СОПОСТАВИТЕЛЬНЫЙ АНАЛИЗ ДООБУЧЕНИЯ МУЛЬТИЯЗЫЧНЫХ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ МАШИННОГО ПЕРЕВОДА ДЛЯ УЗБЕКСКО-РУССКОЙ ЯЗЫКОВОЙ ПАРЫ

Авезов Сухроб Собирович,
PhD, преподаватель кафедры русского языка и литературы
Бухарский государственный университет (Бухара, Узбекистан)
1990senigama@gmail.com
https://orcid.org/0000-0003-1137-3403

Аннотация. *В статье представлен сопоставительный анализ шести мультиязычных моделей нейронного машинного перевода (NLLB-200, M2M-100, mBART-50, OPUS-MT, Google Neural MT, SeamlessM4T) применительно к узбекско-русской языковой паре. Описан процесс дообучения (fine-tuning) указанных моделей на параллельном корпусе объёмом 185 тыс. выровненных сегментов. Качество перевода оценивалось по метрикам BLEU, chrF и TER. Результаты экспериментов показали, что наибольший прирост качества после дообучения демонстрирует модель NLLB-200-3.3B, достигшая значения BLEU 31,6 в направлении узбекский-русский. Обсуждены типичные ошибки каждой модели, предложены рекомендации по выбору архитектуры и стратегии дообучения для низкоресурсных тюркских языков.*

Ключевые слова: *нейронный машинный перевод, дообучение моделей, узбекско-русская языковая пара, NLLB-200, M2M-100, mBART-50, низкоресурсные языки, BLEU, трансферное обучение.*

Annotatsiya. *Maqolada o‘zbek-rus tillari juftligiga nisbatan oltita ko‘p tilli neyron mashina tarjimasi modeli (NLLB-200, M2M-100, mBART-50, OPUS-MT, Google Neural MT, SeamlessM4T)ning qiyosiy tahlili taqdim etilgan. Mazkur modellarni 185 mingta moslashtirilgan segment hajmidagi parallel korpus asosida qo‘shimcha o‘qitish (fine-tuning) jarayoni tavsiflangan. Tarjima sifati BLEU, chrF va TER metrikalari asosida baholandi. Tajriba natijalari shuni ko‘rsatdiki, qo‘shimcha o‘qitilgandan so‘ng sifatning eng katta o‘sishini NLLB-200-3.3B modeli namoyish etdi va u o‘zbek-rus yo‘nalishida BLEU 31,6 ko‘rsatkichiga erishdi. Har bir modelning tipik xatolari muhokama qilindi, shuningdek, past resursli turkiy tillar uchun arxitektura va qo‘shimcha o‘qitish strategiyasini tanlash bo‘yicha tavsiyalar taklif etildi.*

Kalit so‘zlar: *neyron mashina tarjimasi, modellarni qo‘shimcha o‘qitish, o‘zbek-rus tillari juftligi, NLLB-200, M2M-100, mBART-50, past resursli tillar, BLEU, transfer o‘qitish.*

Abstract. *The article presents a comparative analysis of six multilingual neural machine translation models (NLLB-200, M2M-100, mBART-50, OPUS-MT, Google Neural MT, SeamlessM4T) applied to the Uzbek-Russian language pair. The process of fine-tuning these models on a parallel corpus comprising 185 thousand aligned segments is described. Translation quality was evaluated using the BLEU, chrF, and TER metrics. The experimental results showed that the NLLB-200-3.3B model demonstrated the greatest improvement in quality after fine-tuning, reaching a BLEU score of 31.6 in the Uzbek-to-Russian direction. Typical errors of each model are discussed and recommendations are proposed for selecting an architecture and a fine-tuning strategy for low-resource Turkic languages.*

Keywords: *neural machine translation, fine-tuning of models, Uzbek-Russian language pair, NLLB-200, M2M-100, mBART-50, low-resource languages, BLEU, transfer learning.*

Введение. Машинный перевод для языков с ограниченными цифровыми ресурсами остаётся одной из нерешённых задач компьютерной лингвистики. Узбекский язык, на котором говорят более 35 миллионов человек, по-прежнему относится к категории низкоресурсных в контексте технологий обработки

естественного языка. Разрыв между качеством перевода для высокоресурсных пар (например, английский и немецкий) и качеством перевода для пары узбекский и русский превышает 15 пунктов BLEU по данным бенчмарка FLORES-200. Появление крупных мультиязычных моделей, таких как mBART, M2M-100, NLLB-200 и SeamlessM4T, открыло перспективы для преодоления этого разрыва. Данные модели предварительно обучены на сотнях языков и допускают тонкую настройку (fine-tuning) на относительно небольших параллельных корпусах, что делает их особенно привлекательными для исследователей, работающих с малоресурсными языковыми парами.

