

УДК: 630\*722004.981'322.4

# Состояние, тенденции и перспективы развития систем машинного перевода

В публикации рассматривается состояние дел, современные тенденции и перспективы развития систем машинного перевода. Автором обозначены три уровня выполнения переводов на мировом рынке переводов. Проанализированы вопросы эффективности систем машинного перевода и использования машинного обучения для построения цифровой платформы в контексте новой экономики и технологичности процесса перевода. Обозначены направления по повышению качества перевода разных уровней сложности на рынке переводов как в Республике Беларусь, так и на мировых рынках.

**Т.И. МАКАРЕВИЧ,**

магистр управления, магистр филологических наук,  
зам. зав. кафедрой англ. языка  
гуманитарных специальностей (АЯГС)  
ФМО БГУ по НИР

Белорусский государственный университет

**Ключевые слова:**

*цифровая платформа перевода, международный рынок переводов, экономика перевода, поставщики лингвистических услуг, технологичность процесса перевода, искусственный интеллект, машинный перевод, машинное обучение, нейронные сети, новая экономика.*

**Введение.** В последнее время в мире возросла потребность перехода оказания услуг из режима offline в online, повысились требования к системам машинного перевода (МП), в том числе в связи с пандемией, начавшейся в 2020 г. Современный рынок переводческих услуг огромен в мировом измерении и характеризуется как динамично развивающийся, высококонкурентный, с большим запросом по развитию программного обеспечения (ПО) для автоматизации переводческой деятельности и очень неоднородный для заказчиков переводческих услуг и их потребителей.

Темп жизни ускоряется и трансформируется в новые формы взаимодействия, возникают новые тренды в мировой экономике, меняется и традиционная отрасль перевода, которая должна оформиться в эффективно развивающуюся экономику перевода. В этой области сегодня профессия переводчика не защищена, так как в экономике перевода присутствуют не только профессиональные переводчики, но и те участники рынка, для кого перевод является второстепенным заработком без подтверждения соответствующего квалификационного диплома.

На современном рынке переводов переводчиков крупных переводческих агентств и бюро переводов сегодня именуют «поставщиками лингвистических услуг», что звучит широко и многообещающе для развития мирового рынка переводов и систем МП для осуществления разных видов переводческой деятельности.

Актуальность научной работы заключается в недостаточной изученности экономики перевода как таковой в Республике Беларусь, развитии институциональной основы индустрии перевода в эпоху глобальной цифровой трансформации путем формирования цифровой экосистемы индустрии перевода.

Одна из современных тенденций развития систем МП – это трансформация рынка переводческих услуг в области экономики перевода для эффективного ведения хозяйственной деятельности. В этом сегменте есть запрос на установление рациональных правил поведения для всех участников индустрии перевода, в том числе и сегмента развития ПО процесса перевода, спрос на которое не изучен, поэтому его появление на рынке не представляется системным и урегулированным.

