

# АССОЦИАЦИЯ ПЕРЕВОДЧЕСКИХ КОМПАНИЙ (АПК)

## СТАНДАРТ АССОЦИАЦИИ ПЕРЕВОДЧЕСКИХ КОМПАНИЙ. СРЕДСТВА АВТОМАТИЗИРОВАННОЙ И АВТОМАТИЧЕСКОЙ ГЕНЕРАЦИИ ТЕКСТА В ПРОФЕССИОНАЛЬНОМ ПЕРЕВОДЕ. МЕТОДЫ ОБРАБОТКИ РЕЗУЛЬТАТОВ РАБОТЫ СРЕДСТВ И ОЦЕНКА ИХ ЭФФЕКТИВНОСТИ (ПАМЯТЬ ПЕРЕВОДОВ, МАШИННЫЙ ПЕРЕВОД, БОЛЬШИЕ РЕЧЕВЫЕ МОДЕЛИ)

(СТАНДАРТ АПК по АГТ)

\*\*\*

### ***NON-HUMAN TEXT GENERATION FOR AUGMENTED PROFESSIONAL TRANSLATION AND INTERPRETATION SERVICES (TM, MT, LLM). STANDARD DEVELOPED BY THE RUSSIAN ASSOCIATION OF TRANSLATION COMPANIES***

Ключевые слова: перевод, межъязыковой перевод, технологии перевода, автоматизированный перевод, автоматический перевод, автоматизированная генерация текста, автоматическая генерация текста, машинный перевод, большая речевая модель, большая языковая модель, предобработка, предредактирование, постобработка, постредактирование, постредактор, переводчик-постредактор, услуги по переводу, качество перевода, нормативы перевода, память переводов.

## ВВЕДЕНИЕ

Стандарт «Средства автоматизированной и автоматической генерации текста в профессиональном переводе. Методы обработки результатов работы средств и оценка их эффективности (память переводов, машинный перевод, большие речевые модели)» разработан комитетом Ассоциации переводческих компаний «Профессиональное применение машинного перевода и прочих средств автоматической генерации текста» (Комитет АПК по МП) в составе представителей компаний (в алфавитном порядке):

- ООО «АКМ-Вест»;
- ООО «АЛЬБА»;
- ООО «Амберит СНГ»;
- ООО «НЕОТЭК»;
- ООО «СТАР СПб»;
- ООО «Т-Сервис».

Стандарт «Средства автоматизированной и автоматической генерации текста в профессиональном переводе. Методы обработки результатов работы средств и оценка их эффективности (память переводов, машинный перевод, большие речевые модели)» одобрен общим собранием членов Ассоциации переводческих компаний 18 марта 2026 года.

*Разработчики (в алфавитном порядке):*

Акопян Полина Павловна;

Берендяев Максим Викторович;

Бикматов Ренат Рустемович;

Ларин Александр Александрович;

Светова Светлана Юрьевна.

*Эксперты (в алфавитном порядке):*

Сант Сергей Альфредович.

*По всем вопросам обращаться по адресу:*

[standard@pemt.ru](mailto:standard@pemt.ru).

## СОДЕРЖАНИЕ

<b>ВВЕДЕНИЕ</b> .....	<b>2</b>
<b>1. Область применения</b> .....	<b>4</b>
<b>2. Используемые сокращения</b> .....	<b>4</b>
<b>3. Термины и определения</b> .....	<b>4</b>
<b>4. Автоматизированная и автоматическая генерация текста в переводе</b> .....	<b>10</b>
4.1. Автоматизированная генерация текста в переводе (CAT-инструменты на основе памяти переводов).....	10
4.2. Автоматическая генерация текста в переводе .....	11
4.2.1. Машинный перевод.....	11
4.2.2. Большие речевые модели.....	12
<b>5. АГТ в переводческом процессе</b> .....	<b>14</b>
5.1. Профессиональный переводческий процесс .....	14
5.2. Место АГТ в переводческом процессе .....	14
<b>6. Подготовка к АГТ (предобработка)</b> .....	<b>16</b>
<b>7. Постобработка результатов (выдачи) АГТ</b> .....	<b>17</b>
7.1. Этапы постобработки .....	18
7.2. Глубина постобработки .....	19
7.2.1. Техническая постобработка .....	19
7.2.2. Выборочная постобработка .....	20
7.2.3. Полная постобработка.....	20
<b>8. Процесс перевода и постобработки АГТ без взаимодействия с исполнителем на подготовительном этапе</b> .....	<b>20</b>
<b>9. Оценка рисков. Технические требования заказчика</b> .....	<b>22</b>
<b>Приложение А. Основные требования к качеству перевода</b> .....	<b>23</b>
<b>Приложение Б. Методология аналитической оценки качества перевода с типологией ошибок и формулой формирования балльной оценки качества с калибровкой на основе метрики Multidimensional Quality Metrics (MQM)</b> .....	<b>26</b>
<b>Приложение В. Пример требований к профессиональной квалификации переводчиков-постредакторов</b> .....	<b>28</b>
<b>Приложение Г. Пример требований к программам обучения переводчиков-постредакторов</b> .....	<b>29</b>
<b>Приложение Д. Пример профиля должности переводчиков-постредакторов</b> .....	<b>30</b>
<b>Приложение Е. Факторы риска, влияющие на применимость средств АГТ</b> .....	<b>31</b>
<b>Приложение Ж. Пример рекомендаций докладчику, речь которого подлежит распознаванию или переводу с использованием средств АГТ</b> .....	<b>32</b>
<b>Приложение З. Пример требований к системам АГТ (для разработчиков)</b> .....	<b>33</b>
<b>Приложение И. Нормативы постобработки и расчет размера оплаты</b> .....	<b>34</b>
<b>Приложение К. Библиография</b> .....	<b>35</b>

## 1. Область применения

Настоящий Стандарт устанавливает требования к задачам, процессам и ресурсам, а также определяет иные аспекты оказания услуг с применением средств **автоматизированной и автоматической генерации текста** (далее — АГТ) и его **дальнейшей постобработки** в целях профессионального перевода с одного естественного языка на другой на основании исходного материала, представленного в виде письменного текста или устной речи.

Цель Стандарта заключается в минимизации рисков наступления негативных последствий имущественного и репутационного характера, в уменьшении опасности для жизни и здоровья людей, окружающей среды, для безопасности государства и его суверенитета, правосудия, возникающих в результате автоматической генерации перевода текста и речи.

Стандарт способствует сохранению кадрового потенциала специалистов в области профессионального перевода, их обучению и повышению квалификации при применении средств автоматизированной и автоматической генерации текста.

Стандарт описывает **процессы и правила использования АГТ при профессиональном переводе**, следование которым минимально необходимо для достижения качества перевода, отвечающего всем потребностям и требованиям заказчика в конкретной коммуникативной ситуации, и обеспечивает пригодность перевода для применения, распространения или публикации.

Стандарт предназначен для применения исполнителями и заказчиками услуг перевода (лингвистических услуг), специалистами в области перевода, операторами (провайдерами) и разработчиками средств и систем автоматизированной и автоматической генерации текста для перевода.

Стандарт рекомендован к использованию заказчиками для формирования закупочной документации на услуги профессионального перевода, лингвистические услуги, услуги локализации, услуги межъязыкового посредничества, на системы автоматизированного и автоматического перевода.

Стандарт может также быть использован учебными заведениями при разработке и реализации образовательных программ.

## 2. Используемые сокращения

АГТ — автоматизированная или автоматическая генерация текста.

БРМ — большая речевая модель, тж. большая языковая модель (БЯМ), тж. LLM.

МП — машинный перевод.

ТЗ — требования заказчика или требования, изложенные в техническом задании.

САТ-инструменты (англ. CAT tools) — программное обеспечение для перевода с применением средств автоматизации.

ТМ (англ.) — память переводов или база переводов (от англ. translation memory).

## 3. Термины и определения

Термины, используемые в области автоматизированной и автоматической генерации текста в профессиональном переводе:

**Автоматизированный перевод:** выполнение перевода текста человеком с помощью компьютерных программ.

**Автоматизированная генерация текста:** работа, осуществляемая CAT-инструментами, заключающаяся в формировании интерактивных подсказок для переводчика, составленных на базе TM, результатом которой является в том числе претрансляция.

**Автоматическая генерация перевода, тж. автоматическая генерация текста:** создание переводов и текстов компьютерными программами в автоматическом режиме

**Автоматическая оценка качества перевода:** оценка качества перевода с применением специальных метрик, технологий и компьютерных программ.

**Автоматически сгенерированный текст, автоматически сгенерированный перевод:** текст или перевод, которые получены с помощью компьютерной программы без участия или при минимальном участии человека.

**Автоматический перевод:** перевод текста с одного естественного языка на другой компьютерными программами без вмешательства человека.

**Аналитическая оценка качества перевода:** оценка качества перевода, при которой осуществляется тщательный анализ текста перевода, выявляются ошибки, которые классифицируются по категориям и степени критичности.

**База переводов:** база данных специального формата (TM), содержащая пары сегментов на исходном и целевом языках.

**Большая речевая модель, тж. большая языковая модель (англ. LLM):** комплексная нейромодель, использующая искусственные нейронные сети на основе трансформеров, обученные на очень больших объемах текстовых данных методом самостоятельного обучения, с модулем интерактивного диалога с пользователем, обученным с подкреплением, для генерации ответа на запрос на естественном языке.

