

**Сравнительный анализ
результатов двух опросов о
постредактировании МП:
европейские и российские
заказчики и исполнители**

Clara Ginovart, Datawords

Светлана Светова, Т-Сервис

Постредактирование машинного перевода (Machine Translation Postediting, РЕМТ, МТРЕ...)

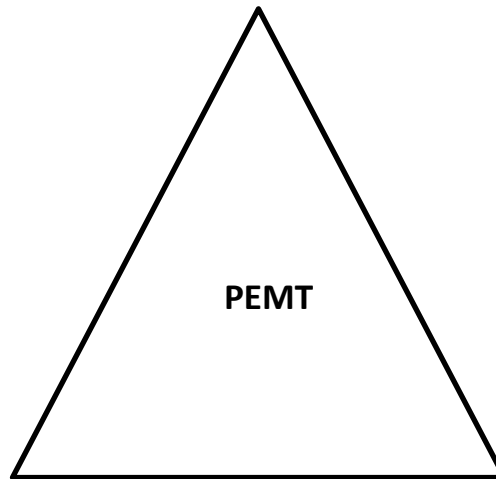
– это процесс доведения переводчиком/постредактором результата машинного перевода до требуемого уровня качества.

Очень краткая история

- 1950-1960 – большой интерес исследователей-теоретиков к постредактированию, формулирование основных его критериев, актуальных до сегодняшнего времени (постредактор – специалист по языку или по предметной области?), первые инструкции
- 1970 – первые коммерческие проекты по постредактированию, Еврокомиссия
- 1980-1990 – множество статей и конференций, но в фокусе не РЕМТ, а сам МП, а также появляется технология ТМ, которая все затмевает, плюс идеи об интеграции МП+ТМ
- С 2010 (примерно?) – мощный запрос на постредактирование

3 ГРУППЫ ЗАДАЧ

**Заказчик (прямой
или БП)**



Разработчик

1. Получить опыт внедрения
2. Улучшить качество МП на конкретном опыте (обратная связь и настройка)
3. Расширить применение (включая языковые пары)
4. Расширить клиентскую базу
5. Продать услуги по настройке и разработке
6. Продать услуги по обучению

1. Внедрить МП для сценариев, когда недостаточно ресурсов для выполнения перевода по традиционной схеме/ требуется увеличить объем перевода в рамках ограниченного бюджета/когда требуется увеличить объем перевода в пределах того же количества переводческих ресурсов
2. Купить и настроить систему (обратная связь)
3. Расширить применение (включая сферы и языки)
4. Обучить МП-специалистов
5. Повысить эффективность процессов

**Переводчик/
постредактор**

1. Получить больше объемов на перевод (включая те, которые не были бы получены без МП)
2. Выработать новые навыки (в т.ч. переводить больше и быстрее, не теряя в деньгах)
3. Выработать подход к новым условиям

Два опроса о постредактировании машинного перевода в Европе



- В 2019 году **Клара Джиноварт (Clara Ginovart)** разработала и провела два опроса, которые отразили представления европейской переводческой отрасли в целом - и практикующих постредакторов МП в частности - о современном процессе постредактирования машинного перевода (РЕМТ).
- Клара Джиноварт – консультант по технологиям перевода в Datawords, а также аспирант Университета Помпеу Фабра (Испания). Получив степень бакалавра по переводам в УПФ, а также степень магистра в Женевском университете, Клара работает в сфере локализации и занимается исследованиями профессиональной переводческой деятельности с применением МП (известной как постредактирование).
- Первый опрос был разработан **для всех участников переводческой отрасли**, второй – непосредственно **для постредакторов**.
- Результаты опросов были обработаны летом 2019 года, и с их обзором можно ознакомиться [по этой ссылке](#).

Два опроса о постредактировании машинного перевода в России

- В начале 2020 года [Светлана Светова](#), автор сайта pemt.ru, адаптировала оба опроса для российского рынка переводческих услуг.
- Опросы на русском языке:
- [опрос для всех участников переводческой отрасли \(73 вопроса\)](#);
- [опрос для постредакторов \(81 вопрос\)](#).
- Результаты, а также сравнительный анализ будут впервые представлены на ТФР 2020.