Между тем сопоставительные исследования, посвящённые дообучению мультиязычных моделей машинного перевода именно для узбекско-русской языковой пары, пока немногочисленны. Хотя отдельные работы и практические реализации уже существуют, систематические сравнительные эксперименты для данного направления остаются ограниченными. Как справедливо отмечают А.М.Хусаинова, В.А.Романов, А.М.Хан «выбор стратегии распределения параметров между языками в моделях многоязычного машинного перевода определяет то, насколько оптимально используется пространство параметров. Следовательно, выбранная стратегия напрямую влияет на конечное качество перевода» [1, с. 125]. Именно эта зависимость обосновывает актуальность экспериментального сравнения различных архитектур при дообучении.

Цель данной статьи состоит в проведении сопоставительного анализа шести моделей нейронного машинного перевода при их дообучении на параллельном узбекско-русском корпусе. Задачи исследования включают подготовку и очистку обучающих данных, проведение экспериментов по fine-tuning, количественную и качественную оценку результатов, а также формулирование рекомендаций по выбору модели для данной языковой пары. Как отмечается в обзоре Ж.Ван, С.Тан, Ж.Ло, Т.Цинь, Т.Я.Лю типичный подход к трансферному обучению в низкоресурсном нейронном машинном переводе состоит в предварительном обучении модели на вспомогательных, как правило, высокоресурсных языковых парах с последующим дообучением на низкоресурсной паре [2, с. 4639]. При этом, как показывают Э.Синьорони, П.Рыхлы и Р.Синьорони, низкоресурсный машинный перевод весьма чувствителен к гиперпараметрам и архитектурным настройкам модели [3, с. 39-40].

Материалы и методы. Основу экспериментального материала составил параллельный корпус, собранный из нескольких открытых источников. Корпус OPUS предоставил значительную долю сегментов, извлечённых из подкорпусов Tatoeba, GNOME, Ubuntu и QED. Дополнительно были использованы материалы из набора данных FLORES-200, разработанного командой Meta AI для оценки качества многоязычного перевода. Общий объём собранного корпуса до очистки составил

247 тыс. пар предложений. Процедура очистки включала удаление дубликатов, фильтрацию по длине (исключались пары, где одно из предложений содержало менее 3 или более 200 токенов), автоматическую проверку языковой принадлежности средствами библиотеки fastText и ручную верификацию случайной выборки из 2000 пар. После очистки в корпусе осталось 185.412 выровненных сегментов. Распределение по тематическим доменам оказалось неравномерным.

Таблица 1. Распределение параллельного корпуса по тематическим доменам

Тематический домен	Число сегментов	Доля (%)
Новостные тексты	52.430	28,3
Юридические документы	38.760	20,9
Техническая документация	31.215	16,8
Художественная литература	25.890	14,0
Субтитры и диалоги	21.347	11,5
Научные и учебные тексты	15.770	8,5
Итого	185.412	100,0

Корпус был разделён на обучающую (90%), валидационную (5%) и тестовую (5%) выборки методом стратифицированного разбиения, обеспечивающего пропорциональное представительство каждого домена во всех трёх подмножествах. Тестовая выборка (9.271 сегмент) использовалась исключительно для финальной оценки и не участвовала ни в обучении, ни в подборе гиперпараметров.

Для сопоставительного анализа были выбраны шесть моделей, представляющих различные архитектурные подходы к многоязычному машинному переводу. Перечислим их с указанием ключевых характеристик.