**Буквальный перевод, тж. дословный перевод, подстрочный перевод:** осуществление в процессе перевода подстановки слов языка перевода на место слов исходного языка; часто характеризуется сохранением синтаксической структуры и порядка слов исходного языка, а также результат такого перевода.

**Верстальщик:** специалист по верстке.

**Верстка:** процедура формирования документов из текстовых и графических объектов.

**Выдача, тж. выдача МП, выдача АГТ:** результат работы систем автоматической генерации текста или перевода.

**Генерация:** создание в автоматическом режиме.

**Гладкость:** свойство текста перевода быть легко читаемым.

**Глоссарий, тж. терминологическая база, терминологический словарь, термбаза:** словарь узкоспециализированных терминов в какой-либо отрасли знаний с переводом на другой язык, возможными комментариями и примерами, в том числе в электронном виде.

**Глубина редактирования, постредактирования, постобработки:** разница между двумя сегментами текста, определяемая как минимальное число односимвольных вставок, удалений и замен, необходимых для преобразования одного сегмента в другой.

**Естественный язык:** язык человеческого общения.

**Заказчик:** организация или лицо, инициирующие процесс перевода с установлением определенных требований.

**Исполнитель:** организация или лицо, выполняющие перевод с учетом установленных стандартов, требований и с применением определенных инструментов.

**Исходный текст:** текст, который подлежит переводу на другой язык.

**Исходный язык:** язык, на котором создан исходный текст.

**Качество перевода:** совокупность свойств результата перевода, обуславливающих его пригодность удовлетворять определенные потребности в соответствии с назначением.

**Качество выдачи:** уровень качества сгенерированного средствами АГТ текста или перевода.

**Количественная оценка качества перевода:** числовой показатель, выражающий уровень качества перевода по заданной шкале; формируется на основе измеримых критериев (например, точности передачи смысла, грамматической и стилистической правильности в языке перевода).

**Коммуникативная ситуация:** система условий взаимодействия участников коммуникации посредством перевода.

**Компетентность:** способность и умение эффективно применять свои знания и умения для решения задач в определенной области.

**Компетентность в предметной области:** осведомленность в определенной специальной предметной (тематической) области и владение профессиональным языком и терминологией этой области, являющиеся результатом получения профессионального опыта переводчиком или специального образования и достаточные для правильного понимания переводчиком исходного текста и его адекватной передачи в переводе.

**Контроль качества:** процесс, выполняемый для обеспечения соответствия перевода установленным стандартам качества и требованиям заказчика.

**Корпус (текстов, для обучения, языковой):** множество текстов и их переводов на один или несколько языков.

**Корректор:** специалист, осуществляющий корректуру текста.

**Корректурa:** выполнение проверки текста на соответствие нормам грамматики, орфографии и пунктуации и исправление выявленных ошибок.

**Лингвистические услуги:** наиболее распространенными лингвистическими услугами являются письменный перевод, локализация, устный перевод, транскрипция, озвучка, оценка качества перевода и текстов.

**Локализация:** перевод интерфейса пользователя, документации и сопутствующих файлов программного обеспечения с одного языка на другой с учетом специфики, законодательства и культуры страны и региона, для которых выполняется перевод.

**Локаль:** информация или набор характеристик, специфичные для языковых, культурных, технических, географических и прочих особенностей целевой аудитории.

**Машинный перевод:** перевод текста с одного естественного языка на другой с помощью компьютерной программы; машинный перевод относится к области автоматической обработки языка и может выполняться с использованием различных инструментов — от систем, основанных на лингвистических правилах либо статистике, до нейросетевых моделей, включая современные системы генерации текста.

**Межкультурная коммуникация:** коммуникация между представителями разных культур, национальностей, стран.

**Межъязыковое посредничество:** содействие в межкультурной коммуникации на разных языках.

**Межъязыковой перевод:** перевод между разными языками.

**Метаданные:** данные, которые предоставляют информацию о других данных.

**Механизм внимания:** механизм, используемый в нейронных сетях для поиска взаимосвязей между различными частями входных и выходных данных.

**Модуль машинного перевода (МП):** компьютерная программа для выполнения машинного перевода.

**Направление перевода:** указание исходного языка и языка перевода.

**Нейронная модель, нейромодель:** математическая модель на основе нейронной сети.

**Нейронная сеть:** математическая модель обработки информации, основанная на принципах работы, сходных с нервной системой человека, в которой отдельные элементы (нейроны) соединены в сеть и обмениваются сигналами; такая архитектура позволяет имитировать процессы обучения, распознавания и принятия решений.

**Нейронный машинный перевод:** тип машинного перевода, при котором используется большая искусственная нейронная сеть.

**Нормативные документы:** документы, устанавливающие правила, общие принципы или характеристики, касающиеся различных видов деятельности или их результатов.

**Нормативы перевода:** количественные характеристики выполнения переводов.

**Оценка качества (целостная, экспертная, с калибровкой):** процесс измерения уровня качества текста или перевода.

**Память переводов, тж. переводческая память:** технология создания, поддержки и использования баз специального формата (TM), содержащих сегменты на исходном и целевых языках перевода и, возможно, метаданные, а также сама такая база данных.

**Переводчик:** специалист, осуществляющий перевод текста.

**Переводческая компания:** предприятие сферы услуг, занимающееся переводами.

**Переводчик-постредактор, тж. постредактор:** специалист, который выполняет постобработку результатов работы средств автоматической генерации текста в переводе.

**Пилотный проект:** проект для предварительного изучения возможностей применения средств АГТ в разных режимах и коммуникативных ситуациях.

**Полное совпадение:** полное совпадение сегмента обрабатываемого исходного текста с сегментом в выбранной базе переводов (TM).

**Пользовательская настройка, пользовательское обучение:** настройка компьютерной программы для выполнения переводов конкретного типа текстов в определенной предметной области.

**Постобработка:** редактирование и корректура результатов АГТ с целью доведения их качества до требуемого уровня.

**Постредактирование:** возможный этап постобработки, заключающийся в редактировании и корректуре результатов АГТ с целью доведения их качества до требуемого уровня.

**Предобработка:** обработка исходного текста перед применением средств автоматической генерации текста для получения результата более высокого качества.

**Предредактирование:** возможный этап предобработки, заключающийся в редактировании и корректуре результатов АГТ с целью доведения их качества до требуемого уровня.

**Предметная область:** сфера человеческой деятельности, выделенная и описанная согласно установленным критериям; в описываемое понятие должны входить сведения об ее элементах, явлениях, отношениях и процессах, отражающих различные аспекты этой деятельности.

**Премиальный перевод:** перевод самого высокого уровня качества.

**Претрансляция, предперевод:** автоматическая подстановка переводов для сегментов или из базы ТМ, или из иного источника (например, из МП).

**Промпт:** текст на естественном языке, который служит «запросом» или «инструкцией» нейронной сети для выполнения конкретных задач по генерации требуемого результата.

**Профессиональный перевод:** процесс создания творческого результата интеллектуальной деятельности автора (группы авторов), осуществляемый в условиях свободы выбора технических средств, стратегии, тактики и приемов осуществления их деятельности.

**Разверстка:** приведение исходного текста к электронному виду, в том числе пригодному для использования средств автоматизации перевода.

**Расшифровка:** процесс нахождения исходного сообщения в закодированном (зашифрованном) материале и сам результат такого процесса.

**Редактор:** специалист, осуществляющий редактирование (редактуру) текста.

**Результат машинного перевода:** результат перевода текста с одного естественного языка на другой с помощью системы МП.

**Руководство по постобработке:** инструкции по выполнению постобработки автоматически сгенерированного текста (указания, приемы, действия, инструменты, списки типовых ошибок).

**Ручной перевод:** перевод, выполненный человеком.

**Сегмент:** одноязычный или двуязычный фрагмент текста для обработки с помощью систем автоматической или автоматизированной генерации текста.

**Система ТМ:** программное обеспечение, работающее с базами данных особого формата, в которых хранятся сегменты исходного текста вместе с результатами перевода, а также дополнительная информация.

**Система МП:** программное обеспечение для выполнения машинного перевода.

**Словарь:** структурированный набор статей, имеющих значение, соответствующее каждой статье, идентифицирующей одно значение, в том числе в электронном виде.

**Сопоставление, выравнивание (элайн):** создание и пополнение базы переводов путем обработки переводов, выполненных ранее без использования CAT-инструментов.

**Специализированный модуль МП:** модуль машинного перевода после настройки для перевода в определенной предметной области (областях).

**Специалист в области перевода:** лицо, участвующее в организации и оказании услуг перевода на основании трудового или гражданско-правового договора.

**Стилистическое редактирование, стилистическая редакция:** улучшение стиля текста перевода, структуры предложений с целью сделать его более понятным и стилистически приемлемым.

**Стратегия постобработки:** выработка сценария работы с результатами АГТ с целью наиболее эффективного доведения их качества до требуемого уровня.

**Текст:** языковое содержание, любая последовательность графических или звуковых языковых знаков, ограниченная единым назначением.

**Техническая гарантия:** оценка возможных технических требований к качеству и предоставление юридических гарантий соответствия результата ТЗ с участием человека.

**Техническая документация:** набор документов, используемых при проектировании (конструировании), создании (изготовлении) и использовании (эксплуатации) каких-либо технических объектов: зданий, сооружений, промышленных товаров, программного и аппаратного обеспечения.

