

Исламов Роман Сергеевич, Фомин Андрей Геннадьевич

**АНАЛИЗ СОВРЕМЕННЫХ СИСТЕМ МАШИННОГО ПЕРЕВОДА ТИПА SMT И RBMT**

Статья раскрывает суть работы современных систем машинного перевода двух типов SMT и RBMT. Основное внимание уделяется анализу и сравнению этих систем, а также выявлению достоинств и недостатков каждой. Кроме этого, в статье отражены некоторые практические рекомендации по улучшению качества машинного перевода.

Адрес статьи: [www.gramota.net/materials/2/2013/3-1/18.html](http://www.gramota.net/materials/2/2013/3-1/18.html)

Источник

**Филологические науки. Вопросы теории и практики**

Тамбов: Грамота, 2013. № 3 (21): в 2-х ч. Ч. I. С. 69-73. ISSN 1997-2911.

Адрес журнала: [www.gramota.net/editions/2.html](http://www.gramota.net/editions/2.html)

Содержание данного номера журнала: [www.gramota.net/materials/2/2013/3-1/](http://www.gramota.net/materials/2/2013/3-1/)

**© Издательство "Грамота"**

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: [www.gramota.net](http://www.gramota.net)

Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: [voprosy\\_phil@gramota.net](mailto:voprosy_phil@gramota.net)

## STATUS, STRUCTURAL AND STRUCTURAL-SEMANTIC TYPOLOGY OF ENGLISH TERMS AND SLANGISMS OF SHOW BUSINESS SPHERE

**Ioakimidi Georgii Aleksandrovich**  
Pyatigorsk State Linguistic University  
vatoslocos7@gmail.com

The author reveals the status and structural-semantic typology of terms and slangisms of show business sphere in the modern English language, considers the status characteristics of terms and slangisms of corresponding sublanguage, classifies them according to structural and semantic features, revealing their percentage ratio, and basing on the analysis of a large number of examples concludes that word-formative pattern based on rethinking is priority for the terms and slangisms of show business sphere.

*Key words and phrases:* status; structure; semantics; typology; term; sublanguage; slang; show business; semantic feature; slangism.

УДК 81'322.4

### Филологические науки

*Статья раскрывает суть работы современных систем машинного перевода двух типов SMT и RBMT. Основное внимание уделяется анализу и сравнению этих систем, а также выявлению достоинств и недостатков каждой. Кроме этого, в статье отражены некоторые практические рекомендации по улучшению качества машинного перевода.*

*Ключевые слова и фразы:* перевод; машинный перевод; статистический перевод; перевод «по правилам»; тождество при переводе; SMT; RBMT.

### Исламов Роман Сергеевич

**Фомин Андрей Геннадьевич**, док. филол. наук  
Кемеровский государственный университет  
r.islamov87@gmail.com; andfomin67@mail.ru

## АНАЛИЗ СОВРЕМЕННЫХ СИСТЕМ МАШИННОГО ПЕРЕВОДА ТИПА SMT И RBMT<sup>®</sup>

*Исследование выполнено при финансовой поддержке Министерства образования и науки Российской Федерации в рамках ФЦП "Научные и научно-педагогические кадры инновационной России" на 2009-2013 годы (тема: "Профессиональный иностранный язык в формировании универсальной компетентностной модели выпускника технического вуза"; Государственное соглашение: 14.В37.21.2129)*

Машинным переводом называют процесс использования программного обеспечения для осуществления переводческой деятельности, то есть переноса информации с одного языка, который является исходным языком (ИЯ), на другой, являющийся переводящим (ПЯ), без вмешательства человека. Существуют два основных типа машинного перевода: статистический машинный перевод (*Statistical Machine Translation* или SMT) и перевод «по правилам» (*Rule-Based Machine Translation* или RBMT).

В данной статье мы рассмотрим оба типа этих систем на примере двух программ: американской *Google Translate* и российской ПРОМТ. Методом сопоставительного анализа переводов, выполненных программами, на материале английских текстов деловой документации мы, анализируя собственные примеры, выделим достоинства и недостатки каждой из них, а также представим некоторые приемы, способствующие улучшению качества автоматического перевода.