Таблица 2. Характеристики исследуемых моделей

Модель	Параметры	Число языков	Архитектура	Тип данных	Токенизация
NLLB-200-3.3B	3,3 млрд	200	Transformer (dense)	Не англо-центричные	SPM-200
M2M-100-1.2B	1,2 млрд	100	Transformer (dense)	Не англо-центричные	SentencePiece
mBART-50-MM	611 млн	50	Transformer (denoising)	Англо-центричные	SentencePiece
OPUS-MT	~74 млн	Пара (uz→ru)	Transformer (MarianNMT)	Двунаправленные	SentencePiece
Google NMT	Неизвестно	130+	GNMT (seq2seq)	Проприетарные	WordPiece
SeamlessM4T v2	2,3 млрд	196	Transformer (multimodal)	Не англо-центричные	SPM

Модель NLLB-200, разработанная в Meta AI, представляет собой семейство мультиязычных трансформерных моделей машинного перевода, включающее как dense-, так и MoE-варианты. В официальной документации отмечается, что NLLB – это «многоязычная модель машинного перевода, обученная на данных с использованием техник майнинга, адаптированных для малоресурсных языков, и поддерживающая более 200 языков» [4]. Согласно основной публикации команды NLLB, база добытого битекста содержит общедоступные веб-данные для 148

англоцентричных и 1465 неанглоцентричных языковых пар. Это позволяет рассматривать NLLB-200 как одну из наиболее репрезентативных архитектур для низкоресурсных языков, включая узбекский.

M2M-100 стала первой моделью, способной переводить между любыми двумя из 100 языков без посредничества английского. Размер её обучающего корпуса составляет 7,5 млрд пар предложений по 2200 направлениям. Для узбекского языка объём доступных данных существенно скромнее. Модель mBART-50-many-to-many является результатом многоязычного дообучения mBART на англоцентричном корпусе ML50. Это обстоятельство потенциально ограничивает её эффективность при переводе между неанглийскими парами. OPUS-MT (Helsinki-NLP) представляет собой серию двунаправленных моделей на базе MarianNMT, обученных на данных проекта OPUS. В отличие от предыдущих моделей, OPUS-MT для пары uz-ru является однонаправленной и значительно уступает по количеству параметров. Google Neural Machine Translation (GNMT) не допускает дообучения пользователем, поэтому для данной модели фиксировалось лишь базовое качество перевода *из коробки*. SeamlessM4T v2, напротив, является мультимодальной моделью Meta AI, допускающей fine-tuning. Её мультимодальный характер (текст, речь) придаёт ей уникальное положение среди рассматриваемых систем.

Дообучение проводилось на сервере с двумя GPU NVIDIA A100 (80 ГБ VRAM). Для всех моделей (за исключением GNMT) использовалась библиотека HuggingFace Transformers версии 4.38.0 совместно с DeepSpeed ZeRO Stage 2 для оптимизации памяти. Общие гиперпараметры были унифицированы в целях обеспечения корректности сопоставления. Скорость обучения составляла 2e-5 с линейным прогревом в течение первых 500 шагов. Размер батча равнялся 16 (с градиентным накоплением по 4 шага, эффективный размер батча 64). Максимальная длина входной последовательности была установлена в 256 токенов. Обучение продолжалось 10 эпох с ранней остановкой по метрике BLEU на валидационной выборке (терпение 3 эпохи). Регуляризация обеспечивалась посредством dropout (0,1) и снижения весов (weight decay 0,01).

Как отмечают О.О.Негматулоев, Д.О.Жорник и А.В.Мельников, «создание моделей нейронного машинного перевода основывается на использовании больших параллельных корпусов, что позволяет достичь высокого качества перевода для языковых пар с достаточным объёмом данных» [5, с. 31]. Для низкоресурсных языковых пар, однако, существенное значение имеют не только объём данных, но и выбор архитектуры модели, а также стратегия обучения. Именно поэтому в нашем эксперименте гиперпараметры сохранялись постоянными, а варьировалась только архитектура модели.