**Трансформер:** архитектура глубоких нейронных сетей, основанная на механизме внимания.

**Уровень качества:** относительная характеристика качества перевода, основанная на сравнении значений показателей качества оцениваемого перевода с базовыми значениями соответствующих показателей.

**Целевая аудитория:** потенциальные потребители перевода.

**Юридическая гарантия:** набор средств для обеспечения правовой охраны перевода в качестве объекта авторского права, а также ответственность за содержательную эквивалентность текста перевода исходному тексту.

**Язык перевода:** язык, на который выполняется перевод.

**Язык:** здесь — естественный язык общения.

**Языковая пара, тж. направление перевода:** комбинация исходного языка и языка перевода.

## 4. Автоматизированная и автоматическая генерация текста в переводе

Принятое в Стандарте сокращение «АГТ» служит для обозначения комбинации двух способов генерации текста — **автоматизированной и автоматической**. Общим для них является использование специальных компьютерных программ, а различие заключается в степени вовлеченности человека.

### 4.1. Автоматизированная генерация текста в переводе (CAT-инструменты на основе памяти переводов)

Традиционной технологией для автоматизации профессионального перевода является «память переводов» (от англ. translation memory, TM).

Использование технологии TM наиболее эффективно для решения задач перевода документов, имеющих высокую степень повторяемости (техническая документация, описание продуктов, договоры, контракты, инструкции и подобные тексты).

Память переводов, или база переводов, представляет собой базу данных специального формата, содержащую двуязычные или многоязычные сегменты текстов в виде «исходный текст — перевод — метаданные».

Процесс перевода с использованием TM построен по принципу накопления: после подтверждения перевода в CAT-инструменте в базе переводов сохраняется исходный текст сегмента и его перевод, а также метаданные — дополнительная информация о характеристиках проекта, файле, исполнителе, дате, времени и т. п.

Технология TM основана на сравнении текста, который необходимо перевести, с сегментами текстов, ранее выполненные переводы которых хранятся в базе памяти, а также на сравнении сегментов переводимого текста между собой в целях расчета потенциальной экономии от использования технологии TM. Если идентичный или частично совпадающий с текущим сегментом исходный текст найден, то его соответствующий перевод отображается в CAT-инструменте вместе с указанием степени совпадения (обычно в процентах или условных обозначениях) или (в зависимости от настроек) автоматически подставляется в поле для перевода.

Кроме того, в CAT-инструментах реализована дополнительная функциональность, позволяющая вести и подключать при переводе глоссарии специального формата для унификации терминологии и автоматизации процедур контроля качества. В современных CAT-инструментах предусматривается возможность подключения таких систем АГТ, как МП и БРМ, для дополнения перевода, выполненного на основе TM, автоматически сгенерированной выдачей, в особенности для тех случаев, когда в базе переводов для исходного сегмента не обнаружено совпадение необходимого установленного уровня (интеграция TM и АГТ в CAT-инструментах).

Накопленные в результате длительного применения технологии TM базы памяти переводов, выполненных человеком, являются основным корпусом обучения для настройки систем МП, включая нейронный, а также могут использоваться для дообучения модулей БРМ.

Следует тщательно разграничивать TM, полученные из перевода, выполненного человеком или с участием человека, и TM, накопленные путем записи выдачи АГТ без проверки и редактирования, чтобы не допускать попадания в корпусы для обучения синтетических данных, произведенных не человеком, а автоматическими системами.

Переводческие компании и специалисты в области перевода не только применяют технологии TM в работе, но и оказывают услуги, относящиеся к ним. В частности, они могут выполнять:

- терминологическую работу;
- разработку процедур проверки качества;
- «чистку» и актуализацию баз перевода, в том числе для целей обучения МП/БРМ;
- формирование языковых корпусов путем сопоставления текста на исходном языке и варианта его перевода (элайн);
- разработку сценариев работы с применением баз ТМ (выбор, подключение, ранжирование, назначение приоритетов баз перевода на разных этапах).

## 4.2. Автоматическая генерация текста в переводе

К основным средствам автоматической генерации текста в переводе относятся:

- системы класса «машинный перевод» (МП);
- системы на основе больших речевых моделей (БРМ).

### 4.2.1. Машинный перевод

Системы класса «машинный перевод» (МП) — это компьютерные программы, с помощью которых может выполняться как полностью автоматический перевод с одного естественного языка на другой, так и автоматический перевод с участием человека (с последующей постобработкой).

Полностью автоматический перевод используется как для ознакомления с текстом, представленным на иностранном языке, так и для создания перевода для ограниченного применения.

При автоматическом переводе с участием человека результат (выдача) МП доводится до необходимого (чаще всего — сопоставимого с профессиональным) уровня качества с возможностью широкого применения и публикации.

При использовании выдачи МП в конкретной коммуникативной ситуации следует уделять особое внимание оценке следующих аспектов:

- прагматическая цель;
- целевая аудитория (ожидания и риски);
- общие риски применения автоматического перевода (как в полностью автоматическом режиме, так и после постобработки);
- отсутствие в режиме полностью автоматического перевода правовой охраны в качестве объекта авторского права, а также отсутствие исполнителя, принимающего ответственность за содержательную эквивалентность текста перевода исходному тексту;
- ожидания от применения МП (рост производительности при сохранении качества);
- достижимые уровни качества МП и уровни настройки (обучения) модулей МП;
- трансформация всего процесса перевода с точки зрения модификации этапов подготовки текста и этапов обеспечения качества.

Особенность технологии МП состоит в том, что для обеспечения приемлемого качества перевода двуязычный корпус для обучения системы МП (чаще всего в формате памяти переводов) должен:

- охватывать тематические области переводимых текстов;
- содержать переводы высокого качества;

- иметь объем, достаточный для конкретной технологии.

Кроме того, результат обучения зависит от применимости выбранной модели, а также от качества обучения. Качество выдачи МП может зависеть и от других факторов, в частности:

- от направления перевода (языковые пары);
- от формата представления исходного текста, типа файла;
- от вида (оформления) и качества исходного текста.

В противном случае генерируемые системой МП переводы не будут соответствовать уровню качества, пригодному для применения как в полностью автоматическом режиме, так и для постобработки.

По указанным причинам процесс внедрения МП должен обязательно включать в себя этап оценки качества выдачи до и после обучения модели, а также этап детального изучения возможностей применения в режимах полностью автоматического перевода и постобработки в связи с уровнями качества МП (так называемый «пилотный проект»).

Спектр услуг, оказываемых переводческими компаниями и специалистами по переводу в отношении профессионального применения МП и постобработки, может включать среди прочего:

- оценку качества выдачи МП;
- оценку общей применимости данной технологии в конкретной коммуникативной ситуации;
- управление рисками, в частности участие в идентификации и обработке рисков;
- разработку сценариев использования МП в полностью автоматическом режиме;
- разработку сценариев постобработки МП;
- настройку, обучение и сопровождение систем МП, в том числе терминологическую работу, работу по «чистке» и актуализации корпусов для обучения и т. п.

Необходимо уделять особое внимание обучению переводчиков применению МП, поскольку задачи, инструменты и приемы переводчиков-постредакторов могут значительно отличаться от традиционных переводческих и редакторских приемов, а основной особенностью выдачи современного нейронного МП является «гладкость» текста, создающая иллюзию качественного перевода и скрывающая значительные смысловые искажения. Пример программы обучения переводчиков-постредакторов приведен в Приложении Г к настоящему Стандарту.

#### **4.2.2. Большие речевые модели**

Большие речевые модели (БРМ), также называемые большими языковыми моделями (БЯМ), — это комплексные нейронные модели, использующие искусственные нейронные сети на основе трансформеров, обученные на больших объемах текстовых данных методом самостоятельного обучения, с модулем интерактивного диалога с пользователем для генерации ответа на запрос на естественном языке.

После обучения на больших объемах данных БРМ могут решать разнообразные задачи в области обработки естественных языков, такие как:

- генерация новых текстов;
- генерация переводов на другие языки;

- извлечение информации;
- обобщение информации;
- классификация текстов;
- внесение изменений в тексты и их адаптация для решения конкретных задач, в том числе для целей автоматической постобработки, и т. п.

В отличие от систем МП, при обучении используется не двуязычный, а одноязычный или многоязычный корпус текстов, и различные языки представлены в нем неравномерно. Кроме того, для обучения требуются чрезвычайно большие объемы корпусов, что неизбежно приводит к необходимости использования так называемых «синтетических» данных (наборов данных, как созданных человеком, так и сгенерированных автоматическими системами). Глубоко присущим, неустранимым свойством БРМ являются т. н. «галлюцинации», т. е. не соответствующие исходному тексту, инструкциям и действительности ответы, сгенерированные автоматически на основании текстов в корпусе обучения и настроек моделей.

Так же, как и у специализированных модулей нейронного машинного перевода, обученных на двуязычных корпусах, качество выдачи БРМ зависит от множества факторов, среди которых тип и устройство модели, языковая пара, качество и объем корпуса для обучения, соответствие предметной области и целевой аудитории, вид и качество исходного материала, технология применения и т. п. К перечисленным выше факторам, влияющим на качество результата, добавляется степень умения пользователей создавать точные и непротиворечивые запросы на естественном языке («промпты») для БРМ.