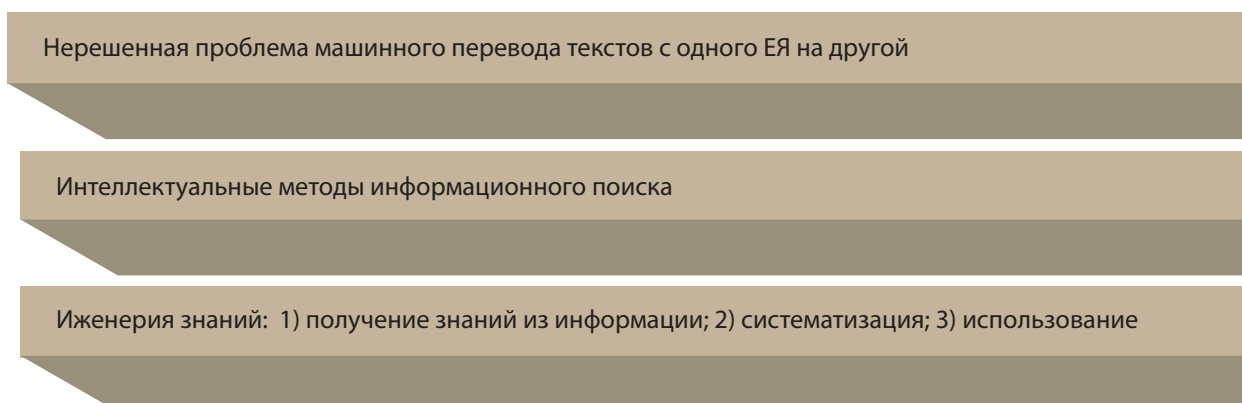


Рисунок 1 – МП и проблемы непереводимости текстовой информации

Необходимым этапом развития рынка переводов является создание механизма эффективной цифровой инфраструктуры. Таким образом, одна из задач работы – разработка концепции универсальной цифровой платформы переводческих услуг, ведь чем быстрее и качественнее осуществляется перевод в различных профессиональных отраслях, тем быстрее проходят международные сделки [1]. Поиск оптимального решения создания цифровой платформы актуален для профессиональных переводчиков-практиков, работающих в переводческих компаниях и бюро переводов, представителей вузов, ведущих подготовку переводчиков, и структур регулирования взаимодействия участников переводческой отрасли для оптимизации работы всей индустрии перевода, которая на настоящий момент не выделена как отдельный вид экономической деятельности.

**МП и машинное обучение в контексте построения цифровой платформы.** Технология машинного обучения (МО) – многосложная и многогранная – находится в основе искусственного интеллекта (ИИ) [2, с. 144].

МП может быть применен только к переводимым текстам, то есть к тем, у которых есть возможность быть переведенными на другой язык. В связи с этим встает серьезный вопрос о переводимости – непереводимости текста, который на данном этапе развития МП не способен решить. Проблема непереводимости сегодня во многих случаях устраняется традиционным путем: умением переводчика пользоваться различными приемами перевода, как показано на рис. 1.

Перед МП по теории несоответствий стоит задача максимальной приближенности к адекватному переводу переданной информации [3, с. 119], вычлняемой в переводе при сравнительном анализе текстов. Переданная информация представляет собой сведения, содержащиеся в исходном тексте и отсутствующие в тексте перевода.

Возвращаясь от искусства перевода к ИКТ-разработчикам МП, важно различать традиционный не-МО алгоритм (жестко запрограммированный, идентифицирующий естественный язык (ЕЯ), грамматические структуры и т. п.) и МО-алгоритм (без жесткого программирования, имеет нелинейный характер, учится на

примерах, предназначен для решения целого ряда конкретных практических задач, таких как перевод с одного ЕЯ на другой). По сути, МО алгоритм – это нейронная сеть, которая проходит тренировку или «обучение» на образцах данных, в результате чего появляется натренированная модель. Полагаем, что в будущем проблемы непереводимости текстовой информации будут решать посредством нейронных сетей.



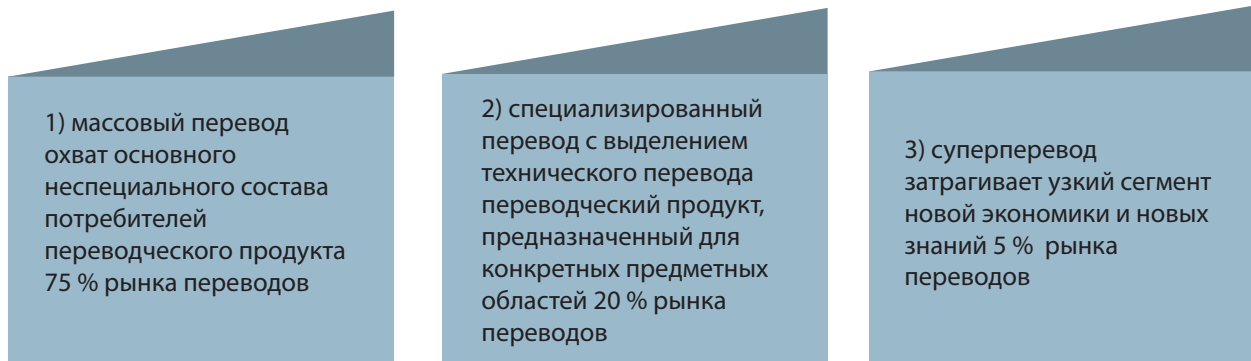


Рисунок 2 – Построение трехуровневой платформы перевода

**Создание цифровой трехуровневой платформы перевода.** В предыдущих публикациях мы обозначили три значимые сферы перевода по охвату аудитории (заказчиков) [4, с. 88] и типам текстов оригинала перевода, которые продемонстрированы на рис. 2.