Качество перевода оценивалось по трём метрикам. BLEU (BiLingual Evaluation Understudy) измеряет совпадение n-грамм между машинным переводом и эталоном,

принимая значения от 0 до 100. chrF (Character F-score) вычисляет совпадение на уровне символьных n-грамм и менее чувствительна к морфологическим расхождениям, что особенно ценно для агглютинативных языков. TER (Translation Edit Rate) измеряет минимальное число правок, необходимых для приведения машинного перевода к эталону; чем ниже значение, тем лучше. Все метрики вычислялись при помощи библиотеки SacreBLEU версии 2.3.1. Дополнительно была проведена экспертная оценка 200 случайно выбранных переводов тремя лингвистами, владеющими обоими языками. Эксперты оценивали адекватность (сохранение смысла) и беглость (грамматическая корректность и естественность) по пятибалльной шкале.

Результаты. Результаты автоматической оценки для направления узбекский-русский сведены в таблице 3. Показатели до дообучения (baseline) и после fine-tuning позволяют судить как об исходном качестве каждой модели, так и о потенциале её адаптации.

Таблица 3. Результаты автоматической оценки (направление uz-ru)

Модель	BLEU (base)	BLEU (FT)	chrF (base)	chrF (FT)	TER (base)	TER (FT)
NLLB-200-3.3B	18,4	31,6	42,1	58,3	72,3	51,8
M2M-100-1.2B	12,7	24,3	35,8	50,1	79,6	60,4
mBART-50-MM	14,2	26,8	38,5	53,7	76,1	56,2
OPUS-MT	9,8	19,4	28,4	43,9	85,2	66,7
Google NMT	22,1	н/д	48,6	н/д	65,4	н/д
SeamlessM4T v2	20,5	29,2	45,3	56,1	68,7	54,3

Примечание. FT означает fine-tuned (после дообучения); н/д указывает на невозможность дообучения проприетарной модели.

Наибольший абсолютный прирост BLEU зафиксирован у NLLB-200 (+13,2 пункта). Второй по величине прирост показала mBART-50 (+12,6). Модель M2M-100 продемонстрировала прирост в +11,6 пункта, а OPUS-MT прибавила +9,6. SeamlessM4T прибавила +8,7, что ожидаемо, учитывая её более высокий стартовый уровень. Google NMT без дообучения показала BLEU 22,1, что объясняется использованием проприетарных данных и непрозрачной процедурой обучения. Показательно, что ранжирование моделей изменилось после дообучения. Если до fine-tuning лидировал Google NMT (22,1), за ним SeamlessM4T (20,5) и NLLB-200 (18,4), то после дообучения первая позиция безоговорочно перешла к NLLB-200 (31,6), обогнавшей даже проприетарную систему Google.

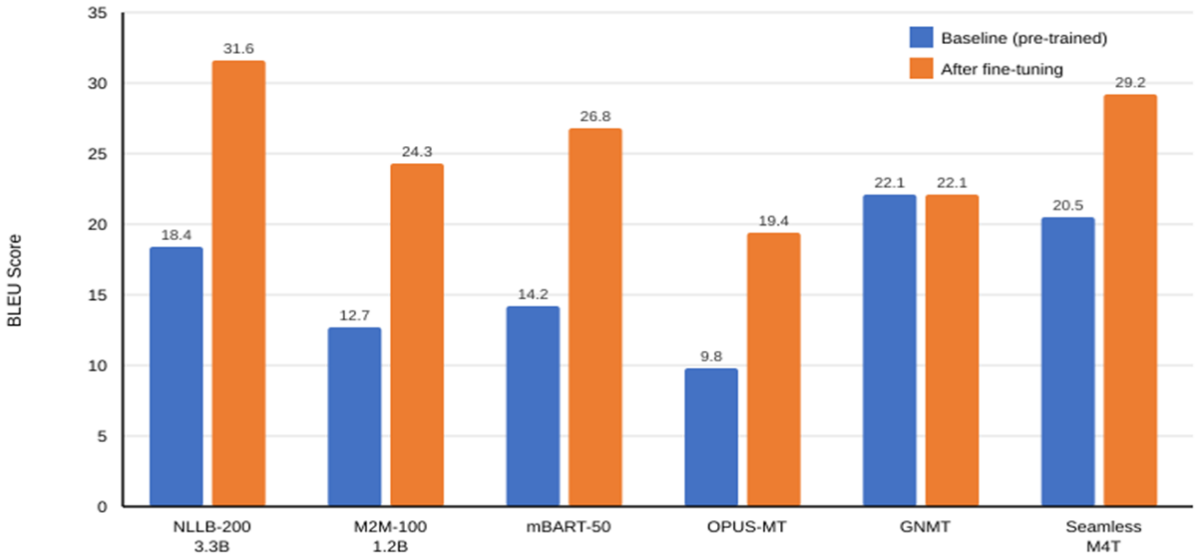

Рисунок 1. Сопоставление значений BLEU до и после дообучения (uz-ru)

