



PROMT®

Машинный перевод и
постредактирование

Содержание

- ✓ Место МП в переводческом процессе
- ✓ Экспресс-оценка применимости МП
- ✓ Улучшение качества МП
- ✓ Стратегии постредактирования МП
- ✓ Оценка эффективности интеграции МП

Место МП в переводческом процессе

Классическая схема:

Источник → переводчик → реципиент

Схема с МП:

Источник → МП → переводчик → реципиент

Эффекты от изменения схемы:

1. Фокус работы переводчика смещается с перевода на редактирование
2. В отличие от типичного процесса с живым переводчиком и редактором, здесь имеем дело с механизированной письменной речью (робот)

Место МП в переводческом процессе

Особенности работы с роботом-переводчиком

Преимущества:



1. Возможность настройки
2. Предсказуемость поведения
3. Отсутствие ошибок из-за человеческого фактора

Недостатки:



1. Требуется навыков и умения в обращении
2. Требуется понимания алгоритмов работы
3. Вероятность трудноисправимых «ошибок»

Содержание

- ✓ Место МП в переводческом процессе
- ✓ Экспресс-оценка применимости МП
- ✓ Улучшение качества МП
- ✓ Стратегии постредактирования МП
- ✓ Оценка эффективности интеграции МП

Экспресс-оценка применимости МП

Подход к оценке:

1. Акцент на удобство редактирования МП
2. Используемая технология МП и ее особенности
(перевод по правилам/статистика)
3. Применимость МП к текстам определенного типа
4. Возможности для повышения степени доверия к МП
(предсказуемость, настройка, feedback)



Технологии МП

Rule-based machine translation (RBMT)	Statistical machine translation (SMT)
Компоненты	Компоненты
<ul style="list-style-type: none">▶ двуязычные словари▶ морфологические таблицы▶ списки префиксов▶ базы имен	<ul style="list-style-type: none">▶ фразовая таблица▶ языковая модель
Преимущества	Преимущества
<ul style="list-style-type: none">▶ синтаксическая и морфологическая точность▶ стабильность и предсказуемость результата▶ возможность настройки на предметную область	<ul style="list-style-type: none">▶ быстрая настройка (параллельные тексты)▶ гладкость перевода
Недостатки	Недостатки
<ul style="list-style-type: none">▶ необходимость поддерживать и актуализировать лингвистические базы данных▶ «машинный акцент» при переводе	<ul style="list-style-type: none">▶ многочисленные грамматические ошибки▶ нестабильность перевода

Экспресс-оценка применимости МП

Эквивалентность по числу компонентов в предложении и их переводы

Исходный текст:

*Being a leading **group** in the field, we are seeking for **best** methods*

Машинный перевод:

*Будучи одной из ведущих в этой области, мы ищем наиболее **эффективные** методы*

Пропущенные составляющие: **group**

Неэквивалентный перевод: **best**

Экспресс-оценка применимости МП

Устойчивость соблюдения терминологии/знание предметной области

Исходный текст:

*The document contains a **heading** with the **heading data** inserted by the system.*

Машинный перевод:

*Документ содержит **заголовок** с **возглавляющими данными**, вставленными системой.*

Несоблюдение терминологии: **heading**

Экспресс-оценка применимости МП

Связность: слова-связки, служебные части речи, окончания (флексии)

Исходный текст:

*The **level** set by the user is incorrect.*

Машинный перевод:

***Уровня**, установленного пользователем является неверным.*

Ошибки в связности: падеж слова «уровень» и зависимых от него, пропущенная запятая после «пользователем»

Экспресс-оценка применимости МП

Критерий	Оценка	Вывод
Эквивалентность	Нет	МП придется использовать с особой осторожностью, проверяя каждое слово
	Да	Можно перейти к более серьезным вещам, чем подсчет значимых слов в исходном и переводном предложении
Терминология	Нет	Выделить время на проверку каждого термина по глоссарию
	Да	Найти более удачные решения по сочетаемости, тщательно выстроить логические связи между предложениями
Связность	Нет	МП придется использовать с особой осторожностью, проверяя каждое слово
	Да	Подумать о возможности разнообразить текст синонимами и/или тоньше проработать стиль

Содержание

- ✓ Место МП в переводческом процессе
- ✓ Экспресс-оценка применимости МП
- ✓ **Улучшение качества МП**
- ✓ Стратегии постредактирования МП
- ✓ Оценка эффективности интеграции МП

Улучшение качества перевода

Настройка МП – отдельный этап перед началом процесса перевода, который имеет смысл при:

1. Большом объеме однородных текстов
2. Жестких требованиях к терминологии и стилю
3. Работе в команде переводчиков
4. Крайне сжатых сроках



Улучшение качества перевода

Основные шаги:


1. Определение тематики подлежащего переводу текста и подбор подходящих (специализированных) словарей
2. Получение списка наиболее частых терминов-оборотов в переводимых текстах, определение их переводных эквивалентов и занесение в словарь
3. Получение списка наиболее часто встречающихся слов и проверка их текущего перевода системой МП
4. Установка требуемых правил перевода*
**технологии PROMT*



Улучшение качества перевода

Результаты:

Повышение оценки применимости МП по вышеперечисленным критериям, в частности:
выдерживание терминологии и предметной области



Снижение объема редактирования МП переводчиком:
за счет настройки предметной области и терминологии – с 80% до 30%

Содержание

- ✓ Место МП в переводческом процессе
- ✓ Экспресс-оценка применимости МП
- ✓ Улучшение качества МП
- ✓ **Стратегии постредактирования МП**
- ✓ Оценка эффективности интеграции МП

Стратегии постредактирования МП

Полное редактирование с проверкой
каждого компонента текста по исходному фрагменту

Исходный текст:

To this end, final analysis results are not quite obvious.