В связи с этим процесс внедрения средств на основе БРМ должен включать обязательный этап оценки качества выдачи и качества последующей обработки результатов с верификацией специалистами по переводу и непрерывным контролем качества («пилотный проект»).

Спектр услуг, оказываемых переводческими компаниями и специалистами по переводу, в отношении профессионального применения средств БРМ и постобработки может включать, среди прочего:

- оценку качества выдачи БРМ;
- оценку общей применимости данной технологии в конкретной коммуникативной ситуации;
- участие в управлении рисками, в частности в идентификации и обработке рисков;
- постановку задач;
- терминологическую работу;
- разработку сценариев использования в полностью автоматическом режиме;
- разработку сценариев постобработки;
- осуществление постобработки;
- написание промптов (запросов) для оптимизации процесса, в том числе по созданию многоязычного контента;
- оценку качества полученного результата.

## **5. АГТ в переводческом процессе**

### **5.1. Профессиональный переводческий процесс**

Профессиональный перевод представляет собой процесс создания творческого результата интеллектуальной деятельности автора (группы авторов) в форме связного производного текста и на основании связного исходного текста, осуществляемый в условиях свободы выбора технических средств, стратегии, тактики и приемов осуществления их деятельности, поэтому главным отличием процесса профессионального перевода от процесса АГТ является наличие автора или группы авторов, берущих на себя профессиональную ответственность за качество результата перевода.

Современный профессиональный перевод является сложным технологическим процессом, который осуществляется специалистами в области перевода, лингвистики, межкультурной коммуникации, информатики, с применением средств автоматизации и управления качеством, процессами и рисками; он представляет собой комплекс ручного труда с применением автоматизированных систем (CAT-инструментов на основе ТМ) в сочетании с автоматической генерацией текста в переводе (МП и БРМ).

Полностью ручной перевод без применения АГТ может быть целесообразен в литературном, художественном, маркетинговом переводе, когда особенно важна творческая свобода переводчика, результатом работы которого будет оригинальный текст, обеспечивающий воздействие на получателя перевода, задуманное автором исходного текста.

### **5.2. Место АГТ в переводческом процессе**

Автоматизированная и автоматическая генерация текста применяется в целях оптимизации переводческого процесса в соответствии с задачами и ресурсами заказчика. Использование АГТ требует добавления (включения) в традиционный переводческий процесс специфических этапов.

В большинстве случаев любое использование средств АГТ должно сопровождаться оценкой уровня качества выдачи систем АГТ, предобработкой исходных материалов и постобработкой результата их выдачи переводчиками-постредакторами с учетом ТЗ, стандартов и инструкций. При этом обязательным условием является соблюдение правил выполнения всех этапов технологического процесса профессионального перевода, в том числе полноценная проверка качества.

В исключительных случаях, например в условиях недостаточности ресурсов и времени для выполнения всех этапов технологического процесса, постобработка выдачи АГТ может исключаться из процесса, но при обязательной должной оценке возможности возникновения рисков и их последствий в конкретной коммуникативной ситуации.

Выбор и комбинирование средств АГТ, а также видов предобработки и постобработки могут быть согласованы между заказчиком и исполнителем, а в отсутствие таких договоренностей осуществляются на усмотрение исполнителя.

**Таблица 1. Этапы работы с текстом при применении АГТ в переводческом процессе и их возможное содержание**

<b>Предобработка (возможно комбинирование)</b>	<b>Генерация</b>	<b>Постобработка (возможно комбинирование)</b>
<p>Разверстка, расшифровка, преобразование исходного текста в редактируемый цифровой формат, устранение избыточной фрагментации текста и т. п., добавление служебной разметки.</p> <p>Проверка правильности автоматически распознанных текста и речи.</p> <p>Вычленение не подлежащих автоматизированному и автоматическому переводу фрагментов, например индикации, подписей к физическим переключателям и т. п.</p> <p>Скрытие текста, не подлежащего переводу.</p> <p>Переписывание исходного текста и подготовка дополнительных материалов, написание «промпов».</p> <p>Создание исходного текста на упрощенном языке.</p> <p>Предредактирование для автоматического перевода на язык или группу языков, имеющих культурную общность (с учетом локалей).</p> <p>В случае перевода устной речи — специальная подготовка устных выступлений докладчиков (контролируемая устная речь). Приведение подлежащих переводу демонстрационных материалов в соответствие с устной речью</p>	<p>Автоматически выбранный сценарий АГТ, например выбор и подключение МП без участия человека.</p> <p>Выбранный экспертами исполнителя сценарий перевода из числа готовых решений с возможными этапами тестирования и выполнения экспериментов.</p> <p>Разделение исходного контента на тематические блоки с индивидуальным подбором модулей, глоссариев и подобных ресурсов.</p> <p>То же, но с индивидуальным обучением на контенте заказчика или на корпусах тематической области, с добавлением глоссариев (терминологических баз), памяти переводов, с учетом контекста.</p> <p>Формирование ранжированных по значимости глоссариев (терминологических баз) для настройки учета контекста</p>	<p>Отсутствие этапа ручной постобработки или же полностью автоматическая постобработка (как правило, в условиях жесткой экономии времени и ресурсов).</p> <p>Техническая постобработка, то есть минимально необходимое редактирование по формальным критериям.</p> <p>Выборочная постобработка на основании набора лингвистических и иных требований, критериев и уровней ответственности, содержание которых определяется сторонами (заказчиком и исполнителем) по согласованию.</p> <p>Полная постобработка, направленная на получение перевода, отвечающего всем требованиям к качеству профессионального перевода</p>

## 6. Подготовка к АГТ (предобработка)

Технологии автоматического перевода (МП и БРМ) в настоящее время не способны обеспечивать безошибочный, а также единообразный по смыслу и качеству результат (выдачу), не обладают способностями к осмыслению и логике, а лишь имитируют их на основе имеющегося корпуса обучения со всеми присущими ему ограничениями (по времени, актуальности, наличию ошибок и т. п.).

В связи с этим указанные технологии сами по себе не дают гарантии полного и точного перевода, отсутствия в нем ошибок и неточностей (в т. ч. перенесенных из исходного текста), а также соответствия требованиям по качеству, стандартам и ТЗ заказчика. Следствием этого является невозможность возложить на исполнителя юридические гарантии в отношении качества работы, выполненной без участия человека.

При применении технологий МП и БРМ возникает необходимость предварительной подготовки материала к автоматическому и автоматизированному переводу, в том числе выполнение следующих операций:

- распознавание изображений;
- разметка и изменение форматирования текста с последующим восстановлением;
- анонимизация конфиденциальной информации;
- подготовка глоссариев для обеспечения единообразия генерации или перевода;
- приведение к единообразию терминологии и обозначений в исходном тексте (для обеспечения отсутствия ошибок в переводе и единообразия перевода);
- исправление ошибок в исходном тексте с целью более качественной обработки материала автоматическими средствами;
- настройка промптов БРМ и прочих средств, позволяющих добиться максимально высокого качества машинной выдачи;
- модификация определенных этапов процесса перевода для максимально эффективного применения технологий;
- оценка качества машинной выдачи после проведения всех перечисленных выше подготовительных и производственных работ.

Оценка возможных технических требований к качеству и предоставление юридических гарантий соответствия результата ТЗ может быть выполнено только с участием человека.

При применении средств АГТ исполнитель самостоятельно или по согласованию с заказчиком может выполнять:

- оценку уровня настройки средства АГТ и качества выдачи АГТ;
- оценку применимости АГТ и возможности выполнения предобработки или постобработки данного типа текстов;
- подбор инструментов для повышения качества АГТ;
- подготовку и очистку данных для обучения АГТ;
- тестирование разных моделей для выбора и подключения наиболее соответствующей задачам перевода (на основе представлений о применимости и критериях оценки качества);
- оценку рисков применения АГТ, в том числе с учетом этических аспектов и требований безопасности;

- предредактирование текстов для обработки с помощью средств автоматической генерации.

Этапы формирования проекта с применением средств АГТ:

- подбор подходящих баз переводов (TM), а также глоссариев для подключения к инструментам перевода при выполнении перевода или постобработки в CAT-инструменте;
- при отсутствии базы TM, но при наличии ранее выполненных переводов для данного типа документов — создание базы переводов с помощью операции сопоставления для создания базы переводов (элайн);
- составление или обновление глоссария;
- при отсутствии инструкций по переводу и постобработке со стороны заказчика — составление их исполнителями, причем таким образом, чтобы они позволяли добиться желаемого результата; по возможности такая инструкция должна быть согласована с заказчиком в виде ТЗ;
- сравнение с имеющейся в распоряжении исполнителя базой TM для получения статистики совпадений с ранее выполненными качественными переводами и предоставления расчета заказчику;
- если выбран процесс с автоматической генерацией текста, подключение модуля МП или БРМ для получения перевода для сегментов, для которых в базе TM не найдено совпадение (сегменты No Match);
- составление списка типовых ошибок АГТ в данной сырой выдаче и подготовка инструкции по постобработке, включающей основные требования, сценарии работы, стратегии, справочные материалы и ресурсы.

## 7. Постобработка результатов (выдачи) АГТ

После подготовки исходных текстов, выбора средств АГТ и сценариев работы с ними, а также предварительной оценки качества выдачи запускается производственный цикл, включающий постобработку выдачи АГТ с целью обеспечения уровня качества, соответствующего стандартам и ТЗ.

Постобработка предполагает работу с двумя источниками: исходным текстом и результатом генерации, в который вносятся правки.