Как отмечает М.Пост, в исследованиях машинного перевода измерение результата «чаще всего осуществляется с помощью BLEU» [6, с. 186]. При этом автор подчёркивает, что сопоставимость и воспроизводимость результатов зависят от единообразия процедуры подсчёта метрики. Вместе с тем, как показывает М.Попович, метрика chrF является особенно перспективной для языков с богатой морфологией [7, с. 499]. Поэтому расхождения между BLEU и chrF, зафиксированные для агглютинативного узбекского языка, подтверждают целесообразность использования комплементарных метрик в оценке качества машинного перевода.

Приведём несколько характерных примеров, иллюстрирующих поведение моделей до и после дообучения. Исходное предложение на узбекском языке и его эталонный перевод на русский даны первыми, за ними следуют варианты каждой модели.

Пример 1. Исходный текст (uz): «O‘zbekiston Respublikasi Prezidenti yangi ta’lim islohotlarini e’lon qildi.» Эталонный перевод (ru): «Президент Республики Узбекистан объявил о новых реформах в сфере образования.» NLLB-200 (baseline): «Президент Республики Узбекистана объявил о новой образовательной реформе.» NLLB-200 (FT): «Президент Республики Узбекистан объявил о новых реформах в образовании.» M2M-100 (baseline): «Президент Узбекской Республики объявил новую реформу образования.» M2M-100 (FT): «Президент Республики Узбекистан заявил о новых реформах в сфере образования.» mBART-50 (baseline): «Президент Узбекистана объявил реформы нового образования.» mBART-50 (FT): «Президент Республики Узбекистан объявил о новых образовательных реформах.» В этом примере все модели после дообучения корректно передали ключевую информацию. Однако до fine-tuning mBART-50 допустила нарушение порядка определений

(«реформы нового образования» вместо «новые реформы образования»), а M2M-100 использовала нестандартное наименование «Узбекская Республика».

Пример 2. Исходный текст (uz): «Bu kitob bolalar uchun juda qiziqarli hikoyalarni o‘z ichiga oladi.» Эталонный перевод (ru): «Эта книга содержит очень интересные рассказы для детей.» NLLB-200 (FT): «Эта книга содержит очень интересные рассказы для детей.» M2M-100 (FT): «Эта книга включает в себя очень интересные истории для детей.» mBART-50 (FT): «Данная книга содержит интересные для детей истории.» OPUS-MT (FT): «В этой книге есть интересные рассказы для детей.» SeamlessM4T (FT): «Эта книга содержит очень интересные истории для детей.» NLLB-200 дала перевод, практически совпадающий с эталоном. Модель mBART-50, хотя и передала смысл верно, утратила усилительное наречие «очень» (juda) и переставила определение «для детей» ближе к существительному «истории», что характерно для англоцентричных моделей, стремящихся к порядку слов, типичному для английского языка.

Пример 3. Исходный текст (uz): «Paxta terimi mavsumi boshlanishi bilan dehqonlar dalaga chiqishdi.» Эталонный перевод (ru): «С началом сезона сбора хлопка дехкане вышли в поле.» NLLB-200 (FT): «С началом сезона сбора хлопка фермеры вышли в поле.» M2M-100 (FT): «Когда начался сезон хлопка, фермеры вышли на поля.» mBART-50 (FT): «С начала сезона хлопка фермеры пошли в поле.» OPUS-MT (FT): «Фермеры вышли в поле с началом сезона хлопка.» SeamlessM4T (FT): «С наступлением сезона сбора хлопка дехкане вышли в поле.» Этот пример особенно показателен. Слово «dehqon» (дехканин, традиционное наименование крестьянина в Центральной Азии) верно передали лишь SeamlessM4T и частично эталон. Все прочие модели использовали нейтральное «фермеры», что свидетельствует о недостаточной представленности культурно-маркированной лексики в обучающих данных.