Итак, статистический машинный перевод (SMT) использует в своей основе статистическую модель, основанную на анализе корпуса, для генерации текста на ПЯ. Перевод «по правилам» (RBMT) же использует лингвистические правила для анализа переводимого материала на ИЯ для того, чтобы сгенерировать текст на ПЯ. Вышесказанное можно проиллюстрировать следующей схемой:



Одним из примеров *SMT* является программное обеспечение компании *Google – Google Translate*. Для осуществления перевода *Google* применяет различные алгоритмы, основанные на теории вероятности и статистике, обладая при этом обширным и постоянно растущим корпусом, насчитывающим на настоящий момент свыше триллиона слов. Кроме этого, *Google* осуществляет привязку своей «переводческой памяти» к выполняемому машинному переводу, что позволяет повышать качество генерируемого текста благодаря готовым переведенным ранее переводчиком-человеком фразам, содержащимся в корпусе. Таким образом, комбинация чистого машинного перевода с «переводческой памятью» увеличивает соответствие и качество перевода.

В переводе «по правилам», как в случае с программой ПРОМТ, используется лингвистическая информация для математического анализа ИЯ и ПЯ, что делает его более предсказуемым, иногда слишком буквально, но грамматически превосходящим *SMT*. Этот тип машинного перевода может быть оснащен системой управления терминологией, чтобы генерируемый текст удовлетворял специфике переводимого первоисточника в терминологическом смысле.

Попытаемся проанализировать два типа систем на конкретных примерах. *SMT* требуется анализ параллельных корпусов для генерации перевода. *RBMT* необходимы «знания», которые вносятся в нее специалистом: грамматика, синтаксис, семантика, лексика для ИЯ и ПЯ соответственно. То есть, речь идет о передаче информации между ИЯ и ПЯ с учетом разницы структур этих языков. И чем выше разница, тем ниже качество конечного перевода. Например, рассмотрим следующее предложение: *This is definitely an indication of an intelligent and sensible approach to your entry into the dental profession*. Получаем следующий результат: *Это, безусловно, показатель интеллектуальных и разумный подход к вступлению в стоматологии*. Результат требует исправлений, поскольку члены сгенерированного предложения оказались плохо согласованными друг с другом. Рассмотрим теперь вариант перевода на один из европейских языков, например французский: *C'est certainement une indication d'une approche intelligente et sensible à votre entrée dans la profession dentaire*. Результат перевода с английского на европейский язык оказался намного лучше перевода на русский как представителя восточнославянских языков, что подтверждает наше вышеуказанное утверждение о том, что эквивалентность при переводе тем выше, чем теснее родство между языками.

*SMT* автоматически адаптируется к новой ситуации в переводе. Системы учитывают перевод новых слов, отдельных выражений, если их перевод поменялся относительно предыдущего раза. Им требуется необходимый корпус для качественной работы. Системы «запоминают» новые версии перевода и используют эти варианты в дальнейшем. Программа «живет» с развитием перевода, что ускоряет ее «обучение» и повышает качество конечного перевода. *RBMT* требуется вмешательство специалиста для внесения поправок во внутренние правила и вокабулярий программы. Зачастую сложно перепрограммировать модели работы программы, что существенно тормозит ее развитие в области выполнения качественного перевода.

*SMT* генерирует более «плавный» перевод, несмотря на то, что статистические системы, в отличие от *RBMT*, могут предложить менее предсказуемый вариант перевода, если используемый корпус в каждом отдельном случае является достаточно обширным. С другой стороны, системы *RBMT* могут не найти необходимого правила среди заложенных в них или подходящего лексического эквивалента для слова ИЯ, пропустив лексическую единицу, тем самым исказив смысл оригинального материала. Например, *Please contact me at this number, so that we can arrange a convenient time to meet*. Перевод *RBMT* системы: *Пожалуйста, свяжитесь со мной в этом числе, так, чтобы мы могли устроить удобное время, чтобы встретиться*. Общий смысл передан верно, но, как мы видим, система неверно интерпретировала слово *number* как число, а не как телефонный номер, следовательно, русскоговорящий читатель, незнакомый с английским языком, оказывается дезинформирован. Еще один пример: *The buyer hereby undertakes to accept the goods and pay for them in accordance with the terms of the contract*. Перевод *RBMT* системы: *Покупатель тем самым обязуется принять товары и плату в соответствии с условиями контракта*. Естественно, речь идет о получении товара и его последующей оплате, а не о получении товара вместе с оплатой. Система *SMT* предлагает нам следующий вариант: *Покупатель обязуется принять товар и оплатить их в соответствии с условиями договора*. Смысл передан верно, но система пропустила слово *hereby*, видимо, не найдя его в своем корпусе, что, как мы упоминали выше, говорит о непредсказуемости результатов ее работы.