При выполнении постобработки в CAT-инструменте на основе TM результат автоматической генерации может быть как предварительно внесен в текст перевода («предперевод»), так и предложен системой в интерактивном режиме. Переводчик-постредактор может использовать выдачу в режиме подсказки или подстановки либо совсем отказаться от нее. Постобработка преимущественно заключается в проверке выдачи АГТ, верификации перевода и исправлении ошибок при необходимости.

Дополнительно переводчик-постредактор может составлять списки типовых ошибок АГТ с целью выработки сценариев их ручной или автоматической обработки, а также для возможности дополнительного обучения систем.

Постобработка может быть полностью ручной или частично автоматизированной.

## 7.1. Этапы постобработки

Производственный процесс постобработки состоит из нескольких этапов, количество которых зависит от качества исходного текста, применяемых технологий и технических требований заказчика.

Варианты реализации процесса постобработки можно разделить на две большие группы: проверка и исправление предварительно сгенерированного текста (претрансляции или предперевода) и внесение правок по интерактивным рекомендациям и подсказкам.

### Основные этапы технологического процесса перевода с применением АГТ

#### а) Выполнение постобработки АГТ

Выполнение постобработки АГТ является основной частью технического процесса перевода и заключается в постобработке полученных из базы ТМ совпадений и вариантов, сгенерированных МП или БРМ, с одновременным сохранением результата в базу ТМ и обновлением глоссария.

По завершении этапа постобработки переводчик-постредактор обязан проверить переведенный документ на соответствие ТЗ, инструкциям и соблюдение проектных правил, далее могут быть задействованы дополнительные редакторы и корректоры.

При любых изменениях текста перевода требуется актуализация баз ТМ.

#### б) Выполнение редактуры и (или) корректуры

После постобработки текст перевода поступает на этап редактуры и (или) корректуры, в ходе которых текст проверяется на соответствие ТЗ заказчика и сверяется с исходным текстом. Редактор при необходимости приводит текст перевода в соответствие с ТЗ заказчика, может помечать смысловые, логические и прочие ошибки, в своей работе редактор может использовать средства АГТ. Корректор при необходимости приводит текст перевода в соответствие с ТЗ заказчика и нормами языка перевода, может помечать выявленные ошибки, в том числе с применением средств АГТ.

В соответствии с нормами авторского права о неделимом соавторстве смысловая правка должна быть согласована с переводчиком-постредактором, выполнявшим постредактирование АГТ. Способ и форма согласования устанавливаются соглашением сторон.

В зависимости от количества и серьезности выявленных ошибок по завершении редактуры текст перевода может поступить на этап контроля качества или вернуться на один или несколько предыдущих этапов.

После выполнения всех этапов текст перевода обрабатывается верстальщиком по ТЗ заказчика и по действующим нормативным документам.

#### в) Контроль качества

Контроль качества должен являться неотъемлемой частью каждого этапа и вида работ. Конкретная реализация системы качества определяется исполнителем.

Все исправления и добавления должны быть отражены в базах ТМ и глоссариях.

При наличии договоренности заказчику дополнительно предоставляется отчет об обнаруженных проблемах АГТ для дополнительной настройки систем АГТ и улучшения результатов их работы.

В зависимости от выбранного в соответствии с ТЗ технологического процесса последовательность этапов и операций может отличаться у разных исполнителей.

Следует учитывать, что отказ от перечисленных выше этапов может повлечь за собой снижение качества и, соответственно, риски наступления негативных последствий.

## 7.2. Глубина постобработки

Выбор уровня качества при постобработке выдачи АГТ зависит от целей перевода и требований заказчика, в связи с чем выделяются три уровня постобработки: **техническая, выборочная и полная**. Уровни постобработки различаются по степени внимания к тексту на исходном языке, а также по степени вмешательства в выдачу АГТ.

### 7.2.1. Техническая постобработка

Техническая постобработка — это выполнение постобработки с целью получения понятного текста в условиях ограничения времени и ресурсов в том случае, если ТЗ заказчика отсутствует или не содержит требований к высокому качеству перевода.

При согласованной с заказчиком технической постобработке АГТ у исполнителя не возникает авторского права и исполнитель не несет ответственности за полную и точную передачу смысла исходного текста в переводе.

Техническая постобработка подразумевает:

- проверку текста перевода только на языке перевода; проверку наличия перевода для всех единиц текста;
- проверку грамматической и пунктуационной правильности выдачи АГТ;
- проверку и восстановление чисел;
- проверку и восстановление служебной разметки (форматирования);
- проверку на наличие в тексте признаков добавления или упущения существенной информации для передачи в таком случае текста на выборочную или полную постобработку;
- проверку на наличие признаков содержания в тексте перевода неточной или неприемлемой согласно ТЗ информации для передачи в таком случае текста на выборочную или полную постобработку.

Техническая постобработка не подразумевает:

- сверку выдачи АГТ с текстом на исходном языке;
- единство терминологии;
- перифразирование.

При таком уровне постобработки рекомендуется:

- использовать выдачу АГТ в максимально возможном объеме;
- обращать внимание на то, нет ли в тексте перевода признаков добавления или упущения существенной информации;
- проверять, не содержится ли в тексте перевода неточная или неприемлемая информация.

Техническая постобработка не предполагает изменения и переписывания текста перевода в целях совершенствования стиля.

### 7.2.2. Выборочная постобработка

Выборочная постобработка — это выполнение технической постобработки в сочетании с проверками по набору индивидуальных требований или критериев заказчика, с учетом возникновения известных рисков применения полученных текстов, которые установлены в ТЗ или в соглашении (договоре) заказчика и исполнителя: например, проверка заголовков и оглавлений, имен собственных, названий, наименований изделий, подписей, элементов индикации, дат, цифр, определенных терминов и т. п.

Выборочная постобработка подразумевает:

- все проверки, входящие в техническую постобработку;
- проверки по четко обозначенным критериям;
- проверку терминологии по конкретному ограниченному списку.

Выборочная постобработка не подразумевает:

- проверку и приведение к единству терминологии по всему тексту;
- сверку выдачи АГТ с текстом на исходном языке;
- перифразирование.

При составлении списка проверочных критериев рекомендуется использовать риск-ориентированный подход в соответствии с Приложением Е к настоящему Стандарту.

Следует учитывать, что не существует полностью достоверного способа оценить последствия, в том числе финансовые потери, даже от мелкой фактической ошибки. В связи с этим, если необходимо гарантировать полное отсутствие ошибок и соответствие результата работ ТЗ, следует осуществлять оценку требуемой глубины постобработки и полную постобработку.

### 7.2.3. Полная постобработка

Полная постобработка — это процесс, результатом которого является текст такого же качества, как и при профессиональном переводе (см. п. 5.1): он должен быть точным, понятным, стилистически адекватным, синтаксически, грамматически, семантически и пунктуационно корректным.

Полная постобработка включает в себя проверки, описанные для технической и выборочной постобработок; кроме того, должны быть соблюдены следующие основные условия:

- никакая информация не добавлена и не упущена,
- в тексте перевода не содержится неточная и неприемлемая информация,
- соблюдены единство и корректность терминологии.

## 8. Процесс перевода и постобработки АГТ без взаимодействия с исполнителем на подготовительном этапе

В случае если подбор средств АГТ осуществляет заказчик, проведение связанных с выбором АГТ мероприятий (включая настройку, оценку рисков, оценку выдачи и т. п.) становится зоной ответственности заказчика.

В этом случае для обеспечения качества, сравнимого с выполнением предобработки с участием исполнителя, должен быть введен промежуточный этап — этап

согласования, цель которого заключается в оценке выдачи АГТ исполнителем, выработки им сценариев ручной или автоматической обработки, а также формировании рекомендаций для возможного дополнительного обучения системы АГТ.

Альтернативной возможностью является модель расчета оплаты, при которой оплата пропорциональна фактически затраченным усилиям исполнителя.

Для оценки качества выдачи АГТ на этапе согласования исполнителю должна быть предоставлена следующая информация:

- сведения об используемой системе АГТ и подробности ее подготовки:
  - использование ТМ заказчика при обучении и подготовке системы АГТ;
  - использование глоссариев заказчика при подготовке и обучении системы АГТ и непосредственно в ходе АГТ;
  - используемые средства оценки качества выдачи АГТ и итоговые отчеты;
- примеры выдачи АГТ;
- информация о возможности и намерении дообучения системы АГТ на основе постобработанных текстов.

Исполнитель со своей стороны должен произвести оценку качества выдачи АГТ.

Роль инструкции по переводу и постобработке (ТЗ) в ситуации, когда непосредственный исполнитель перевода не участвует в подготовке систем АГТ, возрастает вместе с необходимостью участия в ее составлении обеих сторон процесса (заказчика и исполнителя). Результатами этапа согласования могут стать рекомендации заказчику, необходимые для дальнейшего повышения качества АГТ, корректировки производственного процесса на стороне исполнителя и заказчика, корректировки условий выполнения заказа (время, цена, состав и результат работ, формат представления текстов для постобработки и т. п.).

При невозможности заблаговременной оценки качества выдачи АГТ и привлечения непосредственного исполнителя перевода к подготовке и обучению систем АГТ определение трудозатрат, стоимости и сроков для непосредственного исполнителя работ по постобработке предоставленного заказчиком результата АГТ может осуществляться по итогам выполненных работ с применением технологии оценки глубины редактирования.