Пример 4. Исходный текст (uz): «Ushbu qonun fuqarolarning huquqlari va erkinliklarini kafolatlash maqsadida qabul qilingan.» Эталонный перевод (ru): «Данный закон принят в целях обеспечения гарантий прав и свобод граждан.» NLLB-200 (FT): «Данный закон был принят в целях гарантирования прав и свобод граждан.» M2M-100 (FT): «Этот закон был принят для обеспечения прав и свобод граждан.» mBART-50 (FT): «Данный закон принят с целью гарантирования прав и свобод граждан.» OPUS-MT (FT): «Закон принят для защиты прав и свобод граждан.» Юридическая лексика, как видно, передаётся всеми моделями с приемлемой точностью. Различия проявляются в синтаксическом оформлении (пассивная конструкция «был принят» у NLLB-200 и M2M-100 против безличной «принят» у mBART-50) и в лексическом выборе (OPUS-MT подменила «гарантирование» на «защиту», сузив семантику).

Пример 5. Исходный текст (uz): «Mening yoshligim Buxoroning tor ko‘chalarida o‘tdi.» Эталонный перевод (ru): «Моё детство прошло в узких улочках Бухары.» NLLB-200 (FT): «Моя юность прошла на узких улицах Бухары.» M2M-100 (FT): «Моя молодость прошла в тесных улочках Бухары.» mBART-50 (FT): «Мое детство прошло в узких улочках Бухары.» SeamlessM4T (FT): «Моя юность прошла в узких улицах Бухары.» Слово «yoshlik» допускает перевод и как «юность», и как «детство», и как «молодость». Выбор зависит от контекста, однако ни одна метрика не способна в полной мере уловить эту полисемию. Модели разошлись в интерпретации, что лишний раз подчёркивает ограниченность автоматических метрик для оценки качества перевода в условиях лексической неоднозначности.

Пример 6. Исходный текст (uz): «Samarqand shahridagi Registon maydoni dunyo miqyosida mashhur me’moriy yodgorlikdir.» Эталонный перевод (ru): «Площадь Регистан в городе Самарканде является всемирно известным архитектурным памятником.» NLLB-200 (FT): «Площадь Регистан в Самарканде является всемирно известным архитектурным памятником.» M2M-100 (FT): «Площадь Регистан в городе Самарканд является знаменитым архитектурным памятником мирового масштаба.» OPUS-MT (FT): «Площадь Регистан в Самарканде является известным мировым памятником архитектуры.» Перевод географических названий и культурных реалий оказался адекватным у всех моделей, что может объясняться достаточной представленностью подобных наименований в обучающих корпусах. Тем не менее M2M-100 использовала развёрнутую конструкцию «мирового масштаба» вместо более лаконичного «всемирно известным», что делает перевод менее стилистически выдержанным.

Таблица 4. Средние оценки экспертов (адекватность и беглость, шкала от 1 до 5)

Модель	Адекватность (base)	Адекватность (FT)	Беглость (base)	Беглость (FT)
NLLB-200-3.3B	3,2	4,3	3,0	4,1
M2M-100-1.2B	2,7	3,8	2,5	3,6
mBART-50-MM	2,9	4,0	2,8	3,8
OPUS-MT	2,3	3,3	2,1	3,1
Google NMT	3,6	н/д	3,5	н/д
SeamlessM4T v2	3,4	4,2	3,3	4,0

Экспертные оценки в целом согласуются с автоматическими метриками. NLLB-200 после дообучения получила наивысшие средние баллы по обоим критериям (4,3 за адекватность и 4,1 за беглость). Примечательно, что SeamlessM4T, уступая NLLB-200 по BLEU на 2,4 пункта, получила практически сопоставимые экспертные оценки (4,2 и 4,0 соответственно), что может указывать на более высокую «естественность» генерируемого текста, не всегда улавливаемую n-граммными метриками.