Следующий пример продемонстрирует нам, что статистический машинный перевод хорошо показывает себя в определенной области, при учете наличия в ней двуязычного корпуса: *Unless we receive objection within ten days of your receipt of the order, we shall expect delivery of all ordered goods on the date indicated*. *SMT* перевод: *Пока мы не получили возражений в течение десяти дней после получения вами заказа, мы будем ожидать доставку всех заказанных товаров на указанную дату*. *RBMT* перевод: *Если мы не получим возражение в течение десяти дней после Вашей квитанции заказа, мы будем ожидать поставку всех заказанных товаров, в день обозначенных*. Очевидно, что первый вариант является более «плавным» и связанным при обработке нашего материала, деловой корреспонденции.

Одним из недостатков *SMT* систем является потребность в больших вычислениях, что требует мощного аппаратного обеспечения. Для «обучения» программы и ее специализации в какой-то конкретной области перевода требуются миллиарды вычислений. Для систем *RBMT* требуется внедрение необходимой лингвистической информации в начинку ядра программы специалистом, что требует время только на старте ее создания.

*RBMT* способен генерировать улучшенный вариант перевода с учетом вносимых коррекций в исходный вариант. В то время как *SMT* всегда генерирует похожий перевод, несмотря на вносимые изменения: *Meanwhile, we are holding your order aside with instructions to rush shipment as soon as we receive your payment*. Получаем следующие результаты от систем *SMT* и *RBMT* соответственно: *Между тем, мы ведем Ваш заказ*

с указанием в *сторону* спешить отгрузки, как только мы получим ваш платеж / Тем временем мы *держим* Ваш заказ в *стороне* инструкциями срочно отправить отгрузку, как только мы получаем Вашу оплату. Как видно из оригинального текста, сложность вызвал фразовый глагол *hold aside* – *зарезервировать*. Ради эксперимента заменим его синонимом *reserve* в той же временной форме. Вариант *RBMT*: Тем временем мы *резервируем* Ваш заказ с инструкциями срочно отправить отгрузку, как только мы получаем Вашу оплату. Замена данного глагола немного улучшила конечный вариант, он стал более понятным. Вариант *SMT*: Между тем, мы *резервирования* заказа с указанием спешить отгрузки, как только мы получим ваш платеж. Данный результат нельзя считать приемлемым, попробуем использовать другой синоним *save*: В то же время, мы *экономим* Ваш заказ с указанием спешить отгрузки, как только мы получим ваш платеж. Этот вариант также неудовлетворителен, так как все дальше уводит нас от оригинала. Следовательно, мы подтверждаем, что привязка к корпусу не позволяет *SMT* системе улучшить конечный результат. В то же самое время мы видим, что внесение изменений в *RBMT* системах напрямую влияет на качество перевода.

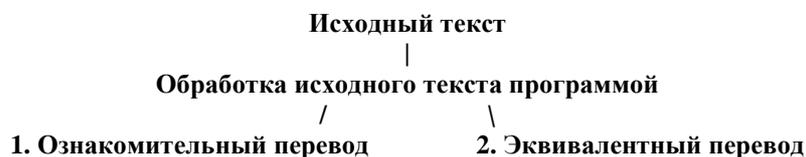
Из всего вышеизложенного можно сделать следующие выводы. Системы *RBMT* не требовательны к аппаратному обеспечению. Они обеспечивают приемлемое качество перевода текстов общего содержания. При этом их алгоритм работы дает пользователю возможность предвидеть конечный результат генерации текста. Возможность использования подключаемых внешних словарей расширяет потенциал программ, особенно при работе со специальной лексикой. Однако подобное расширение возможностей требует больше времени и определенных навыков работы с программой со стороны пользователя. При этом конечный перевод зачастую страдает от буквализма и не всегда оправдывает ожидания.