Технология оценки глубины редактирования может также применяться и в том случае, если непосредственный исполнитель перевода обязан использовать предоставленную заказчиком память переводов не предсказуемого для исполнителя и не контролируемого исполнителем качества.

Однако при постобработке предоставленной заказчиком выдачи АГТ на основе МП и БРМ рекомендуется использовать в качестве пересчетной схемы для оценки глубины редактирования такие схемы учета, которые предусматривают более высокую оплату и больший бюджет времени для той же глубины редактирования, чем в САТ-инструментах на основе ТМ для тех же значений совпадения с ТМ.

В случае «загрязнения» предоставленной заказчиком ТМ результатами систем АГТ на основе МП и БРМ следует использовать те же схемы оценки глубины редактирования, что и для выполнения перевода с применением АГТ на основе МП и БРМ.

## 9. Оценка рисков. Технические требования заказчика

При применении современных средств автоматизированного и автоматического перевода, в особенности построенных на нейронных сетях, возможно возникновение ряда серьезных проблем, связанных с их архитектурой и принципом работы, в результате чего в автоматически сгенерированном тексте и переводе могут содержаться ошибки разного рода.

Ненадлежащее качество перевода чаще всего создает риски возникновения убытков у заказчиков по следующим причинам:

- задержка при выводе на рынок продуктов или услуг;
- необходимость выполнения дополнительного редактирования или даже нового перевода с незапланированными затратами дополнительного времени и средств на исправление ошибок;
- штрафы от регуляторов;
- потеря доверия пользователей, заказчиков и партнеров;
- неверные управленческие и финансовые решения, принятые на основе некорректно переведенной информации.

Эти риски могут возникать тогда, когда задачи оперативного вывода продукта или услуги на рынок и снижения стоимости и сокращения времени работы над текстами и переводами имеют более высокий приоритет, чем соблюдение всех этапов профессионального процесса, гарантирующего отсутствие неверных утверждений, недостоверных фактов или ошибок и некорректных средств взаимодействия с целевой аудиторией в текстах и переводах путем задействования отлаженных процедур с учетом обязательной проверки качества.

При построении процесса перевода текстов, в том числе с использованием средств АГТ, исполнитель обязан руководствоваться требованиями заказчика, по возможности изложенными в форме технического задания (ТЗ). Именно требования заказчика определяют уровень качества конечного материала.

ТЗ должно разрабатываться совместно заказчиком и исполнителем, учитывать бизнес-требования со стороны заказчика, а также условия оказания услуг исполнителем (стоимость и время проведения работ при приемлемом уровне стоимости, сроков и рисков), в результате которых будет обеспечено требуемое качество.

## Приложение А. Основные требования к качеству перевода

### Градации качества перевода

Качество перевода как результата оказания услуг определяется требованиями заказчика, изложенными в техническом задании (ТЗ).

При составлении ТЗ следует сформулировать ответы на следующие вопросы:

- Какова аудитория конечного материала?
- Нужно ли при переводе использовать весь понятийный, языковой и словарный аппарат языка перевода или можно обойтись более простыми средствами?
- Нужна ли адаптация материала к особенностям технической, медицинской, юридической и прочей практики, сложившейся в стране целевой аудитории перевода?
- Достаточно ли буквального перевода?

Эти вопросы соответствуют следующим условным качественным категориям перевода:

- премиальный перевод;
- буквальный перевод.

Каждая из этих категорий требует своего уровня трудозатрат и квалификации исполнителей.

В процессе перевода следует прикладывать максимально возможные усилия, для того чтобы премиальный перевод:

- точно и полностью передавал мысль автора исходного текста;
- соответствовал всем нормам и правилам языка перевода и требованиям заказчика по содержанию, стилю, оформлению;
- не содержал грамматических, орфографических, пунктуационных и прочих ошибок;
- был терминологически последовательным и единообразным, соответствовал отраслевой принадлежности исходного текста;
- соответствовал применимым для данной предметной области справочным материалам (отраслевой терминологии, проектным указаниям, базам памяти переводов, глоссариям, терминологическим базам и т. п.);
- соответствовал региональным стандартам (использование принятых в стране языка перевода символов, условных обозначений, наименований, а также адаптированных к правилам языка перевода форматов дат и времени, единиц измерения, числовых выражений, денежных единиц и пр.);
- по форматированию был идентичен исходному документу;
- являлся пригодным для целей его использования, соответствовал целевой аудитории и назначению документа.

Для проведения такой работы необходимо детальное ТЗ. Следует учитывать, что достижение премиального качества требует значительных трудозатрат. Если применяется АГТ, требуется полная постобработка.

Чаще всего при согласовании ТЗ между заказчиками и исполнителями по умолчанию подразумевается премиальное качество, однако в настоящее время все большее

распространение находят попытки организовать такой производственный процесс, при котором обеспечивался бы определенный заданный уровень достаточно хорошего перевода путем применения средств АГТ, в т.ч. с технической или выборочной постобработкой.

Достаточно хороший перевод — это гарантированный минимум качества текста после технической или выборочной постобработки.

Достаточно хороший перевод:

- максимально точно сохраняет синтаксическую структуру и порядок слов и словоформ исходного языка;
- соответствует требованиям заказчика по содержанию, стилю, оформлению;
- не содержит грамматических, орфографических, пунктуационных и прочих ошибок;
- не содержит неприемлемой лексики и оскорбительных высказываний, если они отсутствовали в исходном тексте;
- содержит единообразную терминологию, соответствующую отраслевой принадлежности исходного текста;
- соответствует локали (символы, условные обозначения, наименования и т. п.)

Выдача современных систем АГТ варьируется по качеству в непредсказуемо широких пределах.

### Методы оценки качества перевода

В настоящее время при разработке средств автоматического перевода используются автоматические метрики, позволяющие получить количественную оценку «качества» строк выдачи без участия человека.

Следует учитывать, что автоматические метрики (BLEU, hLEPOR, Rouge, Meteor и пр.) служат для анализа последовательностей символов без учета цели коммуникации и измеряют не качество перевода, а схожесть выдачи МП и БРМ с некоторым эталонным переводом, образцом. По этой причине метрики могут обнаруживать большую разницу между разными верными вариантами перевода, выражающими одну и ту же мысль.

Существующие способы оценки качества перевода приведены в таблице ниже.

**Таблица 2. Виды метрик измерения качества (\* — автоматические, \*\* — экспертные)**

Методика	Названия	Объект оценки
Сравнение близости с эталоном*	BLEU, Rouge, hLEPOR, NIST, METEOR, TER, BERTScore	Одно предложение
Предсказание специально тренированной нейромодели*	TransQuest, COMET	Одно предложение
Экспертная оценка по балльной шкале**	DA (Direct Assessment)	Одно предложение
	Целостная (Holistic)	Фрагмент текста
Аналитические методы**	ATA Certification Metric, SAE J2450, MQM	Фрагмент текста

Оценка качества перевода может быть **целостной** или **аналитической**. Кроме того, оценка качества перевода может быть **автоматической** или **экспертной**.

**Целостная балльная оценка качества перевода фрагмента текста** делается экспертом по всему объему фрагмента (часто по всему материалу) — с учетом контекста не только самого текста, но и целей применения, требований аудитории и т. п. Такая оценка выставляется по балльной шкале по общему впечатлению, без детального анализа ошибок. Поэтому она обладает большой вариативностью, зависит от качества работы проверяющего и может быть неточной или субъективной, если делается непрофессионально.

Самый точный способ оценки качества перевода — это **аналитическая экспертная оценка**, при которой специалист осуществляет тщательный анализ текста перевода, выявляет ошибки, классифицирует их по категориям и определяет степень их критичности. Ошибкам присваиваются «штрафные очки», по которым подсчитывается общий балл качества перевода оцениваемого фрагмента текста, т. е. формируется количественная оценка качества, которая сравнивается с установленным допустимым количеством ошибок на эталонном объеме текста.

При аналитической оценке количество штрафных очков нормируется на количество слов в стандартном объеме текста, затем по формуле оценки качества вычисляется балл качества, который сравнивается с пороговым допустимым баллом качества в стандартном объеме текста. Пороговое значение определяется требованиями заказчика (ТЗ). Формула вычисления оценки качества может включать в себя калибровку, облегчающую получение человекочитаемой оценки качества и возможность сравнения оценок, полученных для совершенно разных условий.

Наиболее полной и точной аналитической метрикой в настоящее время является метрика MQM (Multidimensional Quality Metrics). Этот подход существует в двух видах: упрощенном и полном.

### **Упрощенная аналитическая метрика качества перевода**

При упрощенном подходе оценка вычисляется по фиксированной «сырой» формуле, без калибровки, часто с подгонкой весов ошибок.

Упрощенный подход широко распространен; чаще всего используются два эмпирически найденных «уровня качества»:

- «премиальный»: от 2 до 10 (чаще — 5) штрафных баллов на 1000 слов;
- «удовлетворительный»: 20 штрафных баллов на 1000 слов.

Попытки уменьшить допустимый порог ошибок до менее чем 2 на 1000 слов не реалистичны и не продуктивны из-за возможной предвзятости проверяющих.

Обычно 1 штрафной балл соответствует одной мелкой ошибке (пунктуация, опечатка и т. п.), а более значительные ошибки получают большее количество штрафных баллов (от 3 до 5 или 10).

