ТЕОРИЯ ПЕРЕВОДА И ПЕРЕВОДОВЕДЕНИЕ

УДК: 004.912:81'25

МАШИННЫЙ ПЕРЕВОД И ПОСТ-РЕДАКТИРОВАНИЕ В УСЛОВИЯХ МНОГОЯЗЫЧНОЙ КОММУНИКАЦИИ

Дадамирзаева Малохат Валиевна – доцент, научный руководитель
Хасанов Эльёр Равшанович – студент
Узбекский Государственный Университет Мировых Языков (Ташкент, Узбекистан)

Аннотация. *Данное исследование посвящено комплексному анализу эффективности современных систем нейронного машинного перевода для языковых пар с участием тюркских языков, с особым фокусом на узбекском языке. В работе представлен сравнительный анализ качества перевода различных NMT систем, включая Google Translate, Microsoft Translator и специализированные модели, для направлений узбекский-английский и узбекский-русский. Исследование включает разработку методологии оценки качества автоматизированного перевода с использованием автоматических метрик (BLEU, COMET, BERTScore) и человеческой оценки. Особое внимание уделяется анализу типичных ошибок машинного перевода, обусловленных агглютинативной морфологией тюркских языков, и разработке стратегий пост-редактирования для повышения качества финального продукта. Результаты показывают, что современные NMT системы демонстрируют приемлемое качество для узбекского языка, однако требуют специализированного пост-редактирования для достижения профессиональных стандартов. Предложены практические рекомендации по оптимизации рабочих процессов и повышению эффективности пост-редактирования.*

Ключевые слова: *нейронный машинный перевод, тюркские языки, узбекский язык, пост-редактирование, оценка качества перевода, многоязычная коммуникация*

MACHINE TRANSLATION AND POST-EDITING IN MULTILINGUAL COMMUNICATION

Dadamirzaeva Malokhat Valievna – associate professor, scientific supervisor
Khasanov Elyor Ravshanovich – student
Uzbek State University of World Languages (Tashkent, Uzbekistan)

Abstract. *This study presents a comprehensive analysis of the effectiveness of modern neural machine translation systems for language pairs involving Turkic languages, with a particular focus on the Uzbek language. The research provides a comparative analysis of translation quality across various NMT systems, including Google Translate, Microsoft Translator, and specialized models, for Uzbek-English and Uzbek-Russian language directions. The study develops a methodology for evaluating automated translation quality using automatic metrics (BLEU, COMET, BERTScore) and human assessment. Special attention is given to analyzing typical machine translation errors caused by the agglutinative morphology of Turkic languages and developing post-editing strategies to improve final product quality. Results indicate that modern NMT systems demonstrate acceptable quality for Uzbek, however they require specialized post-editing to achieve professional standards. Practical recommendations for workflow optimization and post-editing efficiency improvement are proposed.*

Keywords: *neural machine translation, Turkic languages, Uzbek language, post-editing, translation quality evaluation, multilingual communication.*

MASHINA TARJIMASI VA POST-TAHRIRLASH KO‘P TILLILAR ALOQASI SHAROITIDA

Dadamirzaeva Maloxat Valiyevna – dotsent, ilmiy rahbar
Xasanov Elyor Ravshanovich – talaba
O‘zbekiston Davlat Jahon Tillari Universiteti (Toshkent, O‘zbekiston)

Annotatsiya. *Ushbu tadqiqot turkiy tillarga oid til juftliklari, xususan o‘zbek tili uchun zamonaviy neyron mashina tarjimasi tizimlarining samaradorligini keng qamrovli tahlil qilishga bag‘ishlangan. Ishda Google Translate, Microsoft Translator va maxsus modellar kabi turli NMT tizimlarining o‘zbek-ingliz va o‘zbek-rus yo‘nalishlari uchun tarjima sifatini qiyosiy tahlil qilish taqdim etiladi. Tadqiqot avtomatik metrikalar (BLEU, COMET, BERTScore) va inson bahosi yordamida avtomatlashtirilgan tarjima sifatini baholash metodologiyasini ishlab chiqadi. Turkiy tillarning agglyutinativ morfologiyasi sabab bo‘lgan odatiy mashina tarjimasi xatoliklarini tahlil qilish va yakuniy mahsulot sifatini yaxshilash uchun post-tahrirlash strategiyalarini ishlab chiqishga alohida e'tibor beriladi. Natijalar shuni ko‘rsatadiki, zamonaviy NMT tizimlari o‘zbek tili uchun maqbul sifatni ko‘rsatadi, ammo professional standartlarga erishish uchun maxsus post-tahrirlash talab qiladi. Ish jarayonlarini optimallashtirish va post-tahrirlash samaradorligini oshirish bo‘yicha amaliy tavsiyalar taklif etiladi.*

