

УДК 81`322.4

Бусел Татьяна Викторовна

кандидат филологических наук, доцент

Минский государственный

лингвистический университет

г. Минск, Беларусь

Tatyana Busel

PhD in Philology, Assistant Professor

MSLU

Minsk, Belarus

tatsiana-busel@yandex.ru

400

ИСКУССТВО ПЕРЕВОДА В ЭПОХУ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ТЕНДЕНЦИИ И ИННОВАЦИИ

Статья посвящена современным тенденциям и технологическим инновациям в сфере нейронного машинного перевода (НМП) и их влиянию на переводческую деятельность. Актуальность изучения и решения связанных с НМП проблем, и важность его практического применения в преодолении языкового барьера обусловлена значительным увеличением объема переводимого контента, нехваткой профессиональных переводчиков для многочисленных специализированных областей и отдельных языковых пар. В статье раскрывается суть работы современных систем МП, использующих технологии искусственного интеллекта, представлены модели обработки естественного языка в МП, описаны принципы работы систем МП на основе нейронных сетей, а также определены лингвистические особенности, возможности и ограничения использования таких систем в процессе перевода.

К л ю ч е в ы е с л о в а: машинный перевод; искусственный интеллект; нейронная сеть; обработка информации на естественном языке; машинное обучение; большая языковая модель.

THE ART OF TRANSLATION IN THE ERA OF ARTIFICIAL INTELLIGENCE: TENDENCIES AND INNOVATIONS

The article is devoted to recent innovations in neural machine translation (NMT) that has become the new mainstream method in practical MT systems. Today it is impossible to overestimate the significance and prospects of MT due to the rising demand for translations that is intensified by an existing lack of translators for numerous specialized fields and particular language combinations. The author analyzes the possibilities of neural machine translation, shows the models of natural language processing in machine translation and reveals the essence of artificial intelligence based MT systems. NMT impact on the translation industry is undeniable, offering both challenges and opportunities.

Key words: machine translation; artificial intelligence; neural network; natural language processing; machine learning; a large language model.

Вступление общества в современную информационную эпоху определило возможность внедрения инновационных технологий во все сферы человеческой деятельности и сделало актуальным их использование для повышения эффективности труда переводчиков. До наступления эпохи искусственного интеллекта понятие «переводчик» ассоциировалось исключительно с человеком. Перевод художественных и аудиовизуальных произведений, научных публикаций, деловых и юридических документов выполняли профессиональные переводчики. Однако появление технологии искусственного интеллекта оказало огромное влияние на переводческую деятельность. В наши дни «перевод – это искусство, основанное на науке, развивающей технологии» [1, с. 3]. Создание искусственного интеллекта ознаменовало переход общества к новым вызовам, заставляющим задуматься о будущем искусства перевода.

Создание искусственных нейронных сетей позволило эффективно моделировать процессы понимания и генерации текстов на естественном языке, которые традиционно относятся к классу проблем искусственного

интеллекта. Нейронный машинный перевод очень востребован в современном обществе, ежедневно им пользуются миллионы человек. Данные, опубликованные на сайте IMARCGroup (International Market Analysis Research and Consulting Group) [2] свидетельствуют о том, что росту мирового рынка машинного перевода способствует целый ряд факторов:

1) значительное увеличение спроса на переводческие услуги, поскольку они играют ключевую роль в преодолении культурных и языковых барьеров, а также в повышении эффективности работы международных компаний и организаций;

2) растущий объем различных видов цифрового контента: текстового, игрового, видео, аудио, графического и пользовательского;

3) совершенствование технологий машинного обучения (так называемого «искусственного интеллекта») и обработки информации на естественном языке (Natural Language Processing – NLP);

4) все большее распространение облачных услуг машинного перевода (DeepL, Google Cloud Translation API, Alibaba Cloud Machine Translation, IBM Watson Language Translator, SDL Language Cloud Translation Management).

Развитие направления машинного перевода представляет большой научный интерес, что обусловлено комплексностью и сложностью компьютерного моделирования данного процесса. Как вид языковой деятельности перевод затрагивает все уровни языка – от распознавания графем (и фонем при переводе устной речи) до передачи смысла высказывания и текста. Кроме того, для перевода характерна обратная связь и возможность сразу проверить гипотезу об устройстве тех или иных языковых уровней и эффективности предлагаемых алгоритмов. Эта специфическая черта перевода привлекает внимание отечественных ученых, которые разрабатывают теории автоматизации перевода и формализации языковых данных и процессов, которые представлены на платформе искусственного интеллекта BelAI.by.