Что касается статистических систем перевода *SMT*, качество переведенных ими материалов полностью зависит от объема доступных им корпусов. Перевод является довольно «плавным», иногда напоминая результаты работы переводчика-человека, и зачастую оправдывает ожидания пользователя. Однако сами эти результаты очень сложно предсказать заранее. «Обучаемость» на основе корпусов улучшает качество работы программы, однако перевод на общие темы очень сильно уступает *RBMT* системам. Для работы и создания переводческих моделей этим системам требуется очень мощное аппаратное обеспечение.

Ниже в таблице представлены достоинства и недостатки двух типов систем.

Системы типа <i>RBMT</i>	Системы типа <i>SMT</i>
+ Определенная предсказуемость результатов перевода	- Отсутствие возможности предсказать конечный результат перевода
+ Приемлемое качество переводов текстов общей тематики	+ Приемлемое качество переводов текстов специальной тематики
+ «Знание» грамматических правил	- Отсутствие работы «по правилам»
+ Не требовательны к аппаратному обеспечению	- Требовательны к мощному аппаратному обеспечению
- Избыток буквализма	+ «Живой» перевод
- Требовательны к большим вложениям со стороны разработчиков. Требовательны к специальным знаниям со стороны простого пользователя	+ Наличие достаточно объемного корпуса ускоряет и улучшает работу программы, не требуя дополнительных действий
+ Возможность вносить изменения в исходный вариант, повышая качество конечного перевода	- Жесткая привязка к корпусу и отсутствие в нем эквивалента не дает возможности вносить изменения и улучшать качество переводимого текста

Итак, от принципов работы программ машинного перевода перейдем непосредственно к практическому их использованию. Для этого необходимо сформулировать основные цели применения подобных программ, основанные на желаемом конечном результате.



В первом случае это перевод, выполненный в ознакомительных целях, когда важно получить общее представление о содержании оригинала, не вдаваясь в детали. Требования к качеству, стилю и внешнему оформлению такого перевода будут ниже. Здесь допустимы мелкие неточности и пропуски несущественной для понимания информации. На первом месте будет стоять оперативность его выполнения. В качестве исходного текста могут быть четко структурированные документы небольшой сложности и объема.

Во втором случае это перевод, соответствующий оригиналу на всех релевантных уровнях и обеспечивающий решение тех же информационно-коммуникативных задач, на которые был нацелен текст оригинала. Иными словами, речь идет о тождестве при переводе, которое, согласно Г. Д. Воскобойнику, выступает как единство сторон антиномии и практически никогда не достигается.

Процесс преодоления различий между языками и культурами протекает в условиях когнитивного диссонанса, когда большинство решений оценивается переводчиком по степени приближения знака ПЯ к объекту ИЯ. Но остаточный когнитивный «зазор» - творческий стимул, побуждающий переводчика возвращаться к ранее переведенным текстам, когда возросшие профессиональные знания и опыт позволяют искать новые

пути к тождеству. Это также стимул для новых поколений переводчиков, обращающихся к работе над текстами, переведенными их предшественниками.

Движение к тождеству практически никогда не достигает абсолютного результата. Остающиеся различия между содержанием коммуникации на ИЯ и ПЯ отражены в переводческой эпистеме как когнитивный диссонанс. Последний - причина регулярной переоценки прежних результатов перевода, равно как повторных переводов отдельных произведений [2].

Также о невозможности абсолютного тождества в переводе говорит Л. А. Солдатова. Согласно ее определению, эквивалентность содержания двух текстов подразумевает общность их понимания, т.е. определенную степень общности информации (сведений, сообщаемых мыслей), которую могут извлечь из каждого текста лица, владеющие языком, на котором написан текст. Т.е. эквивалентность двух разных текстов предполагает не тождественность, а достаточную общность их содержания для целей коммуникации в конкретных условиях. И в процессе перевода, в зависимости от условий межъязыковой коммуникации, эквивалентность может достигаться на разных уровнях. А в отношении достижения какого-либо из уровней эквивалентности можно говорить также о возможных уровнях вариативности [6].