Машинный перевод:

С этой целью окончательные результаты анализа не совсем очевидны.

Перевод переводчика:

Заключение, полученное в результате анализа, не дало окончательного ответа на этот вопрос.

(При этом в глоссарии: analysis results = результаты анализа)
Объем постредактирования предложения: 100%

Стратегии постредактирования МП

Использование МП исключительно в качестве подсказки
по терминологии

Исходный текст:

To this end, final analysis results are not quite obvious.

Машинный перевод:

С этой целью окончательные результаты анализа не совсем очевидны.

Перевод переводчика:

Окончательные результаты анализа, тем не менее, не дали ответа на этот вопрос.

(При этом в глоссарии: analysis results = результаты анализа)

Объем постредактирования предложения: 82%

Стратегии постредактирования МП

Использование МП в качестве грамматического подстрочника
с подбором синонимичных выражений

Исходный текст:

SWIFT also brings the financial community together to work collaboratively to shape market practice and define standards

Машинный перевод:

SWIFT также примиряет финансовое сообщество для работы совместно, чтобы сформировать практику рынка и определить стандарты

Перевод переводчика:

SWIFT также объединяет финансовое сообщество для совместной работы в целях формирования рыночной практики и определения стандартов

Объем постредактирования предложения: 65%

Стратегии постредактирования МП

Чтение МП и исправление вызывающих сомнения мест с обращением к исходному тексту

Исходный текст:

Distributor may not take legal action against any infringer without SDA's prior written consent.

Машинный перевод:

Дистрибьютор может не подать в суд против никакого нарушителя без предварительного письменного согласия SDA.

Перевод переводчика:

Дистрибьютор не имеет права подавать в суд на какого-либо нарушителя без предварительного письменного согласия SDA.

Стратегии постредактирования МП

Чтение МП и исправление очевидных погрешностей без обращения к исходному тексту

Исходный текст:

Our study focuses on emergent visual rhetorical practices that can inspire advertisers.

Машинный перевод:

Наше исследование сосредотачивается на визуальных риторических методах на стадии становления, которые могут вдохновить рекламодателей.

Перевод переводчика:

*Наше исследование сосредотачивается на визуальных риторических приемах, **находящихся** на стадии становления, которые могут вдохновить рекламодателей.*

Объем постредактирования предложения: 14%

Стратегии постредактирования МП

Сводная таблица по стратегиям редактирования

Стратегия	Объем постредактирования	Имеет ли смысл использовать МП?	Имеет ли смысл настраивать МП?
Полное редактирование	100%	Нет	Нет
Подсказка по терминологии	80%	Только при работе в команде	Да
Грамматический подстрочник	60%	Да	Нет
Анализ сомнительных мест	35%	Да	Да
Корректурa	20%	Да	Да

Стратегии постредактирования МП

Основные критерии для выбора стратегии редактирования:

1. Знание предметной области
2. Степень владения исходным и целевым языком
3. Сроки и предполагаемый объем постредактирования
4. Опыт работы с данной системой МП
5. Степень доверия машинному переводу
6. Наличие технических средств для анализа внутренней структуры МП

Содержание

- ✓ Место МП в переводческом процессе
- ✓ Экспресс-оценка применимости МП
- ✓ Улучшение качества МП
- ✓ Стратегии постредактирования МП
- ✓ Оценка эффективности интеграции МП

Оценка эффективности интеграции МП

Основные эффекты от интеграции:

- ✓ Увеличение скорости работы переводчика
- ✓ Снижение объема работы при редактировании переводов, полученных от команды переводчиков
- ✓ Повышение качества готового перевода

Дополнительные затраты на интеграцию:

- ✓ Техническое и организационное внедрение, обучение
- ✓ Настройка МП по проекту

Влияющие факторы:

- ✓ Квалификация переводчиков
- ✓ Удобство технической инфраструктуры

Компания «Аргонавт»

- ✓ перевод документации в сфере Telecom + IT с английского на русский с 1992 г.
- ✓ основные заказчики: AT&T, Italtel, Samsung, Ericsson, ZTE, Siemens
- ✓ применение системы машинного перевода (MT) с 1993 г.
- ✓ применение систем Translation Memory (TM) с 1997 г.

www.argonaut.spb.ru

Опыт работы с МП

- ✓ Применение МП (PROMT) – в 97-99% случаев
- ✓ Производительность – 15-17 страниц (250 слов) в день (8-9 часов)
- ✓ Предварительная разработка терминологии и настройка словарей – в 100% случаев
- ✓ Техническое редактирование (в большинстве случаев) сводится к финальному контролю стиля перевода
- ✓ Предварительная разработка терминологии и настройка словарей – качественная база для адаптации новых переводчиков

Спасибо за внимание!

Юлия Епифанцева

директор по развитию

Julia.Epiphantseva@prompt.ru