Участники опросов (бизнес)

- Европа

БП – 88%

Компании – 12 %

- Россия

БП – 55%

Компании – 45%

Размер компании (**бизнес**)

- Европа

От 11 до 50 сотрудников – 48%

Свыше 500 – 12%

От 50 до 500 – 20%

Менее 10 – 20%

- Россия

От 11 до 50 сотрудников – 54%

51 до 500 – 45%

Свыше 500 – 1%

Опыт постредактирования МП (бизнес)

- Европа

От 1 года до 5 лет – 54%

Свыше 5 лет – 32%

менее 1 года – 14%

- Россия

От 1 года до 5 лет – 64%

Свыше 5 лет – 9%

Менее 1 года – 27%

Самый распространенный исходный язык для постредактирования (**бизнес**)

- **Европа**

Исходный английский, а перевод на испанский, французский, немецкий, итальянский и нидерландский.

- **Россия**

Исходный английский, второй по популярности русский. Целевой – наоборот. Дополнительно: французский, немецкий.

Пул постредакторов-фрилансеров (бизнес)

- Европа

1-15 человек – 30%

15-30 – 18%

Свыше 30 – 37%

Только штатные – 15%

- Россия

1-15 человек – 37%

15-30 – 27%

Только штатные – 36%

Пул штатных постредакторов (бизнес)

- Европа

1-15 человек – 52%

15-30 – 1%

Свыше 30 – 5%

Только фрилансеры – 42%

- Россия

1-15 человек – 100%

Ожидания по скорости постредактирования (**бизнес**)

- **Европа**

1000-3000 слов в час – 44%

3000-5000 – 3%

Менее 1000 – 53%

- **Россия**

1000-3000 слов в час – 27%

Менее 1000 – 73%

Задание на постредактирование (бизнес)

- Европа

Включает всю информацию об использованном модуле МП, настройке, метриках качества, инструкции - 9%

Некоторая информация о процессе и инструкции – 65%

Ничего такого – 26%

- Россия

Некоторая информация о процессе и инструкции (руководства по постредактированию, уровни, описание контента) - 55%

Ничего такого – 45%

Инструкции постредактору (**бизнес**)

- **Европа**

Общие и отлаженные внутри компании – 22%

С учетом типа контента и языка – 65%

Только указание уровня постредактирования (light или full) – 11%

Инструкции TAUS – 2%

- **Россия**

Разработанные в компании – 55%

Инструкции TAUS – 18%

Только уровни постредактирования – 18%

Никакие – 9%

Ресурсы для выполнения задач по постредактированию (**бизнес**)

- **Европа**

85% – пул постредакторов-фрилансеров

15% – собственные ресурсы

- **Россия**

85% – пул постредакторов-фрилансеров

15% – собственные ресурсы

Кто принимает решение о применении МП (бизнес)

- Европа

Заказчики – 56%

Сами – 41%

- Россия

Заказчики – 82%

Сами – 18%

Участники опроса (постредакторы)

- Европа

Фрилансеры или внештатные переводчики

- Россия

Фрилансеры, штатные редакторы, штатные
руководители групп переводчиков

Самый распространенный исходный язык для постредактирования (**постредакторы**)

- **Европа**

Английский, а целевой испанский, французский, немецкий, итальянский и нидерландский

- **Россия**

Английский, второй русский, перевод на русский, английский, итальянский

Опыт постредактирования МП (постредакторы)

- Европа

От 1 года до 5 лет – 59%

Свыше 5 лет – 17%

менее 1 года – 24%

- Россия

От 1 года до 5 лет – 47%

Свыше 5 лет – 7%

Менее 1 года – 47%

Источник получения работы по постредактированию (**постредакторы**)

- Европа

От БП – 79%

От прямых заказчиков – 21%

- Россия

От БП – 80%

От прямых заказчиков – 20%

Могут ли самостоятельно решать вопрос о применении МП (**постредакторы**)

- Европа

Не могут – 75%

Могут – 18%

- Россия

Не могут – 80%

Могут – 20%

Основной источник МП (постредакторы)

- Европа

Google Translate – 33%

DeepL – 13%

Разработка компании – 37%

- Россия

Google Translate – 53%

DeepL – 53%

SDL MT – 13%

Разработка компании – 7%

На основании чего следует принимать решение о пригодности результата МП для постредактирования (**постредакторы**)?