С помощью современных технологических решений, как правило, можно получить связный и точный текст на другом языке за считанные секунды. Это стало возможным благодаря применению инструментов на базе искусственного интеллекта, в частности, самообучающихся нейронных сетей. ИТ-разработкой занимаются ведущие научные центры и университеты мира: Кембриджский и Оксфордский университеты, на базе которых был создан Центр по изучению искусственного интеллекта и будущего человечества, Массачусетский и Стэнфордский университеты, МГУ имени М. В. Ломоносова и Объединенный институт проблем информатики НАН Беларуси. Последние несколько лет компании, являющиеся мировыми лидерами в области информационных технологий, OpenAI, Prompt, Facebook AI Research Lab, Google и Microsoft внедряют методы машинного обучения и технологии на основе нейронных сетей для улучшения качества систем машинного перевода.

IT-разработчики осознают уязвимость искусственного интеллекта и стремятся максимально приблизить ИИ-перевод к человеческому. Одним из самых многообещающих последних достижений в области NLP является использование «трансформеров (transformers), новой архитектуры нейронной сети, позволяющей достичь самых современных результатов при решении задач машинного перевода» [3]. Особенностью трансформеров является их способность одновременно анализировать все части предложения, что значительно ускоряет процесс обработки и улучшает понимание контекста.

Как правило, система сначала кодирует исходное предложение в абстрактный набор чисел – а потом декодирует из чисел слова, но уже на другом языке. Нейросеть позволяет декодировать и предсказывать слова перевода, причем каждое предсказанное слово используется для предсказания следующего слова. Для более точного выбора слов для перевода используется контекст всего исходного предложения, а также контекст всех предыдущих предсказанных слов. Именно в этом основное отличие нейронных систем от систем, основанных на использовании правил и статистических методов, в которых контекст ограничен предыдущими 4–7 словами. И это более похоже на то, как переводит человек.

Системы с архитектурой последнего поколения Transformer, как правило, оснащены так называемым «механизмом внимания» (attention mechanism), который «при предсказании следующего слова как бы фокусируется на одном или нескольких словах исходного предложения, складывая эту информацию с закодированным полным контекстом» [3]. Контроль внимания в процессе перевода – это сложный многоуровневый процесс, который в работе искусственного интеллекта и в деятельности переводчика реализуется совершенно по-разному. Если внимание искусственного интеллекта – это «встроенные алгоритмы, призванные передать информацию через систему кодирования и декодирования за счет многоуровневых математических функций, то внимание переводчика – это способность концентрироваться на разных задачах, как лингвистических, так и внелингвистических, одновременно и успешно их выполнять» [4, с. 21].

Первые системы машинного перевода строились под конкретные языки, а в настоящее время одним из основных требований к подобному программному обеспечению является мультязычность. По данным, опубликованным в научном журнале *New Scientist* [5], в 2022 г. была создана и успешно протестирована инновационная модель ИИ, которая может осуществлять перевод с 204 языков. Модель, получившая название *No Language Left Behind (NLLB)*, опережает систему *Google Translate*, которая в настоящее время работает на 133 языках, и систему *Microsoft Translator*, которая обслуживает 110 языков. Модель *NLLB* поддерживает такие языки, как ачехский, на котором говорят 4,7 млн человек в Индонезии и Малайзии, и фриульский, на котором говорят всего 600 000 человек в северо-восточной Италии, а также может выполнять переводы на 55 африканских языков. Несмотря на свое

название, модель No Language Left Behind охватывает лишь незначительную часть из почти 7000 языков, существующих во всем мире.

Большие языковые модели, подобные NLLB, обучаются на огромных объемах разнообразных текстовых данных, таких как книги, статьи, веб-страницы и т.д. В процессе обучения модель «учится понимать» языковые закономерности, структуру предложений и контекст. Обучение больших языковых моделей – это сложный и ресурсоемкий процесс, требующий комбинации передовых алгоритмов машинного обучения и больших объемов данных.

Языковая модель NLLB использует трансферное обучение. Это означает, что она может применять полученные знания из одной задачи, чтобы обучиться выполнению другой задачи. Например, модель NLLB, обученная переводить с английского на испанский язык, может использовать полученные знания для перевода с английского на немецкий язык.

Следует отметить, что неоднократно предпринимались попытки использовать ИИ для перевода книг и научных работ. В 2020 г. издательство «Лабиринт» опубликовало книгу английского экономиста и бывшего советника при британском правительстве Д. Сасскинда «A World Without Work: Technology, Automation and How We Should Respond». Книга была переведена с английского на русский язык при помощи системы машинного перевода, чтобы читатели могли убедиться в «невероятных возможностях» современных самообучаемых алгоритмов. При переводе книг и художественных произведений очень важно качество переводов, а ИИ пока в этом отношении далек от совершенства.

