МОЖЕТ ЛИ МАШИННЫЙ ПЕРЕВОД ПРИВЕСТИ К ИСЧЕЗНОВЕНИЮ РЕДКИХ ЯЗЫКОВ?
“ЦИФРОВАЯ ДИСКРИМИНАЦИЯ ИЛИ ИНСТРУМЕНТ СПАСЕНИЯ?”

Омонхонов Элдорхон

Андижанский государственный институт иностранных языков

E-mail: Xon01022004@gmail.com

Преподаватель: Эгамова Н. Т.

E-mail: egamovanilufar82@gmail.com

Аннотация: Работа посвящена исследованию влияния систем нейросетевого машинного перевода (NMT) на языковое разнообразие и судьбу малоресурсных и редких языков в цифровую эпоху. Рассматриваются технологический, социолингвистический, экономический и этический аспекты проблемы. Обосновывается тезис о том, что машинный перевод не является прямой причиной вымирания языков, однако отсутствие цифровой инфраструктуры и качественных корпусов данных для редких языков усугубляет их маргинализацию, в то время как целенаправленный языковой активизм и специализированные корпусные технологии могут превратить ИИ в инструмент их документирования и сохранения.

Ключевые слова: машинный перевод, редкие языки, языковая экология, цифровая дискриминация, малоресурсные языки, корпусная лингвистика, языковой сдвиг.

Введение

В условиях глобализации и стремительного развития технологий искусственного интеллекта (системы Neural Machine Translation — NMT) машинный перевод занял центральное место в межкультурной и международной коммуникации. Однако за впечатляющими успехами перевода между языками-гигантами (английский, китайский, испанский, русский) скрывается глубокая проблема языкового неравенства. Возникает закономерный и острый вопрос: способен ли машинный перевод ускорить исчезновение редких и малоресурсных языков, или же он может стать средством их возрождения? Общий обзор социолингвистических проблем

языкового контакта и глобализации представлен в [1].

Цель исследования: определить влияние современных технологий машинного перевода на языковую экологию и выявить риски цифровой ассимиляции редких языков.

Основной тезис: Сам по себе алгоритм машинного перевода не уничтожает языки, однако инфраструктурное неравенство и коммерческая ориентация ИИ на языки с большими ресурсами создают серьезные угрозы для выживания малочисленных этнолектов, ускоряя языковой сдвиг.

Современные нейросетевые системы перевода функционируют на основе огромных массивов данных (Big Data) — параллельных корпусов текстов. Для редких языков и языков коренных малочисленных народов такие корпусы практически отсутствуют. Возникает феномен цифровой дискриминации (Digital Exclusion):

Поисковые системы, операционные системы, образовательные платформы и интерфейсы не поддерживают редкие языки, вынуждая их носителей переходить на языки межэтнического общения (лингва франка), что стимулирует процесс языкового сдвига (language shift) [2].

Даже при попытке создания моделей для языков с низким объемом ресурсов (Low-Resource NMT) нехватка размеченных данных приводит к серьезным лексическим и синтаксическим искажениям, навязывающим чужеродные кальки и стирая уникальные грамматические структуры автохтонных языков [3].

С точки зрения социолингвистики, главным условием сохранения языка является его активная межпоколенная передача (intergenerational transmission) и расширение сферы функционирования. Машинный перевод влияет на эти факторы двойственно:

С одной стороны, мгновенный доступ к качественному переводу глобального контента на доминирующие языки снижает у молодежи практическую мотивацию изучать и использовать редкие родительские диалекты. Зачем сохранять или учить редкий язык, если алгоритм позволяет общаться и потреблять контент на глобальном языке? Это разрушает естественную среду бытования языка [4].

С другой стороны, исследователи отмечают потенциал языкового активизма:

энтузиасты используют технологии машинного обучения для создания цифровых словарей, автоматической транскрипции фольклорных аудиозаписей и перевода учебных материалов, что повышает престиж языка в цифровом пространстве [5].

Для наглядности приведем сопоставительный анализ того, как технологическое игнорирование контекста и прагматики отражается на передаче культурно-специфической информации в уязвимых языковых системах:

Категория материала	Риск машинной обработки	Адекватное решение / Подход
Идиоматика и фольклор	Буквальный кальковский перевод, утрата образности и культурного кода	Привлечение носителей для создания качественных глоссариев и датасетов
Лексика коренных ремесел	Подмена специфических терминов общими гиперонимами	Фиксация автохтонных терминов через корпусные исследования
Устная речь диалектов	Ошибки распознавания акустических моделей (Speech-to-Text)	Разработка адаптированных акустических моделей для конкретного диалекта



Коммерческие разработчики ИИ инвестируют прежде всего в высокодоходные языковые пары. Редкие языки не приносят коммерческой прибыли, из-за чего остаются вне зоны внимания технологических корпораций. Это порождает проблему лингвистической справедливости (linguistic justice). Ответственность за сохранение языкового многообразия в эпоху ИИ ложится не на автоматические алгоритмы, а на государственные институты, академическое сообщество и языковых активистов, объединяющих усилия по документированию исчезающих языков.

Заключение

На вопрос «может ли машинный перевод привести к исчезновению редких языков?» следует ответить следующим образом: прямой причинно-следственной связи здесь нет, так как вымирание языков обусловлено социально-экономическими и историческими факторами. Тем не менее, цифровое игнорирование и отсутствие редких языков в экосистеме машинного перевода косвенно ускоряют их ассимиляцию. Выживание редких языков в XXI веке зависит от того, смогут ли лингвисты и разработчики интегрировать технологии малоресурсного перевода в программы сохранения культурного наследия.

Список использованных источников

1. Бацевич Ф. С. Основы коммуникативной лингвистики. М.: Академия, 2009.

2. Климкин И. И. Цифровая социолингвистика и судьба малочисленных языков. СПб.: Лань, 2021.

3. Koehn P. Statistical Machine Translation. Cambridge University Press, 2010.

4. Crystal D. Language Death. Cambridge University Press, 2000.

5. Austin P. K., Sallabank J. Endangered Languages: Issues and Methods. Cambridge University Press, 2011.