- **Европа**

- Ручная оценка в целом – 39%
- **Автоматическая оценка – 48%**
- И то, и другое – 13%

- **Россия**

- **Ручная оценка на основе типологии ошибок – 53%**
- Ручная оценка расстояния редактирования на основе постредактирования – 27%
- Автоматическая оценка (такие показатели, как BLEU, TER, METEOR и пр.) – 13%

Сравнение бизнес-постредакторы

	Европа (бизнес)	Европа (постредакторы)	Россия (бизнес)	Россия (постредакторы)
РЕМТ составляет до 25% от общего объема работ	73%	62%	91%	100%
РЕМТ составляет от 26 до 50% от общего объема работ	21%	16%	9%	0%
Постредактирование контента для публичного распространения		80%	45%	53%
Постредактирование контента для ограниченного распространения		20%	55%	47%
Заказчик требует полное постредактирование	73%		82%	60%
Заказчик требует частичное постредактирование	21%		18%	27%

Сравнение бизнес-постредакторы

	Европа (бизнес)	Европа (постредакторы)	Россия (бизнес)	Россия (постредакторы)
Задача постредактирования - исправление плохого качества до достижения идеального		56%	55%	40%
Обратная связь по качеству МП требуется	73%		73%	60%
Обратная связь по качеству МП не требуется		70%	27%	40%
Инструкции по постредактированию предоставляются в подробном виде	87%		91%	47%
Инструкции по постредактированию не предоставляются вообще		46%	9%	53%

Сравнение бизнес-постредакторы

	Европа (бизнес)	Европа (постредакторы)	Россия (бизнес)	Россия (постредакторы)
Отслеживание производительности	50% - собственные средства и Excel, а также TMS	55% постредакторов отслеживают производительность, половина в Excel, 10% используют собственные средства, менее 10% - системы TMS	18% - TMS, 18% - нет, Excel - 55%, собственное - 9%	60% постредакторов отслеживают производительность, 40% - нет
Предпочитаемая модель расчета: за исходное слово	71%	57%	55%	80%
Предпочитаемая модель расчета: почасовая	6%	25%	45%	20%

Сравнение бизнес-постредакторы

	Европа (бизнес)	Европа (постредакторы)	Россия (бизнес)	Россия (постредакторы)
Самая популярная система МП для выполнения постредактирования	SDL Trados Studio	SDL Trados Studio	SDL Trados Studio	SDL Trados Studio
Самое популярное средство для контроля качества (QA)	Verifika Xbench и QA Distiller	Verifika Xbench	Verifika Xbench и QA Distiller	Verifika Xbench и чек-листы заказчика

Сравнение бизнес-постредакторы

	Европа (бизнес)	Европа (постредакторы)	Россия (бизнес)	Россия (постредакторы)
Считают ли адекватными имеющиеся курсы по обучению РЕМТ	42% - нет, 36% - да	41% - нет, 45% - да	73% - нет, 27% - да	47% - нет, 53% - да
Никаких курсов по постредактированию	67% не организовывали	53% не проходили	18%	47%
Знание стандартов ISO 18587:2017 и ASTMWK46396	45% - да	14% - да, 45% - нет	55%	60% - да

Оценка навыков и компетенций согласно их важности для профессионального постредактора (бизнес, Россия, топ-5 ответов)

- Умение принимать решение, редактировать или отвергать (переводить с нуля) результат МП
- Умение редактировать согласно руководствам по постредактированию
- Умение выявлять ошибки МП
- Умение добиваться качества "ручного перевода" при постредактировании
- Умение предоставлять обратную связь для разработчиков решений МП

Какие критерии должны учитываться при найме или отборе постредакторов (постредакторы, Россия, топ-7 ответов)

- Знание предметной области
- Опыт постредактирования МП
- Производительность (скорость постредактирования)
- Навыки редактирования и коррекции текста
 - Навыки в области контроля качества
 - Навыки использования технологий
 - Технические навыки: макросы и пр.

To Be Continued...

В следующей серии ответы на вопросы...

- Для каких типов документов применяется постредактирование?
- Каков самый распространенный сценарий применения постредактирования?
- В каком порядке вы читаете оригинал и перевод при постредактировании?
- Проверяет ли второй переводчик/редактор тексты после постредактирования МП?
- В каких случаях качество вашего постредактирования признавалось неудовлетворительным?
- Какие данные постредакторы присылают по завершении проекта по постредактированию МП?
- Должны ли постредакторы предоставлять заказчику отзывы по качеству МП для дальнейшей настройки?
- Выполняют ли штатные постредакторы постредактирование 100% своего времени?
- Задачи для штатных и внештатных постредакторов?
- Проходили ли вы обучение и получали ли какие-либо из следующих сертификатов (TAUS, ProZ, SDL...)?
- Какие элементы включает в себя обучение постредактированию МП?
- Кто проводит обучение постредактированию МП?

Вопросы?