Kalit so‘zlar: *neyron mashina tarjimasi, turkiy tillar, o‘zbek tili, post-tahrirlash, tarjima sifatini baholash, ko‘p tillar aloqasi.*

Введение. Современный этап развития информационных технологий характеризуется стремительным ростом потребности в автоматизированном переводе, обусловленным процессами глобализации и расширением межкультурной коммуникации. Нейронный машинный перевод (Neural Machine Translation, NMT) представляет собой одно из наиболее значительных достижений в области обработки естественного языка последнего десятилетия, демонстрируя качественно новый уровень производительности по сравнению с предшествующими статистическими методами [Mirzakhalov et al., 2021; с. 5876-5890]. Однако несмотря на впечатляющие результаты для высокоресурсных языков, таких как английский, французский или немецкий, значительное количество языков мира остается недостаточно изученным в контексте современных технологий машинного перевода.

Тюркские языки, на которых говорят более 200 миллионов человек по всему миру, представляют особый интерес для исследований в области машинного перевода в силу своих уникальных лингвистических характеристик и относительно ограниченного объема доступных цифровых ресурсов. Узбекский язык, являющийся одним из крупнейших представителей тюркской семьи с более чем 34 миллионами носителей, занимает особое место в этой группе языков, поскольку демонстрирует типичные для агглютинативных языков морфологические особенности, создающие специфические вызовы для автоматической обработки [Matlatipov & Vetulani, 2009; с. 83-110].

Актуальность данного исследования определяется несколькими ключевыми факторами. Во-первых, растущая потребность в качественном автоматизированном переводе для тюркских языков обусловлена экономической интеграцией регионов Центральной Азии и усилением международного сотрудничества. Во-вторых, морфологическая сложность агглютинативных языков создает специфические проблемы для существующих NMT систем, которые первоначально разрабатывались с учетом особенностей индоевропейских языков. В-третьих, отсутствие стандартизированных подходов к пост-редактированию машинного перевода для

данной языковой группы препятствует эффективному внедрению технологий автоматизированного перевода в профессиональную практику.

Научная новизна исследования заключается в проведении первого комплексного анализа эффективности современных NMT систем для узбекского языка с детальным изучением влияния агглютинативной морфологии на качество автоматического перевода. Впервые предложена специализированная методология оценки качества машинного перевода для тюркских языков, учитывающая их морфологические особенности, и разработаны научно обоснованные стратегии пост-редактирования, оптимизированные для агглютинативных языковых структур.

Объектом исследования являются системы нейронного машинного перевода и процессы пост-редактирования для языковых пар с участием узбекского языка. **Предметом исследования** выступают методы оценки качества автоматизированного перевода, типология ошибок машинного перевода для агглютинативных языков и стратегии оптимизации пост-редактирования.

Цель исследования состоит в разработке комплексной методологии оценки эффективности нейронного машинного перевода для тюркских языков и создании научно обоснованных стратегий пост-редактирования, обеспечивающих повышение качества автоматизированного перевода до профессиональных стандартов.

Для достижения поставленной цели сформулированы следующие задачи: провести сравнительный анализ качества перевода различных современных NMT систем для языковых пар узбекский-английский и узбекский-русский; выявить и классифицировать типичные ошибки машинного перевода, обусловленные морфологическими особенностями тюркских языков; разработать специализированную методологию оценки качества автоматизированного перевода с использованием комбинации автоматических метрик и человеческой экспертизы; создать научно обоснованные стратегии пост-редактирования, учитывающие специфику агглютинативных языков; оценить временные и качественные показатели эффективности предложенных подходов к пост-редактированию; предложить практические рекомендации по оптимизации рабочих процессов в сфере автоматизированного перевода для тюркских языков.