Обозначенные уровни перевода выделены для действенного и эффективного поиска переводческих инструментов на рынке переводов с целью адекватной и быстрой передачи следующих типов текстов оригинала на язык перевода, что важно не только для переводческого бизнеса, но и для экономики перевода в целом. Сегодня МП не всегда способен справиться с текстами оригинала перевода любой сложности, как показано в табл. 1.

Из таблицы следует, что переводческие инструменты избираются целенаправленно для сопровождения текстов определенного типа. Наглядно показано, в каком направлении разработчикам МП стоит стремиться в создании новых продуктов для индустрии перевода.

Характеризуя уровни цифровой трехуровневой платформы перевода, обозначим *уровень 1 (массовый перевод)* как стремление людей автоматизировать процесс перевода при активном использовании различных ИКТ для жизни человека в информационном обществе. Для развития мировой экономики массовый перевод представляет значительную серьезную нишу в плане увеличения доходов, которая на данный момент не имеет четкого структурного регулирования и, соответственно, массовый перевод обусловлен преимуществами в использовании

МП: 1) скорость: перевод слова или текста можно получить в течение нескольких секунд; 2) доступность услуги: большинство онлайн-переводчиков бесплатны, и единственным условием доступа является наличие интернета.

*Уровень 2 (специализированный перевод)* – это перевод текстов разных предметных областей с присущей ему терминологической системой и особой номенклатурой документов. Отличительной чертой этого вида перевода, как и другого специального перевода, является «предельно точное и четкое изложение материала при почти полном отсутствии образно-эмоциональных выразительных средств, присущих художественному и общественно-политическому текстам» [5, с. 5].

К *уровню 3* относится суперперевод, который включает в себя наиболее сложные и непосильные для МП устные и письменные тексты новых отраслей знаний, речей политиков, ученых, дипломатов, ценных художественных произведений, осуществляемых профессиональными переводчиками с использованием автоматизации при устном, последовательном и письменном видах перевода [6, с. 58].

При суперпереводке наиболее актуальна в настоящее время проблемы с недостатками МП. Выделим некоторые наиболее очевидные: 1) во многих онлайн-переводчиках слабо учитываются грамматические правила, художественные языковые приемы; 2) большое количество ошибок вызвано неправильными вариантами перевода; онлайн-переводчики редко учитывают особенности сочетаемости и употребления слов; много ошибок допускается при

Таблица 1 – Типы текстов оригинала перевода предметных областей

Типы текстов оригинала	МП общий	МП специальный (технический)	МП с участием человека	Перевод с использованием компьютера	Традиционный перевод, осуществляемый переводчиком
Высокотворческий	Не подходит	Не подходит	Не подходит	Не подходит	Подходит
С творческими элементами	Не подходит	Не подходит	Не подходит	Не подходит	Подходит
Общего характера	Подходит	Не подходит	Возможно	Подходит	Подходит
Специализированный	Возможно	Подходит	Подходит	Подходит	Подходит
Чисто технический	Подходит	Подходит	Подходит	Подходит	Подходит

Таблица 2 – Системы МП с различной степенью автоматизации

Степень автоматизации МП	Применение МП к уровням перевода
Полностью автоматический перевод	Уровень 1
Перевод, осуществляемый человеком с использованием компьютера	Уровень 2 Уровень 3
Автоматизированный МП при участии человека	Уровень 1 Уровень 2 (технический перевод)

переводе многозначных слов и омонимов; 3) выбранное переводчиком слово может не подходить по контексту и стилевой принадлежности; 4) особое значение при МП имеет перевод специализированной терминологии [6, с. 16] и, если в словарной базе отсутствует определенный термин или слово, оно вообще не переводится или толкуется с искажением его изначального значения.

Приведенные аргументы свидетельствуют об актуальности создания цифровой трехуровневой платформы для адекватности процесса МП по трем обозначенным уровням перевода.