В настоящее время очень востребованы наиболее сложные системы по распознаванию и синхронному переводу человеческой речи. В 2023 г. было разработано программное обеспечение Deepdub Go – Hollywood Grade Generative AI-powered Localization [6] для автоматического дубляжа видео на 65 языков, включая региональные диалекты. Deepdub Go осуществляет сквозной процесс дубляжа, включающий автоматическую транскрипцию, перевод, генерацию голоса и микширование аудио. Особенностью Deepdub Go, как отмечают разработчики, является точный учет контекста оригинальной аудиодорожки. Во-первых, при дубляже воспроизводятся эмоции говорящего. Во-вторых, при переводе, как утверждают разработчики, языковая модель учитывает и корректно переводит шутки и идиомы. В-третьих, Deepdub позволяет переводить на различные диалекты, уникальные для каждого региона. Многоязычные страны, такие как Индия (447 языков и 2000 диалектов), страны Латинской Америки (8 диалектов испанского языка) смогут в будущем получать контент, адаптированный к каждому конкретному региону.

По мнению представителей научного сообщества и Ассоциации переводческих компаний, «качество нейронного машинного перевода достигло такого уровня, что применение данных технологий коренным образом изменило профессию переводчика и всю отрасль переводов и локализации в целом» [7]. В связи с этим в 2023 г. Ассоциацией переводческих компаний

был принят «меморандум, регулирующий профессиональное применение средств автоматической генерации текста (включая машинный перевод и генеративные языковые технологии) с учетом интересов представителей профессии и потребителей (заказчиков) перевода, в том числе для оценки качества результата и рисков применения этих технологий» [7].

Кроме того, эксперты отмечают «острую необходимость выработки общего подхода к пониманию основ и принципов применения средств автоматической генерации текста в области профессионального перевода, как среди потребителей переводческих услуг, так и среди переводческих компаний, а также, что крайне важно, у представителей научного и академического сообщества и IT-разработчиков» [7].

В исследовании «The future of employment: how susceptible are jobs to computerisation» [8], опубликованном на сайте Оксфордского университета, экономисту К. Б. Фрей и эксперту по компьютерному обучению М. Осборн удалось определить количественное значение потенциального воздействия технологических инноваций на безработицу. Авторы исследования полагают, что развитие интеллектуальных технологий призвано оптимизировать работу переводчиков. Профессии письменных и устных переводчиков они отводят 265 место из 702 профессий, которые в будущем наиболее подвержены риску полной или частичной автоматизации.

Знания, опыт и творческий подход профессиональных переводчиков помогают им успешно конкурировать с ИИ. Подготовка в учреждениях высшего образования специалистов к обеспечению эффективной межкультурной коммуникации сейчас происходит «по тем моделям и представлениям о переводческой деятельности, которые сложились десятилетия тому назад и пока еще не утратили своей актуальности, однако формируемое сегодня новое поколение переводчиков необходимо подготовить к грядущим изменениям в обществе» [1, с. 21], обусловленным цифровизацией многих видов человеческой деятельности.

ЛИТЕРАТУРА

1. Гарбовский Н. К., Костикова О. И. Интеллект для перевода: искусный или искусственный? // Вестник Московского ун-та. Сер. 22. Теория перевода, 2019. № 4. С. 3–25.

2. Machine Translation Market Report 2024–2032 [Electronic resource] // URL: <https://www.imarcgroup.com/machine-translation-market> (accessed: 20.03.2024).

3. Прошина М. В., Виноградов А. Н. Анализ эффективности трансформеров для решения некоторых задач NLP [Электронный ресурс] // Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем. 2023. URL: <https://repository.ru-dn.ru/ru/records/article/record/102138> (дата обращения: 02.06.2024).

4. Шебаршина Д. Ю. Проблема внимания при синхронном переводе как один из ключевых факторов, обуславливающих «конкуренцию интеллектов» // Вестник Московского ун-та. Сер. 22. Теория перевода, 2021. № 2. С. 21–30.

5. Sparkes M. Meta's AI can translate between 204 languages, including rare ones [Electronic resource] // New Scientist. 2022. URL: <https://www.newscientist.com/article/2327061-metas-ai-can-translate-between-204-languages-including-rare-ones/> (accessed: 20.03.2024).

6. Kim E. AI-dubbing-startup-makes-your-favorite-movie-stars-multilingual [Electronic resource] // NoCamels – Tech & Innovation News. 2023. URL: <https://nocamels.com/2023/07/ai-dubbing-startup-makes-your-favorite-movie-stars-multilingual/> (accessed: 20.03.2024).

7. Берендяев М. В., Сант С. А., Светова С. Ю. 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 [Электронный ресурс] // Документы Ассоциации переводческих компаний. 2023. URL: <https://atcru.org/upload/iblock/ff5/fp58d29psjlsr3minttu0nuqqntlay50.pdf> (дата обращения: 02.06.2024).

8. Frey C. B., Osborne M. A. The future of employment: how susceptible are jobs to computerisation? [Electronic resource] // URL: <https://www.fhi.ox.ac.uk/wp-content/uploads/The-Future-of-Employment-How-Susceptible-Are-Jobs-to-Computerization.pdf> (accessed: 02.03.2024).