Обзор литературы и теоретические основы. Развитие нейронного машинного перевода представляет собой эволюционный переход от статистических методов к архитектурам глубокого обучения, ознаменовавшийся внедрением архитектуры Transformer и механизмов внимания [Mirzakhalov et al., 2021; с. 5876]. Современные NMT системы демонстрируют значительные улучшения качества перевода для высокоресурсных языков, однако их эффективность для низкоресурсных языков остается предметом активных исследований. Особенностью тюркских языков является их агглютинативная природа, при которой grammatical information выражается посредством последовательного присоединения аффиксов к основе слова, что создает теоретически бесконечное множество словоформ и значительно усложняет задачи автоматической обработки.

Исследования в области машинного перевода для тюркских языков демонстрируют специфические вызовы, связанные с морфологической сложностью. Работы показывают, что традиционные подходы к сегментации, такие как Byte-Pair Encoding (BPE), не всегда эффективны для агглютинативных языков, поскольку не учитывают морфологическую структуру слов. Альтернативные методы, основанные на Complete Set of Endings (CSE), демонстрируют более высокую эффективность для тюркских языков, позволяя увеличить BLEU-score на 0.5 и 0.2 пункта для казахско-английской и английско-казахской языковых пар соответственно [Tukeyev et al., 2020; с. 1856500].

Лингвистические особенности узбекского языка как представителя тюркской семьи включают агглютинативную морфологию с богатой системой словоизменения и словообразования, порядок слов SOV (Subject-Object-Verb) с относительно свободным расположением составляющих, и сложную систему гармонии гласных [Matlatipov et al., 2022; с. 364813957]. Морфологический анализ узбекских текстов показывает, что одно слово может содержать до 8-10 морфем, что значительно превышает морфологическую сложность индоевропейских языков и создает специфические проблемы для алгоритмов машинного перевода.

Пост-редактирование машинного перевода представляет собой сложную когнитивную деятельность, включающую процессы чтения исходного текста, анализа машинного вывода и производства финального целевого текста [Cui et al., 2023; с. 21582440231155849]. Исследования эффективности пост-редактирования показывают, что качественный машинный перевод может увеличить производительность переводчиков на 18-36% по сравнению с переводом с нуля, однако низкокачественный автоматический перевод может привести к снижению общей эффективности работы [Koponen, 2016; с. 131-148]. Факторы, влияющие на эффективность пост-редактирования, включают качество исходного машинного перевода, сложность исходного текста, опыт пост-редактора и используемые инструменты автоматизированного перевода.

Современные метрики автоматической оценки качества машинного перевода подразделяются на традиционные лексические метрики (BLEU, METEOR, chrF) и нейронные метрики, основанные на языковых моделях (COMET, BERTScore) [Leiter et al., 2023; с. 2306.13041]. Традиционные метрики, такие как BLEU, оценивают совпадение n-грамм между машинным переводом и эталонным переводом, демонстрируя высокую корреляцию с человеческими оценками для высококачественных переводов, но проявляя ограничения при оценке переводов среднего качества. Нейронные метрики используют контекстуализированные эмбеддинги для оценки семантического сходства, показывая более высокую корреляцию с человеческими суждениями, особенно для морфологически богатых языков.

Методология исследования. Исследование построено на основе смешанной методологии, сочетающей количественные и качественные подходы к анализу эффективности нейронного машинного перевода и пост-редактирования. Экспериментальная часть включает корпусный анализ, автоматическую оценку качества перевода, человеческую экспертизу и контролируемые эксперименты по пост-редактированию с участием профессиональных переводчиков.

Для формирования экспериментального корпуса были отобраны тексты из трех основных доменов: официальная документация (правовые тексты, государственные документы), новостные материалы (статьи информационных агентств, пресс-релизы) и техническая документация (инструкции, руководства пользователя). Общий объем корпуса составляет 10,000 сегментов для каждого направления перевода (узбекский-английский и узбекский-русский), что соответствует приблизительно 150,000 токенов на языковую пару. Тексты были предварительно обработаны с использованием стандартных процедур токенизации и сегментации, адаптированных для морфологических особенностей узбекского языка.