Данную методологию оценки качества целесообразно применять также и в отношении результатов АГТ, однако в этом случае допустимые пороги качества могут быть иными.

Более подробно методология аналитической оценки качества с типологией ошибок и формулами формирования балльной оценки качества с калибровкой и без таковой приведена в Приложении 2 к настоящему Стандарту.

## Приложение Б. Методология аналитической оценки качества перевода с типологией ошибок и формулой формирования балльной оценки качества с калибровкой на основе метрики Multidimensional Quality Metrics (MQM)

Чтобы узнать, соответствует ли качество перевода установленным требованиям, наиболее точным подходом является подсчет количества ошибок по установленной метрике. Все прочие способы намного менее точны и оставляют широкий простор для несогласия сторон.

Наиболее тщательно проработанной и широко применяемой является методика Multidimensional Quality Metrics (MQM). MQM — это набор инструментов для аналитической оценки качества перевода (Translation Quality Estimation, TQE). Этот подход можно использовать для оценки ручного перевода, машинного перевода, а также перевода, сгенерированного ИИ.

Аналитическая модель MQM основана на детальной оценке образцов переведенного текста с использованием типологии ошибок и формулы оценки MQM. Ошибки классифицируются по типам и серьезности, в соответствии с выбранной метрикой присваиваются штрафные очки, которые с помощью формулы оценки преобразуются в итоговый балл оценки качества, отражающий степень соответствия перевода конкретным спецификациям заказчика (ТЗ). Именно ТЗ определяют требуемый заказчиком уровень качества, так как при его составлении были выявлены и количественно оценены риски (юридические, финансовые, репутационные и пр.), связанные с некачественным переводом.

MQM включает набор типов ошибок и модель оценки. Типы ошибок организованы в иерархическую систему, называемую **типологией ошибок MQM**, по 7 основным «измерениям» типов ошибок старшего уровня, которые делятся на несколько подтипов ошибок второго и третьего уровней. Метрика MQM содержит нужное подмножество категорий ошибок, систему весов и параметров, назначенных типам ошибок и уровням серьезности, а также формулу оценки, используемую для расчета числового балла, который представляет качество оцениваемого фрагмента перевода в соответствии с согласованными спецификациями.

Достоверная оценка качества может быть произведена на фрагменте текста не менее 2500 слов (около десяти страниц), поскольку при уменьшении размера фрагмента экспоненциально возрастает погрешность измерения, а кроме того, многие ошибки проявляются только в контексте документа. И хотя для грубой выборочной проверки минимальный размер фрагмента текста может составлять как 500, так и даже 250 слов, обязательно следует учитывать, что делать обобщающие выводы по измерению качества на основании столь небольшого фрагмента нельзя. Размер фрагмента для оценки варьируется в диапазоне от 500 до 20 000 слов (в зависимости от размера проекта и ресурсов, доступных для оценки).

Для оценки могут быть разработаны собственные модели и формулы оценки на основе выбора тех типов ошибок и уровней их критичности, которые важны с учетом спецификаций проекта. Эти значения используются в формуле оценки для расчета показателя качества. Расчет баллов определяет окончательный показатель качества. Балл может быть рассчитан с калибровкой или без нее.

Модель **сырой оценки (оценка без калибровки)** использует следующие значения и параметры:

1. Объем проверяемого текста в словах (EWC).

2. Абсолютная сумма штрафных очков (APT) — общая сумма штрафных очков, рассчитанных для количества слов оценки. Штрафные очки рассчитываются на основе количества ошибок, указанных для отдельных типов ошибок по критичности ошибок.

Формула балльной оценки качества перевода в модели сырой оценки (без калибровки) приведена ниже:

$$Score = 1 - \frac{APT}{EWC}$$

Модель **оценки с калибровкой** использует те же значения и параметры и, кроме того, штрафные баллы, рассчитанные для оцениваемого фрагмента текста, по предварительно выбранному порогу прохождения или пределу допуска, на балльной шкале от 0 до 100. Метод калибровки существенно упрощает установку нужного порогового значения балла качества и допустимого предела ошибки, и балльная оценка становится намного проще и гибче, а процесс принятия решения о прохождении или непрохождении проверки качества — более ясным и понятным. Оценки текста разного типа становятся человекочитаемыми и сравнимыми по единой балльной шкале. Калибровка позволяет выразить значения балльной оценки таким образом, чтобы заинтересованные стороны могли легко интерпретировать их в соответствии со своими ожиданиями и спецификациями.

Формула балльной оценки качества перевода в модели с калибровкой приведена ниже:

$$Score = 100 - \frac{APT * 1000}{EWC} * \frac{100 - PFT}{c},$$

где **APT** и **EWC** определены выше, **c** — допустимое количество ошибок на эталонном объеме текста в 1000 слов, **PFT** — условный порог допустимого балла качества на балльной шкале от 0 до 100.

### Процесс оценки MQM

1. Проанализировать ТЗ проекта перевода с учетом потребностей конечного пользователя.
2. Выбрать типы ошибок из типологии ошибок MQM, которые соответствуют определенным спецификациям.
3. Выбрать формулу оценки (с калибровкой или без).
4. Создать собственную метрику MQM.
5. Определить размер выборки и процедуру отбора образцов.
6. Применить метрику: аннотировать ошибки, вычислить итоговый балл качества.

## Приложение В. Пример требований к профессиональной квалификации переводчиков-постредакторов

### 1. Основные требования, образование и опыт

- Высшее образование или опыт работы не менее 5 лет.
- Переводческие навыки.
- Владение на профессиональном уровне языками, используемыми для перевода.
- Знание требований к переводу, редактированию и постобработке.
- Исследовательские навыки.
- Культурная компетенция.

### 2. Специализация

- Компетенция в предметной области.
- Владение специализированной терминологией на языках, используемых для перевода.

### 3. Технические навыки

- Знание основ технологий CAT на основе TM, МП, БРМ.
- Представление о типовых ошибках и проблемах перевода и постобработки.
- Представление о преимуществах, недостатках, ограничениях переводческих технологий.
- Владение переводческими инструментами на уровне уверенного пользователя.

### 4. Опыт и квалификация

- Опыт постобработки (постредактирования) не менее 1 года, желательно подтвержденный положительным результатом тестирования по метрике, применимой для выполнения проекта.
- Тестирование при отсутствии релевантного опыта.

## Приложение Г. Пример требований к программам обучения переводчиков-постредакторов

Основная цель обучения — снабдить специалистов по постобработке результатов (выдачи) АГТ необходимыми навыками, помимо навыков перевода и редактирования переводов, а также улучшить восприятие и понимание технологий АГТ.

Специалистов по постобработке желательно обучать также предобработке текстов для АГТ, обучению и настройке систем АГТ и комплексным технологическим процессам, включающим в себя не только постобработку, но и настройку систем АГТ и предобработку текстов.

Основные аспекты обучения: технологии, средства, умение следовать инструкциям, стратегии постобработки, типология ошибок, терминологическая работа, навыки обработки текста, инструменты проверки качества, обратная связь.

Уровень обучения может быть базовым (общее представление о деятельности оператора АГТ) и проектным (выполнение постобработки в конкретных условиях).

Примерный набор тем приведен ниже.

1. Как работают системы АГТ, обучение и переобучение.
2. Виды и области применения АГТ в устном и письменном переводе, реферировании, субтитровании, копирайтинге и т. п.
3. Преимущества и ограничения, применимость для разных сценариев.
4. Редактирование сегментов (подсказок) из ТМ, постобработка МП и БРМ — в чем разница.
5. Постобработка, отличие от редактирования результатов ручного перевода.
6. Уровни постобработки. Недостаточная постобработка и излишняя постобработка.
7. Типы контента, подходящего для постобработки.
8. Типовые ошибки и способы их исправления или недопущения.
9. Оценка качества.
10. Предредактирование текстов.
11. Автоматизированная постобработка.
12. Создание промптов (инструкций) для систем на базе ИИ.
13. Практическое применение.

## Приложение Д. Пример профиля должности переводчиков-постредакторов

### Обязанности

- Постобработка МП до получения текста с качеством, соответствующим переводу, выполненному квалифицированным переводчиком.
- Работа в CAT-инструментах.
- Изучение универсальных и специально обученных систем МП и АГТ.
- Подбор проектов (текстов), пригодных для машинного перевода.
- Подготовка текстов для МП (хорошее распознавание, возможно предредактирование — исправление опечаток, корректировка разбивки на сегменты и т. п.).
- Участие в сравнительном тестировании (оценка с использованием метрики hLEPOR и др.) и обучении в МП.
- Прочая работа лингвиста (создание и «чистка» ТМ, в том числе для МП, работа с глоссариями).
- Оценка работы внешних переводчиков-постредакторов, предоставление обратной связи и заключений по качеству.
- Расчет стоимости и скидок для проектов с применением МП и АГТ (на основе метрики hLEPOR — 0–30 % в зависимости от объема внесенных правок).
- Мониторинг новейших разработок в области МП и АГТ.

### Требования

- Высшее образование в соответствующей предметной области проектов.
- Грамотность, владение русским языком, исходным языком и языком перевода.
- Опыт выполнения переводов или редактирования переводов по меньшей мере одного из следующих типов: юриспруденция, менеджмент, финансы, медицина, техника, маркетинг (не менее 3 лет).
- Успешное выполнение тестового задания.
- Внимательность, добросовестность и аккуратность.
- Опыт работы в CAT-инструментах, понимание особенностей постобработки (постредактирования, редактирования машинного перевода и АГТ и т. п.).