**Классификация автоматизированных систем МП.** Рассмотрение классификации автоматизированных систем МП необходимо для создания цифровой трехуровневой платформы, как показано в табл. 2.

1. Полностью автоматический перевод. Пока в профессиональном переводе на него нельзя полагаться, так как текст, полученный в результате автоматического перевода, содержит достаточно много ошибок, поэтому совершенствование данного вида перевода является задачей будущего.

2. Перевод, осуществляемый человеком с использованием компьютера. При данной методике центральное место в процессе перевода принадлежит человеку, а разнообразное ПО служит только инструментом, чтобы процесс перевода становился более быстрым.

3. Автоматизированный МП при участии человека. При таком виде МП идет обработка текста, а также редакция текста до осуществления перевода компьютером. После человек вновь корректирует текст для получения наиболее адекватного, верного перевода. Поддержание баз данных слов и терминов в рабочем состоянии – процесс трудоемкий, кроме того, специалист, осуществляющий перевод, должен быть обучен корректуре предварительных и переведенных текстов.

Мы также анализируем классификацию МП в зависимости от используемых в нем технологий, в которой выделяют следующие модели МП, как показано в табл. 3.

Полагаем, что в направлении создания систем МП для цифровой трехуровневой платформы наибольший интерес представляет гибридная система, которая сможет объединить три разноплановых уровня перевода.

Таблица 3 – Модели МП

Название модели	Описание функционирования модели
Статистический перевод	Предложение, которое нужно перевести, в данной модели разделяется на отдельные слова и фразы, затем машина перебирает все варианты перевода для каждого фрагмента и оценивает вероятность каждого существующего варианта
Нейросетевой перевод	Нейронная сеть проводит анализ параллельных текстов, обнаруживая в них закономерности и при переводе она работает с более крупными единицами – предложениями. Данный подход позволяет учитывать смысловые связи внутри предложения и контекст, поэтому перевод получается более адекватным, так как проходит через модель языка, однако полностью избежать ошибок не удается
Гибридная система	Объединяет статистическую и нейросетевую модели, что означает, что текст, вводимый для перевода, обрабатывается и нейронной сетью, и статистическим переводчиком. Затем алгоритм выделяет лучший перевод и показывает его пользователю

Для экономики перевода при создании цифровой трехуровневой платформы перевода важным с точки зрения пользователя являются типы систем МП по эксплуатационным параметрам, среди которых выделяют следующие типы [2, с. 18]: 1) домашний тип систем МП; 2) онлайн-тип систем МП; 3) профессиональный тип систем МП; 4) корпоративный тип систем МП. Данные типы систем МП по эксплуатационным параметрам учтены при разработке концепции цифровой трехуровневой платформы перевода.

При решении проблемы точности и адекватности перевода, в частности, терминологии различных предметных областей – уровень перевода 2 и 3, необходимо провести самую трудозатратную работу – осуществить разметку данных чтобы достоверно и безошибочно разметить тысячи единиц терминов для их последующего правильного перевода. Для решения этой задачи создаются специализированные сервисы разметки данных для МО. Данные сервисы работают по принципу нанесения разметчиками из разных уголков мира разметок данных, для подготовки обучающей выборки для ИИ, при котором алгоритм, переняв знания людей, успешно может решать аналогичные задачи. При этом отбор данных имеет колоссальное значение: данные, на которых будет обучаться нейросеть, надо готовить. По мнению А. Лосева, ИИ пока не создан [7], – это пока лишь отдельные методы и технологии МО. Если данные неполные, неправильно подготовлены, то на выходе нейросеть даст абсолютно непредсказуемый результат [7].

При создании функциональной цифровой платформы перевода как эффективного направления развития экономики перевода важно понимать, что

на основе перевода мы приходим к новому знанию. Экономическая выгода от осуществления процесса перевода, возможно, будет не столь заметна на первый взгляд, и это тормозит процесс инвестирования в развитие переводческой отрасли, которая остается не выделенной как самостоятельный вид экономической деятельности. Тем не менее важно замечать не прямые экономические выгоды и общую всеобъемлемость экономики перевода, которая, будучи отдельным видом профессиональной деятельности, способна оказать мощное влияние на процессы во всех других отраслях мировой экономики. Перевод – это часть мировой экономической системы и может принести существенную прибыль, создать репутацию страны интеллектуально развитого информационного общества.