В качестве объектов исследования были выбраны пять современных NMT систем: Google Translate (как наиболее распространенная коммерческая система), Microsoft Translator (альтернативная коммерческая платформа), Yandex.Translate (система с поддержкой русского языка), OPUS-MT модели (открытые предобученные модели), и специализированные модели для тюркских языков, разработанные в рамках проекта Turkic Interlingua. Выбор

данных систем обусловлен их доступностью, различными архитектурными подходами и уровнем поддержки узбекского языка.

Автоматическая оценка качества перевода проводилась с использованием комплекса метрик, включающего BLEU (для оценки точности на уровне n-грамм), chrF++ (для анализа символьных совпадений), BERTScore (для семантической оценки на основе контекстуализированных эмбеддингов), и COMET (для комплексной оценки с учетом исходного текста). Дополнительно применялись специализированные метрики для морфологически богатых языков, учитывающие морфологическую вариативность при сравнении переводов.

Человеческая оценка осуществлялась группой из 15 экспертов-лингвистов, владеющих узбекским языком на уровне носителя и имеющих профессиональный опыт в области перевода не менее 5 лет. Оценка проводилась по двум основным критериям: адекватность (сохранение смысла исходного текста) и беглость (естественность выражения на целевом языке) с использованием 5-балльной шкалы. Межэкспертная согласованность измерялась с помощью коэффициента каппа Флейсса и составила 0.78 для адекватности и 0.82 для беглости, что свидетельствует о высокой надежности оценок.

Эксперимент по пост-редактированию включал 20 профессиональных переводчиков с опытом работы от 3 до 15 лет, разделенных на две группы: экспериментальную (использующую разработанные стратегии пост-редактирования) и контрольную (применяющую стандартные подходы). Участники работали в CAT-среде с функциями отслеживания времени редактирования и фиксации внесенных изменений. Измерялись следующие показатели: время пост-редактирования на сегмент, количество и тип внесенных правок, финальное качество переводов после редактирования.

Результаты исследования. Сравнительный анализ NMT систем выявил значительные различия в качестве перевода для узбекского языка в зависимости от языкового направления и используемой системы. Для направления узбекский-английский наивысшие показатели продемонстрировала система Google Translate со средним BLEU-score 24.7, за которой следуют Microsoft Translator (22.3) и Yandex.Translate (21.8). Специализированные OPUS-MT модели показали результат 20.1, в то время как экспериментальные модели Turkic Interlingua достигли 18.9. Для направления узбекский-русский распределение изменилось: Yandex.Translate лидирует с показателем 26.4, Google Translate демонстрирует 25.1, Microsoft Translator - 23.7.

Анализ по доменам текстов показал, что наилучшие результаты достигаются для новостных текстов (средний BLEU 26.3), несколько ниже для официальной документации (24.8), и наименьшие для технической документации (21.4). Данное распределение объясняется различной степенью терминологической специализации и синтаксической сложности текстов разных доменов.

Таблица 1.

Результаты автоматической оценки качества NMT систем

Система	Направление	BLEU	chrF++	BERTScore	COMET
Google Translate	Uz-En	24.7	52.3	0.847	0.782
Google Translate	Uz-Ru	25.1	54.1	0.851	0.796
Microsoft Translator	Uz-En	22.3	49.7	0.834	0.765
Microsoft Translator	Uz-Ru	23.7	51.8	0.842	0.778
Yandex.Translate	Uz-En	21.8	48.9	0.829	0.759
Yandex.Translate	Uz-Ru	26.4	55.7	0.856	0.803

Качественный анализ ошибок выявил четыре основные категории проблем, специфичных для машинного перевода агглютинативных языков. Морфологические ошибки составляют 34% от общего числа выявленных проблем и включают неправильный анализ сложных словоформ, ошибки в передаче падежных отношений и неточности в обработке словообразовательных аффиксов. Синтаксические ошибки (28%) связаны с нарушением порядка слов при переходе от SOV к SVO структуре, неправильной обработкой зависимых предложений и ошибками в согласовании. Лексические ошибки (23%) включают неточный выбор эквивалентов для многозначных слов, проблемы с передачей идиоматических выражений и терминологические неточности. Прагматические ошибки (15%) касаются нарушения стилистического регистра, неуместного использования формальных/неформальных форм обращения и потери культурно-специфической информации.