Обсуждение. Полученные результаты позволяют сделать несколько наблюдений. Прежде всего, доминирование NLLB-200 не случайно и объясняется

сочетанием трёх факторов. Первый фактор состоит в масштабе предобучения (3,3 млрд параметров против 1,2 млрд у M2M-100 и 611 млн у mBART-50). Второй заключается в не-англоцентричном характере обучающих данных, что критически важно для пары, не включающей английский язык. Третий связан с использованием специализированного токенизатора SPM-200, более адекватно обрабатывающего латинизированный узбекский текст. Любопытным оказалось поведение OPUS-MT. Эта модель, обладая наименьшим числом параметров (~74 млн), показала самый скромный результат. Однако относительный прирост BLEU после дообучения (+98%) у OPUS-MT оказался наиболее впечатляющим среди всех моделей. Данный парадокс объясняется тем, что исходная модель OPUS-MT для пары uz-ru обучена на крайне ограниченном объёме данных из репозитория OPUS, и любой дополнительный параллельный материал оказывает на неё непропорционально сильное влияние. Google NMT, не подлежащая дообучению, продемонстрировала конкурентоспособный baseline (BLEU 22,1). Вместе с тем после fine-tuning четыре из пяти открытых моделей обошли Google по автоматическим метрикам.

Отдельного внимания заслуживает анализ ошибок. На основании экспертного разбора 200 предложений были выделены типичные категории ошибок для каждой модели. У NLLB-200 после дообучения преобладали лексические несоответствия (32% ошибок), связанные преимущественно с культурно-маркированной лексикой и редкими терминами. Грамматические ошибки составили 28%, среди них наиболее частыми были нарушения управления и согласования в роде. Ошибки передачи смысла (пропуски и искажения) зафиксированы в 18% случаев, стилистические несоответствия составили 22%. У M2M-100 картина ошибок оказалась иной. Грамматические нарушения доминировали (41%), что может быть связано с менее детализированным представлением русской морфологии в модели. Лексические ошибки составили 25%, смысловые пропуски встречались в 20% случаев, стилистические отклонения достигли 14%.

Заключение. Проведённое сопоставительное исследование шести моделей нейронного машинного перевода выявило значительный потенциал дообучения для узбекско-русской языковой пары. Модель NLLB-200-3.3B достигла наивысшего качества перевода (BLEU 31,6, chrF 58,3), превзойдя как собственный baseline, так и проприетарную систему Google NMT. SeamlessM4T v2 продемонстрировала наиболее естественный стиль перевода при близких к лидеру количественных показателях. Модель mBART-50, несмотря на англоцентричное смещение, показала третий результат и может рассматриваться как компромиссный вариант при ограниченных ресурсах. Качественный анализ ошибок обнаружил устойчивые закономерности, связанные с архитектурными особенностями каждой модели. Англоцентричные модели склонны к калькированию синтаксических конструкций английского языка. Модели с малым числом параметров допускают больше

грамматических ошибок. Культурно-маркированная лексика остаётся проблемной зоной для всех без исключения систем.

Список использованной литературы:

1. Хусаинова А. М., Романов В. А., Хан А. М. Многоязычный машинный перевод с помощью иерархического трансформера // Вестник ВГУ. Серия: Системный анализ и информационные технологии. – 2022. № 1. – С. 125-138.
2. Wang R., Utiyama M., Sumita E. A Survey on Low-Resource Neural Machine Translation // Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21). – 2021. – P. 4636-4643.
3. Signoroni R., Rychlý P. Efficient Architectures For Low-Resource Machine Translation // Proceedings of the 8th Workshop on Technologies for Machine Translation of Low-Resource Languages. – 2025. – P. 39-64.
4. https://huggingface.co/docs/transformers/en/model_doc/nllb
5. Негматулоев О. О., Жорник Д. О., Мельников А. В. Разработка модели нейронного машинного перевода для мансийского языка // Системная инженерия и информационные технологии. – 2025. Т. 7, № 2(21). – С. 30-47
6. Post M. A Call for Clarity in Reporting BLEU Scores // Proceedings of the Third Conference on Machine Translation: Research Papers. Brussels, Belgium, – 2018. – P. 186-191.
7. Popović M. chrF deconstructed: beta parameters and n-gram weights // Proceedings of the First Conference on Machine Translation. Berlin, Germany, – 2016. – P. 499-504.