Справедливо это и по отношению к машинному переводу, особенно, если его целью является получение результата максимально тождественного оригинальному тексту, то обойтись без вмешательства человека – внесения поправок, редактирования, выбора вариантов перевода и работы над эквивалентностью, нам не представляется возможным.

Здесь мы предлагаем некоторые приемы повышения качества переводов машиной на примере деловой корреспонденции, а именно отрывка из договора между продавцом и покупателем:

*The buyer shall have the right to examine the goods on arrival, and within business days after such delivery he must give notice to the seller of any claim for damages on account of the condition, quality, or grade of the property, and must specify the basis of his claim in detail.*

Вариант перевода SMT: *Покупатель имеет право осмотреть товар по прибытии, и в течение рабочих дней после доставки таким он должен уведомить продавца о любой иск о возмещении ущерба в связи с состоянием, качеством или класса объекта недвижимости, и необходимо указать основу его иска в деталях.*

Вариант перевода RBMT: *Покупатель должен иметь право исследовать товары по прибытию, и в течение рабочих дней после такой поставки он должен дать уведомление продавцу любого требования о возмещении ущерба из-за условия, качества или сорта собственности, и должен определить основание его требования подробно.*

Как мы видим, общий смысл обе системы смогли передать читателю. Но текст нуждается в редактировании. Попробуем повысить качество перевода, для этого мы должны осуществить **предпереводческую** подготовку генерируемого текста, а именно сделаем его реферирование, сохранив основную идею, будем использовать только простые предложения и избегать пассивных конструкций: *The buyer shall have the right to examine the goods on arrival. He must give notice to the seller of claim for damages on account of the condition, quality, or grade of the property within business days after such delivery. He must specify the basis of his claim in detail.*

Вариант перевода SMT: *Покупатель имеет право осмотреть товар по прибытии. Он должен уведомить продавца иск о возмещении убытков в связи с состоянием, качеством или класса собственности в течение рабочих дней после такой доставки. Он должен определить основы своего требования в деталях.*

Вариант перевода RBMT: *Покупатель должен иметь право исследовать товары по прибытию. Он должен дать уведомление продавцу требования о возмещении ущерба из-за условия, качества или сорта собственности в течение рабочих дней после такой поставки. Он должен определить основание своего требования подробно.*

Конечно, оба этих варианта еще далеки от идеала. Есть проблемы на лексическом уровне, так, под словом *grade* понимается *норма качества (продукта или материала)*, но данная проблема может быть решена путем использования дополнительного «специализированного» словаря, который касается только определенной темы и содержит определенную терминологию, в нашем случае это деловая лексика.

Последний предлагаемый нами шаг это **постпереводческий** анализ полученного материала и его редактирование. На этом этапе переводчику-редактору необходимо вручную исправить сгенерированный текст, обратив особое внимание на синтаксис предложений и отдельных словосочетаний: *Покупатель имеет право проверить товар по прибытии и в течение рабочих дней с момента получения уведомить продавца, обратившись с жалобой на любые повреждения вследствие условий транспортировки, соответствии товара заявленным качествам, детально обосновав поданную жалобу.*

Итак, в целом преимущества использования программ машинного перевода любого типа достаточно очевидны. Во-первых, этот процесс требует меньше времени, а во-вторых, обходится дешевле найма переводческих кадров, особенно когда выполняемая работа срочная и имеет небольшой объем.

Итогом становится быстрый результат и относительная простота использования. Но несложно предугадать и его слабые места. Машинный перевод ограничен переводимым материалом, например, технический текст и научный стиль передаются на ПЯ без потери качества. В то время как любой другой «творческий» материал, например художественный текст, можно рассматривать как неподходящий, чтобы полностью доверить его машине. Кроме этого, машинный перевод находится за гранью контроля со стороны человека. Мы вводим данные и получаем сгенерированный текст, весь остальной процесс автоматизирован.