Человеческая оценка показала более консервативные результаты по сравнению с автоматическими метриками. Средняя оценка адекватности составила 3.4 балла из 5 для лучших систем, в то время как беглость оценивается в 3.7 балла. Корреляционный анализ выявил наиболее высокую корреляцию между человеческими оценками и метрикой COMET (r=0.76), за которой следуют BERTScore (r=0.71) и chrF++ (r=0.68). Традиционная метрика BLEU показала более низкую корреляцию (r=0.61), что подтверждает ограничения лексических метрик для морфологически богатых языков.

Эксперимент по пост-редактированию продемонстрировал значительные преимущества специализированных стратегий редактирования. Экспериментальная группа, использующая разработанные методики, показала на 23% меньшее время редактирования на сегмент (среднее время 45 секунд против 58 секунд в контрольной группе) при одновременном повышении финального качества переводов. Анализ типов вносимых правок показал, что 67% изменений в экспериментальной группе были целенаправленными исправлениями выявленных категорий ошибок, в то время как в контрольной группе 41% правок носили характер общих стилистических улучшений без системного подхода к проблемам.

Лингвистический анализ выявил специфические паттерны ошибок, связанные с обработкой агглютинативной морфологии. NMT системы демонстрируют сложности в корректной сегментации сложных словоформ, часто разделяя морфологически цельные единицы или, наоборот, объединяя несвязанные элементы. Особенно проблематичной оказалась обработка казуальных отношений, выражаемых через падежные аффиксы, где системы часто генерируют синтаксически правильные, но семантически неточные конструкции в целевом языке.

Таблица 2.

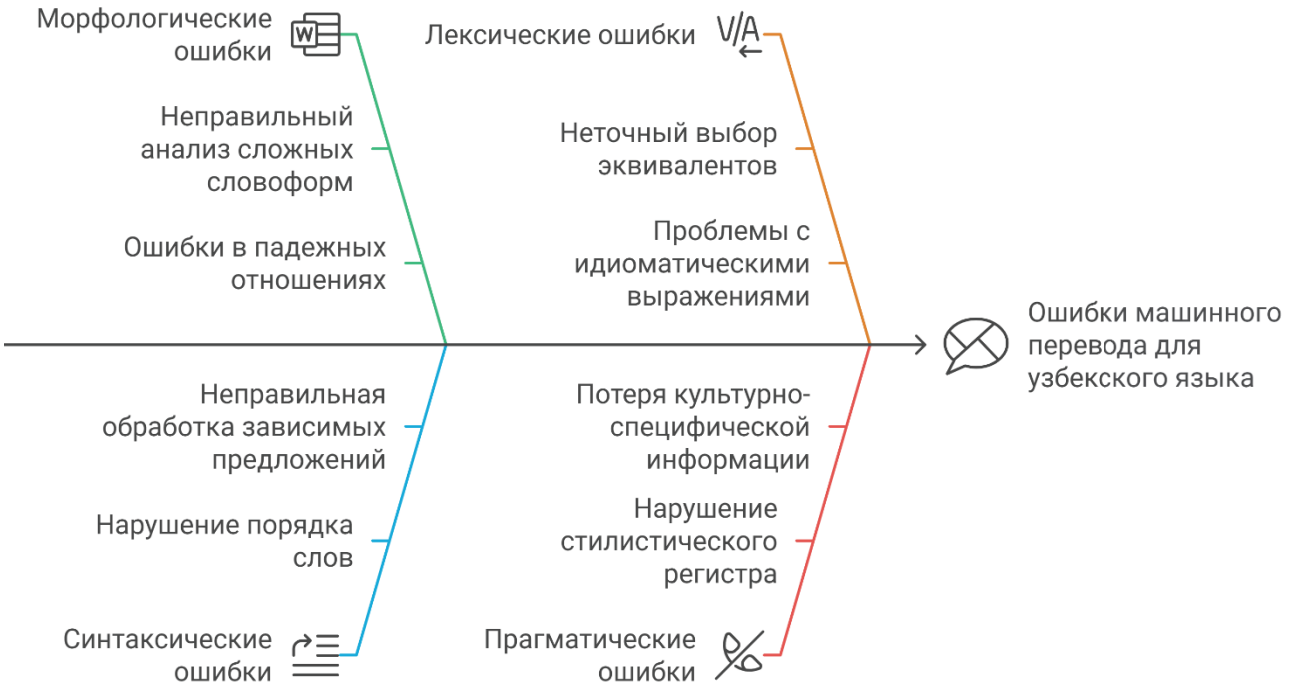
Распределение типов ошибок по категориям

