глаза на местами слишком дословный перевод и некоторую "старомодность" или неточность лексики и фразеологии. Сервис показывает очень ровное качество, "выверенность" переводов во всех немногочисленных, но хорошо проработанных направлениях. Заметно наличие базы имен собственных, то есть они часто хорошо опознаются и передаются, особенно для языковых пар, включающих английский или русский.

Если Google Translate для почти всех поддерживаемых языков переводит сначала на английский (стержневой язык) и постоянно обеспечивает пополнение "статистических эвристик", то есть соответствий слов и фраз, для этих направлений перевода, то PROMT скорее дешифрует текст на некий "метаязык", а потом по этому шаблону находит соответствия в описаниях грамматических правил и словарях языков перевода – часто дословно и излишне громоздко, но "близко к тексту", не отклоняясь от оригинала. В тестовых переводах исходный текст дешифровывался на все рассмотренные языки сходными конструкциями, оптимальными для этих языков, но порой слишком "неживыми".

Программа использует скорее книжный стиль, чем разговорный.

Для русскоязычного пользователя PROMT особенно ценен качественным переводом с русского и на русский, так как для многих хороших западных переводческих сервисов наш язык не является приоритетным и качество работы с ним от этого страдает. Также PROMT отличается хорошим уровнем немецких переводов. Этот язык сложен для машинных переводчиков: типичны ошибки с формами слов (как и на русском), особенно глаголов, с порядком слов в предложении, позицией глагола. В связи с этим сервисы, качественно работающие с типично сложными для машинных переводчиков языками (синтетическими – как немецкий, русский и другие славянские), вызывают особое уважение. Технология PROMT представляется одной из оптимальных именно для синтетических языков. Поэтому удивляет отсутствие поддержки других славянских языков. Это можно объяснить экономическими причинами, как в случае перевода между тюркскими языками: выполняется он относительно просто, но заказчиков на него будет немного.

Юзабилити: 5. Отличный современный интерфейс, все, что нужно для быстрого перевода и копирования результатов. Есть возможность заглянуть в словарь.

SDL FreeTranslation.com: 5 с плюсом, не считая английского

Британская компания SDL известна в России программой для автоматизации и координации переводов Trados. Ее онлайн-переводчик показал превосходные результаты: они отличаются хорошей лексикой и грамматикой, некоторые предложения вовсе не требуют коррекции. Также радует добросовестность подхода: сервис не предлагает суррогатных переводов (через язык-посредник) с целью создания

видимости большого количества поддерживаемых направлений перевода. Если для английского выбор направлений перевода большой, то для остальных языков, рассмотренных в данном обзоре, их 4-5 (как и у PROMT), но во всех перевод на хорошем уровне. Удивляет только худшее качество английского перевода по сравнению с конкурентами и переводом на другие языки в этом же сервисе, ведь обычно английский поддерживается лучше остальных языков или хотя бы "на уровне" (особенно у англоязычных компаний). Юзабилити: 5+. "Все удобства" плюс озвучка перевода.

Reverso: Юзабилити? Нет, не слышали!

Западноевропейский онлайн-переводчик Reverso предоставляет перевод на высоком уровне для западноевропейских языков, для русского – несколько хуже, но достаточно качественный, особенно в паре с английским. Подобные показатели объясняются тем, что с русского на немецкий и французский и наоборот перевод осуществляется через промежуточный перевод на английский, а с немецкого на французский и наоборот – переводы прямые, без посредника. Как и в случае с Google Translate, Bing и другими, не прямые переводы оказываются хуже, чем прямые (в таких многоязычных сервисах без посредника обходятся обычно все направления в паре с английским или между западноевропейскими языками).

Удивляет только тот факт, что не все переводы на немецкий и обратно – лучшие в выборке. Во француско-немецком переводе были сложности с порядком слов: глаголы были вынесены в совсем дальние части предложения, что значительно затемняло смысл, так что даже перевод "слово в слово" был бы здесь удачней, чем подобная "аналитика" структуры предложения.

Юзабилити: 3,5. Прошлый век или даже тысячелетие. Это касается и дизайна. При переводе загружается новое окно, медленно, для выбора нового перевода нужно снова загружать стартовое окно. Иконки с наиболее популярными направлениями скорее мельтешат в глазах, чем помогают. Европейский плюс: есть виртуальная клавиатура для букв с диакритиками.

WordLingo: сила – в баннерах

Сервис WordLingo специализируется скорее на западноевропейских языках, чем славянских. Качество работы с последними хромает, даже в паре с английским. Перевод с русского, польского на немецкий и французский и обратно осуществляется через английский, что обуславливает еще более низкое качество работы с этими направлениями, чем со славяно-английскими. Переводы между английским, немецким и французским неплохие, особенно с точки зрения грамматики. Однако словарь для всех языков слабоват, перевод часто выполняется пословно, лексика неточная, много непереуведенных слов, особенно в славянских направлениях перевода. Немецкие и французские переводы в целом удачны, особенно в типично сложных для машинных переводчиков случаях с немецкими окончаниями, а также артиклями и предлогами. Тем не менее по сравнению с более сильными сервисами у WordLingo структура немецкого предложения местами хуже, порядок слов нарушен, глагол бывает отнесен слишком далеко в конец предложения (как у Reverso). В итоге получается "расшатанное предложение", в котором гораздо меньше улавливается смысл, чем при копировании порядка слов из оригинала у менее старательных переводчиков.

Юзабилити: 3,5. Аналогично Reverso. Дизайн устаревший. При переводе каждый раз обновляется окно с большим количеством баннеров. Есть виртуальная клавиатура для букв с диакритиками.