## Приложение Е. Факторы риска, влияющие на применимость средств АГТ

При наличии нескольких из перечисленных ниже факторов рекомендуется увеличивать долю ручной предобработки или постобработки текстов, полученных с помощью средств АГТ.

Факторы риска	Рекомендации
Отсутствие накопленных языковых корпусов и наборов данных достаточного объема для конкретной языковой пары, низкое число грамотных носителей, профессионально владеющих исходным языком и языком перевода, недостаточный уровень лингвистического описания языка, отсутствие ресурсов для экспертного обучения	Уделять больше внимания постобработке или не применять АГТ без постобработки
Отсутствие накопленных языковых корпусов и наборов данных требуемого объема для конкретной тематической области	Уделять больше внимания постобработке или не применять АГТ без постобработки
Наличие в исходном тексте смешения нескольких языков, нескольких вариантов (диалектов, региональных видов) языков и сочетание различных видов письменности	Уделять больше внимания предобработке и постобработке или не применять АГТ без постобработки
Наличие в исходном контенте графических, аудиовизуальных (нетекстовых) данных, представляющих собой важный для понимания текста контекст	Уделять больше внимания предобработке и постобработке или не применять АГТ без постобработки
Исходный текст для перевода получен путем некачественного ручного перевода или путем машинного перевода с другого языка	Уделять больше внимания предобработке и постобработке или не применять АГТ без постобработки
Исходный текст содержит неконвенциональные или полисемантические сокращения, аббревиатуры	Уделять больше внимания предобработке и постобработке или не применять АГТ без постобработки
Исходный текст включает содержательные изъяны, непоследовательность терминологии, аббревиатур, расхождения с графическими данными	Уделять больше внимания предобработке и постобработке или не применять АГТ без постобработки
Исходный текст является уникальным произведением, содержащим статистически редкие словосочетания, неологизмы, искусственные понятия, термины, абстракцию	Уделять больше внимания предобработке и постобработке или не применять АГТ без постобработки
Потенциальные ошибки в переводе несут высокую опасность для жизни и здоровья людей, окружающей среды, суверенитета и независимости	Уделять больше внимания предобработке и постобработке или не применять АГТ без постобработки

## **Приложение Ж. Пример рекомендаций докладчику, речь которого подлежит распознаванию или переводу с использованием средств АГТ**

В случае использования средств АГТ докладчику следует стремиться:

1. Говорить громко и четко, не иметь дефектов речи.
2. Использовать один язык высказываний, минимизировать использование англицизмов.
3. Упрощать предложения, избегать сложносочиненных связей, сложного синтаксиса.
4. Избегать слов-паразитов и звуков-паразитов в речи.
5. Не использовать аббревиатуры, в особенности неконвенциональные, вместо аббревиатур использовать полные расшифровки.
6. Обеспечить исправность и работу оборудования, включая микрофоны, без перебоев.
7. Обеспечить хорошую акустику помещения, исключить фоновые шумы (примеры фоновых шумов: ветер, тряска в дороге, шум аэропорта, жужжание проектора).
8. Правильно пользоваться микрофоном, находиться на оптимальном расстоянии в зоне действия микрофона.
9. В русском языке избегать причастных и деепричастных оборотов, а также нанизывания существительных в родительном падеже.
10. Избегать местоимений.
11. Избегать многозначности.
12. Избегать фразеологии, отсылок, аллюзий, иных интертекстовых явлений, средств выразительности без разъяснений.
13. Соблюдать единообразие в использовании терминов и имен собственных.
14. Использовать ту же лексику, что и в презентации или визуализации.
15. При использовании редкой или многозначной лексики использовать в том же предложении слова, позволяющие определить контекст (например, не «шасси не выдержало удара о землю», а «шасси самолета не выдержало удара о землю»).
16. Делать смысловые паузы между предложениями.
17. Замолкать, если докладчика перебивают, не допускать одновременного говорения нескольких человек.
18. Начинать фразу заново, если докладчик сбился.

### Приложение 3. Пример требований к системам АГТ (для разработчиков)

1. Учет контекста и ранжированных глоссариев.
2. Возможность настройки систем АГТ, когда при формировании перевода (или текста) учитывается весь проект (весь текст) целиком, без сегментации (разделения на не связанные между собой сегменты) или иной фрагментации.
3. Возможность постоянного пополнения и дообучения систем АГТ глоссариями и памятью переводов на основе обратной связи участников ручного процесса постобработки.
4. Исключение нарушения (авторских) прав на текстовые и речевые корпуса, используемые для обучения систем АГТ.
5. Соблюдение тайны текстов, предназначенных для перевода, самого перевода, метаданных, исполнителей, заказчиков. Очистка текстовых и речевых корпусов, используемых для обучения систем АГТ, от информации ограниченного доступа, включая персональные данные, а также от сведений, составляющих государственную, коммерческую, налоговую, банковскую, врачебную, нотариальную тайну, либо принятие мер по нераспространению такой информации через системы АГТ (деперсонализация, сортировка строк в случайном порядке и другие методы).
6. Учет возможного наличия в тексте разных видов письменности, разных направлений письма и их сочетания, а также нетекстовых элементов.
7. Возможность оставления части текста без перевода.
8. Возможность перевода с нескольких языков, доперевода многоязычного исходного текста при согласовании с теми частями текста, которые в исходном тексте (документе, произведении) уже сформулированы на языке перевода.
9. Обеспечение возможности для комбинирования технологий АГТ с технологиями использования памяти перевода, полностью ручного перевода с возможностью гибкой настройки приоритетов, а также записью авторов тех или иных фрагментов текста с целью конкретизации и разграничения ответственности людей (специалистов) за тот или иной текст, а также определения их авторства и авторских прав, в том числе с целью принятия корректирующих мер, направленных на повышение качества.

## Приложение И. Нормативы постобработки и расчет размера оплаты

Нормативы производительности постобработки зависят от требований к уровню качества, указанных в техническом задании заказчика, и от объема усилий, необходимых для приведения выдачи в соответствие с требуемым уровнем качества.

Следует учитывать, что даже при достаточном уровне качества выдачи и отсутствии требований к стилю в большинстве случаев заказчик все же требует перевод, который не содержит фактических ошибок. Чтобы обеспечить это, исполнитель должен внимательно прочитать весь исходный текст и текст перевода и проверить соответствие чисел, дат, отсутствие фактических ошибок, корректность терминов, сумм, единиц измерения и т. п.

Стоимость такой проверки может превысить стоимость нового перевода того же текста.

Без этого исполнитель не может гарантировать соблюдение требований заказчика, указанных в ТЗ. Так как не существует достоверного способа оценить финансовые потери даже от мелкой фактической ошибки, весь текст перевода должен быть подвергнут постобработке целиком в той степени, чтобы обеспечить и гарантировать соблюдение ТЗ.

## Приложение К. Библиография

- ГОСТ 1.1-2002 «Межгосударственная система стандартизации. Термины и определения».
- ИСО 832:1994 «Информация и документация. Библиографическое описание и ссылки. Правила сокращения библиографических терминов».
- ГОСТ 8.417-2002 «Государственная система обеспечения единства измерений. Единицы величин».
- ГОСТ 2.004-88 «Единая система конструкторской документации. Общие требования к выполнению конструкторских и технологических документов на печатающих и графических устройствах вывода ЭВМ».
- ГОСТ 2.051-2006 «Единая система конструкторской документации. Электронные документы. Общие положения».
- ГОСТ 2.601-2006 «Единая система конструкторской документации. Эксплуатационные документы».
- ГОСТ Р 2.901-99 «Единая система конструкторской документации. Документация, отправляемая за границу. Общие требования».
- ГОСТ 7.36-2006 «Система Стандартов по информации, библиотечному и издательскому делу. Неопубликованный перевод. Общие требования и правила оформления».
- ГОСТ 7.79-2000 «Система Стандартов по информации, библиотечному и издательскому делу. Правила транслитерации кирилловского письма латинским алфавитом».
- ГОСТ 7.11-2004 «Система Стандартов по информации, библиотечному и издательскому делу. Библиографическая запись. Сокращение слов и словосочетаний на иностранных европейских языках».
- ГОСТ 34.003-90 «Информационные технологии. Комплекс Стандартов на автоматизированные системы. Автоматизированные системы. Термины и определения».
- ГОСТ Р 7.0.3-2006 «Система стандартов по информации, библиотечному и издательскому делу. Издания. Основные элементы. Термины и определения».
- ГОСТ Р ЕН 15038-2014 «Услуги населению. Услуги переводческие. Требования к услугам».
- Меморандум Ассоциации переводческих компаний по вопросам институционализации, стандартизации и создания нормативно-правовой базы для применения машинного перевода и прочих технологий автоматической генерации текста в качестве профессиональных инструментов в области перевода и локализации // URL: <https://atcru.org/upload/iblock/ff5/fp58d29psjlsr3minttu0nuqqntlay50.pdf> (дата обращения: 06.02.2025).
- EN 15038:2006 Translation services — Service requirements.
- ISO/DIS 17100 Translation Services — Requirements for translation services.
- ISO 18587:2017 Translation services — Post-editing of machine translation output — Requirements.
- ISO 5060:2024 Translation services — Evaluation of translation output — General guidance.