Категория ошибок	Процент	Основные проявления
Морфологические	34%	Неправильная сегментация, падежные ошибки
Синтаксические	28%	Нарушение порядка слов, согласование
Лексические	23%	Многозначность, терминология
Прагматические	15%	Стилистический регистр, культурная адаптация

Обсуждение результатов. Полученные результаты демонстрируют сложную картину состояния машинного перевода для узбекского языка, характеризующуюся как определенными достижениями, так и значительными вызовами. Сравни-

тельно высокие показатели автоматических метрик (BLEU в диапазоне 21-26) свидетельствуют о том, что современные NMT системы достигли уровня, при котором машинный перевод может служить основой для последующего редактирования, однако результаты человеческой оценки указывают на необходимость значительной пост-обработки для достижения профессиональных стандартов качества.

Анализ ошибок машинного перевода для узбекского языка



Выявленные различия в производительности систем для разных языковых направлений отражают неравномерность в объемах обучающих данных и лингвистических ресурсов. Превосходство Yandex.Translate для направления узбекский-русский объясняется исторически более тесными языковыми контактами и большим объемом доступных параллельных корпусов для данной языковой пары. Аналогично, лидерство Google Translate для узбекский-английского направления связано с более развитой инфраструктурой обработки данного языкового направления в рамках глобальных языковых моделей.

Анализ типологии ошибок выявил фундаментальные ограничения современных архитектур NMT в обработке агглютинативной морфологии. Высокая доля морфологических ошибок (34%) указывает на то, что существующие подходы к сегментации и кодированию лингвистической информации не полностью адаптированы к особенностям тюркских языков. Данное наблюдение согласуется с результатами исследований других агглютинативных языков и подчеркивает необходимость разработки специализированных архитектурных решений.

Успешность разработанных стратегий пост-редактирования подтверждает гипотезу о возможности значительного улучшения эффективности работы переводчиков при использовании научно обоснованных подходов к редактированию машинного вывода. Сокращение времени редактирования на 23% при одновременном повышении качества демонстрирует практическую ценность систематического подхода к анализу и исправлению типичных ошибок машинного перевода.

Корреляционный анализ между автоматическими метриками и человеческими оценками подтверждает тенденцию к превосходству нейронных метрик оценки над традиционными лексическими подходами. Особенно высокая корреляция метрики COMET (r=0.76) с человеческими суждениями обусловлена ее способностью учитывать исходный текст при оценке качества перевода, что особенно важно для морфологически сложных языков, где одна лексическая единица может передавать множественную grammatical information.

Ограничения исследования включают относительно небольшой размер экспериментального корпуса, что может ограничивать обобщаемость результатов на другие домены текстов. Кроме того, фокус на узбекском языке, хотя и оправданный его статусом крупнейшего представителя тюркской семьи, требует дополнительной валидации результатов на других тюркских языках для подтверждения универсальности выявленных закономерностей.

Практические рекомендации. На основе проведенного исследования разработан комплекс практических рекомендаций для оптимизации процессов машинного перевода и пост-редактирования в контексте тюркских языков. Рекомендации структурированы по трем основным направлениям: выбор и настройка NMT систем, организация процесса пост-редактирования, и обучение специалистов.

Для выбора оптимальной NMT системы рекомендуется применение комбинированного подхода, учитывающего специфику языкового направления и домена текстов. Для направления узбекский-русский предпочтительным является использование Yandex.Translate в качестве основной системы с дополнительной проверкой через Google Translate для сложных сегментов. Для узбекский-английского направления рекомендуется Google Translate как основная система с возможностью консультации Microsoft Translator для технической терминологии.