## Bing Translator: "лишнее" вычеркиваем

Переводчик Bing Translator Microsoft, как и Google Translate, работает в связке с технологиями поисковика (Bing) и имеет стержневой язык (английский), к которому подключаются остальные поддерживаемые языки (их также масса) и посредством которого ведется перевод между этими языками. Это, как упоминалось выше, обычно приводит к некоторому искажению смысла во всех направлениях перевода, кроме прямых, то есть кроме направлений с английским. Этот переводчик скорее "англоцентричный", как и Google Translate, WordLingo и многие другие, не вошедшие в обзор. Он мало настроен на русскоязычную аудиторию, качество переводов на русский и другие славянские языки низкое. Проработка немецкого языка (тоже синтетического и традиционно сложного в машинном переводе из-за трудностей согласования, системы падежей и пр.) также хромает. Перевод между западноевропейскими языками выполняется качественно.

Особенность: как синхронист, Bing Translator упрощает "запутанные" фразы, выбрасывает "лишние" слова, старается строить гладкие фразы: зачем в одном предложении слишком много глагольных форм?

Юзабилити: 5. Не хватает только виртуальной клавиатуры.

## Pragma: переводчик для славян

Результаты онлайн-переводчика Pragma украинской компании Trident ("трезубец") расцветены в таблице серым и болотным, кроме вполне хороших переводов между славянскими языками. По-видимому, их и стоит считать специализацией приложения, несмотря на то, что оно поддерживает 56+ языков, включая экзотические. Pragma значительно хуже работает с западноевропейскими языками, чем с родными славянскими. Вопреки заявлениям разработчиков об оригинальной "языконезависимой" архитектуре этого машинного переводчика создается впечатление, что Pragma также переводит через язык-посредник русский или украинский. Во-первых, качество переводов во всех направлениях, где эти языки не участвуют, хуже. Во-вторых, в рассмотренных текстах перевода содержится много кальки с русского и украинского языков, особенно в польском и немецком переводе. В некоторых переводах на польский и немецкий сохраняются непереуведенные украинские слова и те же формы слова, что в украинском (но не русском и не языке оригинала).

Главные минусы – ограниченный словарь, малая база собственных имен и алгоритмов транслитерации. Слабо проработана грамматика западноевропейских языков: очень много славянизмов и мало правильно выбранных форм слова, фраз, грамотно построенных словосочетаний, не "от балды" проставленных артиклей и окончаний. По сути это пословный перевод по русскому/ украинскому грамматическому шаблону. Кроме того, в западноевропейских направлениях перевода лексика выбирается часто неуместно, много слов вовсе не переведено и не транслитерировано.

Юзабилити: 5 б. Все, что нужно для быстрого перевода и выбора/ смены языка. Дизайн несовременный, но "по делу". Есть рекомендации по подготовке теста к машинному переводу.

## Классификация технологий машинного перевода

Попытаемся сопоставить основные черты перевода у рассмотренных в обзоре сервисов, перечислить плюсы и минусы соответствующих технических решений. Классификация основывается на внешних показателях осязаемых, реальных

результатах перевода, а не "внутренней" информации о рабочих процессах в соответствующих компаниях и не на их публичных заявлениях для пользователей.

## Классификация технологий машинного перевода



Источник: CNews Analytics, 2012

Отмеченные в схеме тенденции не проявляются на 100 % и не являются монополией указанных приложений. Во-первых, идеального исполнения ни у кого из них нет. Во-вторых, разработчики вполне могут сочетать "полярные" подходы. Обычно это происходит не в равной степени, делается упор на один из методов как основополагающий, "передовой".

Приведем примеры таких дополняющих комбинаций. В статистическом Google Translate на ранней стадии разработки использовали ядро Systran, традиционного машинного переводчика, который анализирует грамматические конструкции одного языка и преобразует их в конструкции целевого языка. Возможно, и на нынешнем этапе Google без публичных деклараций применяет грамматический подход или упорядочение объемных статистических находок людьми, как это делали и Google, и Яндекс для обучения своих поисковиков.

С другой стороны, PROMT не является "чистым" представителем традиционного машинного перевода. Это гибридная система, сочетающая классический грамматический подход со статистическим. Однако в сравнении с результатами Google Translate у российского сервиса на первый план выступает именно лучшая проработка грамматики, с чего и "начинался" этот машинный переводчик, а вот статистические находки фраз, имен, терминов пока недостаточны для полноценной конкуренции с "самым статистическим" онлайн-переводчиком Google. Возможно, дело за временем или же за приобретением баз параллельных текстов какого-нибудь ведущего переводческого агентства. Так, Google на начальном этапе тренировали свой сервис на документации ООН, а IBM в 2010 году выкупила у Lionbridge массив оригиналов и переводов на множество языков в самых разных тематиках.

Это распространенный ход – интегрировать в свою работу чужой успешный опыт, улучшить свой метод чужими наработками. Однако упор лишь на статистический подход не даст на выходе, даже со временем, полноценные читабельные результаты



перевода, сопоставимые с трудом специалиста, как автор убедительно показала в своей предыдущей статье. Так что ждем с нетерпением дальнейших новинок машинного перевода, которые бы "разорвали шаблон", представленный в нашей классификации.

Елена Тихомирова

Адрес новости: <http://www.cnews.ru/reviews/index.shtml?2012/07/18/496805>

---

Все права защищены © 1995 – 2012 РИА «РосБизнесКонсалтинг»  
По вопросам рекламы и подписки на новости обращайтесь на [marketing@cnews.ru](mailto:marketing@cnews.ru)

ПЕЧАТАТЬ