Разработка систем МП и стимулирование развития цифровых платформ собственной разработки в Республике Беларусь будет служить увеличению позиционирования Беларуси как High-Tech

Nation – высокотехнологичной страны. Данный термин был засвидетельствован как название одноименного международного форума в ноябре 2019 г. в г. Минске, на котором было заявлено о раскрытии и использовании экономического, промышленного, научно-исследовательского и делового потенциала страны для увеличения макроэкономического эффекта различных отраслей экономики нации. Таким образом, развитие систем МП и участие в нем белорусских разработчиков ПО, профессиональных переводчиков и лингвистов как поставщиков лингвистических услуг для экономики переводов Беларуси внесет существенный вклад в получение дополнительной прибыли от роста спроса на качественный продукт МП, подготовки специалистов в области разработки систем МП, подготовки профессиональных переводчиков, организации переводческих форумов, разработке качественных программ повышения квалификации в области перевода и совершенствования систем МП.

#### ЛИТЕРАТУРА

1. Панышин, Б.Н. Интеллектуализация деловых услуг как основной фактор формирования новой экономики / Б. Н. Панышин // Наука и инновации. – № 9 (139). – 2014. – С. 41–54.
2. Кво, Ч.К. Технологии перевода: учеб. пособие для студ. высш. учеб. заведений / Ч. К. Кво; [пер. с англ. и науч. ред. рус. текста А.Л. Семенова]. – Москва, Изд. центр «Академия», 2008. – 256 с.
3. Нелюбин, Л.Л. Введение в технику перевода: (когнитивный теоретико-прагматический аспект): курс лекций / Л.Л. Нелюбин. – Москва: Изд-во МГОУ, 2005. – 152 с.
4. Макаревич, Т.И., Макаревич, И.И. Экономика и управление на рынке перевода: международный аспект [Электронный сборник] // Вестник университета им. С.Ю. Витте. Серия 1: Экономика и управление»: электрон. науч. ж-л / Т. И. Макаревич, И.И. Макаревич //– Москва, 2020. – С. 85–93.
5. Нелюбин, Л.Л. Толковый переводческий словарь / Л. Л. Нелюбин. – 4-е изд., испр. – Москва, Наука, 2006. – 320 с.
6. Макаревич, И.И. Техничко-экономические исследования методических положений и состава ИКТ при создании словаря-справочника по цифровой трансформации / И.И. Макаревич // Практика преподавания иностранных языков на факультете международных отношений БГУ: электронный сборник. – 2020. – Вып. 10. – Минск, БГУ. – С. 12–15.
7. Лосев, А.В. Искусственным интеллектом должны заниматься давшие присягу / А.В. Лосев. [Электронный ресурс]. – Режим доступа: [https://octagon.media/istorii/aleksandr\\_losev\\_iskusstvennym\\_intellektom\\_dolzny\\_zanimatsya\\_davshie\\_prisyagu\\_.html](https://octagon.media/istorii/aleksandr_losev_iskusstvennym_intellektom_dolzny_zanimatsya_davshie_prisyagu_.html). Дата доступа: 29.01.21.
8. Макаревич, Т.И. Автоматизация труда переводчиков и развитие технологий перевода на международном рынке переводов: состояние и тенденции развития / Т. И. Макаревич. – Цифровая трансформация. – № 4. – Минск, 2020. – С. 57–67.
9. Патак, Нишит. Искусственный интеллект для .Net: речь, язык и поиск / Нишит Патак. – Москва, ДМК Пресс, 2018. – 298 с.

*The given paper considers contemporary state, tendencies and perspectives for machine translation systems development. The author has indicated tree levels for translators and interpreters' activity on international translation market. The author has analysed effectiveness issues of machine translation and application of machine learning in the framework of new economy and technological efficiency of translation process. It suggests some tendencies in streamlining translation and interpreting of any kind of difficulty both in the Republic of Belarus and on the world markets.*

*Key words:* digital platform, international translation market, translation economy, language services producers, technological effectiveness of translation process, artificial intelligence, machine translation, machine learning, neural networks, deep learning, new economy.

Получено 08.02.2021.