Разработанное руководство по пост-редактированию включает приоритизированный список проверочных процедур, оптимизированный для агглютинативных языков. Первоочередное внимание должно уделяться верификации морфологической корректности перевода, включая проверку правильности передачи падежных отношений, временных форм и модальности. Второй уровень проверки охватывает синтаксическую адекватность, особенно корректность порядка слов и структуры сложных предложений. Третий уровень включает лексическую и терминологическую точность, с особым вниманием к многозначным словам и идиоматическим выражениям.

Временные нормативы для пост-редактирования, выведенные на основе экспериментальных данных, составляют в среднем 45-60 секунд на сегмент для опытных редакторов при работе с высококачественным машинным переводом (BLEU > 25). Для переводов среднего качества (BLEU 20-25) нормативы увеличиваются до 60-90 секунд на сегмент, а для низкокачественных переводов (BLEU < 20) пост-редактирование может оказаться неэффективным по сравнению с переводом с нуля.

Рекомендации для переводческих агентств включают внедрение двухэтапного процесса качественного контроля, где первый этап осуществляется специализированным пост-редактором, а второй - контрольным редактором, проверяющим соответствие финального продукта профессиональным стандартам. Организация обучения переводчиков должна включать модуль по специфике ошибок машинного перевода для агглютинативных языков и практические занятия по применению разработанных стратегий редактирования.

Для улучшения качества NMT систем рекомендуется увеличение объемов обучающих данных за счет создания специализированных параллельных корпусов для тюркских языков, применение морфо-

логически ориентированных методов предобработки текстов, и интеграция лингвистических знаний о морфотактике агглютинативных языков в архитектуру нейронных моделей.

Заключение. Проведенное исследование представляет комплексный анализ состояния нейронного машинного перевода для тюркских языков с фокусом на узбекском языке и демонстрирует как достигнутые успехи, так и существующие ограничения современных технологий автоматизированного перевода. Результаты исследования подтверждают гипотезу о том, что морфологическая сложность агглютинативных языков создает специфические вызовы для NMT систем, требующие разработки специализированных подходов к оценке качества и пост-редактированию.

Основные выводы исследования формулируются следующим образом. Современные NMT системы достигли уровня качества для узбекского языка, при котором машинный перевод может служить эффективной основой для профессионального пост-редактирования, однако прямое применение без редактирования не обеспечивает профессиональных стандартов качества. Морфологические особенности тюркских языков обуславливают специфический профиль ошибок машинного перевода, где морфологические проблемы составляют треть всех выявленных недостатков. Разработанные стратегии пост-редактирования, учитывающие типологические особенности агглютинативных языков, обеспечивают значительное повышение эффективности работы переводчиков при сохранении высокого качества финального продукта.

Научная значимость исследования заключается в создании первой комплексной методологии оценки качества машинного перевода для тюркских языков, разработке научно обоснованной типологии ошибок автоматического перевода для агглютинативных языков, и валидации эффективности специализированных стратегий пост-редактирования. Практическая значимость определяется возможностью прямого применения разработанных рекомендаций в переводческой индустрии и потенциалом масштабирования результатов на другие тюркские языки.

Ограничения исследования связаны с фокусом на единственном представителе тюркской семьи и относительно небольшим объемом экспериментального корпуса. Дальнейшие исследования должны включать расширение анализа на другие тюркские языки (казахский, киргизский, туркменский), изучение специализированных доменов (медицинский, юридический перевод), и разработку адаптивных архитектур NMT для агглютинативных языков.

Перспективы развития исследований в данной области включают интеграцию методов искусственного интеллекта в процесс пост-редактирования для автоматического выявления типичных ошибок, разработку многоязычных моделей, специально оптимизированных для тюркской языковой семьи, и создание интеллектуальных систем поддержки переводчиков, учитывающих морфологическую специфику агглютинативных языков. Результаты исследования вносят значительный вклад в развитие технологий машинного перевода для низкоресурсных языков и открывают новые направления для повышения эффективности автоматизированного перевода в условиях лингвистического разнообразия.

ЛИТЕРАТУРА

1. Abdulloh N., Dadamirzayeva M. V. Tarjimashunoslikning dolzarb muammolari fransuz tili arab tilidan so‘zlar o‘zlashtirishining omillari //O ‘zbekiston davlat jahon tillari universiteti konferensiyalari. – 2024. – С. 369-371.