Как может показаться на первый взгляд, машинный перевод бросает вызов переводческому рынку. Однако, несмотря на его ценность и потенциал, а также прогресс в его развитии в последние годы, машинный перевод все еще остается простой генерацией текста.

## Список литературы

1. **Беляева Л. Н., Откупщикова М. И.** Автоматический (машинный) перевод // Прикладное языкознание. СПб.: Изд-во Санкт-Петербургского ун-та, 1996. С. 360-388.
2. **Воскобойник Г. Д.** Лингвофилософские основания общей когнитивной теории перевода: дисс. ... д-ра филол. наук. Иркутск, 2004. 296 с.
3. **Марчук Ю. Н.** Проблемы машинного перевода. М.: Наука, 1983. 233 с.
4. **Нелюбин Л. Л.** Компьютерная лингвистика и машинный перевод. М.: ВЦП, 1991. 207 с.
5. **Рябцева Н. К.** Информационные процессы и машинный перевод. М.: Наука, 1986. 167 с.
6. **Солдатова Л. А.** Межязыковая коммуникация и множественность перевода // Альманах современной науки и образования. Тамбов: Грамота, 2010. № 1. Ч. 1. С. 139-143
7. <http://letterworld.narod.ru>
8. <http://translate.google.ru/>
9. <http://www.isampleletter.com>
10. <http://www.promt.ru/>
11. <http://www.4hb.com>
12. **Hutchins W. J.** Machine Translation: Past, Present, Future. Ellis Horwood Series in Computers and their Applications. Chichester (UK): Ellis Horwood, 1986. 382 p.
13. **Hutchins W. J., Somers H.** An Introduction to Machine Translation. London: Academic Press, 1992. 362 p.
14. **Koehn P., Och F. J., and Marcu D.** Statistical Phrase Based Translation // NAACL '03 Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology. HLT-NAACL 2003. Edmonton, Canada, 2003. Volume 1. P. 48-54.

## ANALYSIS OF MODERN MACHINE TRANSLATION SYSTEMS OF SMT AND RBMT TYPE

**Islamov Roman Sergeevich**

**Fomin Andrei Gennad'evich**, Doctor in Philology

*Kemerovo State University*

*r.islamov87@gmail.com; andfomin67@mail.ru*

The authors reveal the essence of the work of modern machine translation systems of two types - SMT and RBMT, pay special attention to the analysis and comparison of these systems, as well as to the identification of advantages and disadvantages of each one, and in addition describe some practical recommendations on the improvement of machine translation quality.

*Key words and phrases:* translation; machine translation; statistical translation; "rule-based" translation; identity in translation; SMT; RBMT.

УДК 37

**Педагогические науки**

*Статья посвящена одной из важнейших проблем языковой и культурной политики – созданию программы непрерывного поликультурного образования, направленного на формирование и развитие межкультурной компетенции детей и учащейся молодежи в образовательных учреждениях Республики Саха (Якутия) всех уровней и форм. Анализируются документы правительства республики, программы и концепции поликультурного образования, восходящего от родной культуры к мировой через общероссийскую, начиная с детского возраста до окончания вуза. Предлагается идея разработки программы республиканского масштаба, способствующей повышению уровня межкультурной компетенции обучающихся и учитывающей особенности региона.*

*Ключевые слова и фразы:* непрерывное поликультурное образование; межкультурная компетенция; поликультурные компетенции; национально-культурная идентификация; система образования Республики Саха (Якутия); программа.

**Корякина Антонина Федоровна**

*НИИ национальных школ Республики Саха (Якутия)*

*mylove@mail.ru*

**ФОРМИРОВАНИЕ МЕЖКУЛЬТУРНОЙ КОМПЕТЕНЦИИ ДЕТЕЙ И УЧАЩЕЙСЯ МОЛОДЕЖИ  
В УСЛОВИЯХ НЕПРЕРЫВНОГО ПОЛИКУЛЬТУРНОГО ОБРАЗОВАНИЯ  
РЕСПУБЛИКИ САХА (ЯКУТИЯ)®**

Республика Саха (Якутия) является исторической родиной для ее коренных народов: якутов, эвенов, эвенков, чукчей, юкагиров. В то же время она является многонациональным регионом, где проживают люди различных