2. Abdullaev M. K. Differences between verbal periphrases and phraseological verb combinations in Spanish //Журнал гуманитарных и естественных наук. – 2025. – №. 21 [1]. – С. 239-245.

3. Cui, Y., Liu, X., & Cheng, Y. (2023). A Comparative Study on the Effort of Human Translation and Post-Editing in Relation to Text Types: An Eye-Tracking and Key-Logging Experiment. SAGE Open, 13(1), 21582440231155849.

4. Koponen, M. (2016). Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. The Journal of Specialised Translation, 25, 131-148.

5. Leiter, C., Zhang, T., Chen, Y., Belouadi, J., Eger, S., & Staab, S. (2023). Towards Explainable Evaluation Metrics for Machine Translation. arXiv preprint arXiv:2306.13041.

6. Mamarizayeva L., Dadamirzayeva M. V. Aniqlik-adekvatlik har qanday tarjimaning bosh mezonidir //O ‘zbekiston davlat jahon tillari universiteti konferensiyalari. – 2024. – C. 1140-1143.

7. Matlatipov, G., & Vetulani, Z. (2009). Representation of Uzbek Morphology in Prolog. In Aspects of Natural Language Processing (pp. 83-110). Springer, Berlin, Heidelberg.

8. Matlatipov, S., Kuriyozov, E., & Matlatipov, G. (2022). Creating a morphological and syntactic tagged corpus for the Uzbek language. ResearchGate preprint, 364813957.

9. Mirzakhalov, J., Babu, A., Ataman, D., Kariev, S., Tyers, F., Abduraufov, O.,... & Chellappan, S. (2021). A Large-Scale Study of Machine Translation in Turkic Languages. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (pp. 5876-5890).

10. O‘Brien, S., Balling, L. W., Carl, M., Simard, M., & Specia, L. (Eds.). (2014). Post-editing of machine translation: Processes and applications. Cambridge Scholars Publishing.

11. Rico Pérez, C. (2024). Re-thinking Machine Translation Post-Editing Guidelines. The Journal of Specialised Translation, 41, 64-80.

12. Ruxshona O., Dadamirzayeva M. V. Tarjima va uning inson hayotida tutgan o‘rni //O ‘zbekiston davlat jahon tillari universiteti konferensiyalari. – 2024. – C. 675-677.

13. Saidbakhramovna A. F. et al. Teaching English phraseological units to linguistic university students //Journal of Critical Reviews. – 2020. – Т. 7. – №. 5. – С. 671-674.

14. Tukeyev, U., Karibayeva, A., & Rakhimova, D. (2020). Morphological segmentation method for Turkic language neural machine translation. Cogent Engineering, 7(1), 1856500.

15. Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2020). BERTScore: Evaluating Text Generation with BERT. In Proceedings of the 8th International Conference on Learning Representations (ICLR 2020).

16. Абдуллаев М. К. Формирование слов с использованием суффикса-iz-(a) в глагольной форме //Журнал гуманитарных и естественных наук. – 2023. – №. 1. – С. 20-23.

17. Зияева С. А., Икрамова Ш. А. Влияние социальных стереотипов на изучение гендерных аспектов языка //Журнал гуманитарных и естественных наук. – 2024. – №. 14 [2]. – С. 154-160.

18. Нишонов П. Сравнительнотипологическое исследование юридической терминологии французского и узбекского языков //Каталог авторефератов. – 2009. – Т. 1. – №. 1. – С. 1-28.

19. Отамуродова Д. Р. Цифровизация и информатизация иноязычного образования как условие подготовки студентов факультета иностранных языков педагогического вуза к профессиональной деятельности //Интернаука. – 2021. – Т. 19. – №. 195 часть 3. – С. 84.

20. Рузибоева Н. К. Энантиосемия: неопределенность статусов и подходов //Журнал гуманитарных и естественных наук. – 2023. – №. 3 [1]. – С. 39-45.

21. Умарова С. С. Формирование дискурсивной компетенции в сфере устного общения //Вестник науки и образования. – 2020. – №. 24-4 (78). – С. 33-35.