

## ПЕРЕВОД И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: ПРОБЛЕМЫ И ПУТИ РАЗВИТИЯ

**С. А. Лосева**

Белорусский государственный университет,  
пр. Независимости, 4, 220030, г. Минск, Беларусь, e-mail: tepinasvetlana@gmail.com

*Внедрение переводческих технологий в программу обучения будущих переводчиков заслуживает большого внимания, так как эти инструменты все больше проникают в повседневную рабочую практику настолько, что вносят существенные изменения в способ работы и взаимодействия переводческой группы. Данная статья является кратким руководством по видам автоматического перевода и онлайн-сервисам, помогающим в нелегком процессе «перехода» с одного языка на другой. Кроме того, в статье приведен анализ проблем и перспектив использования искусственного интеллекта в переводческой деятельности. Материал может быть широко использован в преподавании практического курса перевода. Полученные данные могут представлять интерес для практикующих переводчиков. Результаты работы могут быть использованы на семинарских занятиях по теории и практике перевода, а также в исследовательской практике студентов.*

**Ключевые слова:** переводческие технологии; искусственный интеллект; автоматический перевод.

## TRANSLATION AND ARTIFICIAL INTELLIGENCE: PROBLEMS AND WAYS OF DEVELOPMENT

**S. A. Loseva**

Belarusian State University,  
Niezaliežnasci Avenue, 4, 220030, Minsk, Belarus, e-mail: tepinasvetlana@gmail.com

*The introduction of translation technologies in the training of future translators deserves great attention, as these tools are increasingly penetrating into everyday work practice so much that they make significant changes in the way the translation team works and interacts. This article is a short guide to the types of automatic translation and online services that help in the difficult process of "transition" from one language to another. In addition, the article provides an analysis of the problems and prospects for the use of artificial intelligence in translation. The material can be widely used in teaching a practical translation course. The data obtained may be of interest to working translators. The results of the work can be used in seminars on the theory of translation, as well as in the research practice of students.*

**Keywords:** translation technologies; artificial intelligence; automatic translation.

Переводческий бизнес сильно выиграл от технологического развития, в первую очередь от создания искусственного интеллекта. Быть переводчиком — очень сложная задача, так как су-

ществуют различные виды перевода: от устного, целью которого является установление контакта и интерпретация сообщений двух и более людей, до письменного перевода, который требует определенной специализации (редакционный, научный, коммерческий, юридический и др.). Большие словари прошлого, где тысячи слов вводились в алфавитном порядке для использования в нужное время, требовали много времени, потому что поиск слов производился путем анализа тысяч терминов. Сегодня все это по мере развития технологий стало более оперативным, быстрым и мощным, прежде всего благодаря функциям немедленного поиска. Благодаря Интернету использование словаря не только расширилось и ускорилось, но и стало более полным по функциям, доступным онлайн, которые могут, например, «перехватывать» слова и при необходимости исправлять их. Более того, с помощью Интернета можно не только быстро выполнять переводы, но и узнавать происхождение слов и дополнительную информацию о них.

Чтобы понять, как можно эффективно и правильно выполнять перевод с помощью порталов и онлайн-словарей, стоит начать с определения и распознавания доступных функций и цифровых инструментов.

Цифровой словарь, *e-dictionary*, представляет собой полный справочник по электронной поддержке, в частности, это одноязычный или двуязычный словарь, предназначенный для использования носителями языка или изучающими *L2*. Он включает в себя словари для учащихся, доступные сначала на *CD-ROM* и *DVD*, а затем по сети, доступные на настольных компьютерах, ноутбуках и планшетных компьютерах или через мобильные телефоны. Процесс перевода означает любой процесс, который из первого текста, написанного на исходном языке, называемом «прототекстом», переводится после фазы декодирования и перекодирования на целевом языке, создавая «метатекст» [1].

*Машинный перевод* — это процесс использования искусственного интеллекта (ИИ) для автоматического перевода контента с одного языка на другой без вмешательства человека. Первые эксперименты с машинным переводом были проведены в 1950-х гг., а машинный перевод начал использоваться в 2000 г. Вначале качество перевода было очень низким, первые модели машинного перевода также требовали от разработчиков вручную определять и программировать то, что на самом деле представляло собой

большой набор правил. Появление глубокого обучения, которое работает с искусственным интеллектом, произвело настоящую революцию в переводе.

*Статистический машинный перевод:* статистические модели работают путем анализа огромного количества существующих переводов и многоязычных корпусов и поиска статистических моделей во входных данных, что позволяет программе генерировать гипотезу о том, как она должна переводить другие подобные тексты в будущем. Ресурсы, необходимые для обучения моделей, значительны, но результаты могут быть неплохими, особенно при переводе технических или научных текстов. Статистические модели перевода изначально были основаны на словах, но позже превратились в системы на основе предложений, которые составляют контекст слов.

*Нейронный машинный перевод.* Модели нейронных сетей сильно отличаются от систем, основанных на предложениях. В то время как последние разбивают входную фразу на набор слов и фраз и сопоставляют их со словом или фразой на целевом языке, нейронные сети учитывают всю входную фразу на каждом этапе при генерации выходной фразы. Нейронный машинный перевод использует нейронные сети для перевода исходного текста в целевой текст, а нейронные сети могут работать с очень большими наборами данных и требуют минимального контроля. Нейронные системы машинного перевода состоят из двух основных частей: сети кодирования и сети декодирования. Обе являются нейронными сетями — набором взаимосвязанных узлов, приблизительно смоделированных по образцу человеческого мозга. Это информационная система, в которой входные данные передаются через эти узлы для получения выходных данных. Эта архитектура нейронной сети называется нейронной сетью последовательности к последовательности (*Seq2Seq*), которая работает, просматривая предложение на исходном языке и создавая соответствующее предложение на целевом языке. Эффективность нейронного машинного перевода заключается в его архитектуре нейронной сети, которая позволяет ему учиться на больших объемах данных и адаптироваться к новым контекстам [2]. Это определяет такие результаты, как:

— *высокая точность:* за счет использования постоянно расширяющихся наборов данных и языкового моделирования механизмы *НМТ* могут понимать более широкий контекст слов и фраз,

что позволяет производить более точные и плавные переводы и улучшать их с течением времени;

— *быстрое обучение*: нейронные сети можно быстро обучить с помощью автоматизированных процессов, в отличие от дорогостоящих и в основном ручных методов, необходимых для машинного перевода на основе правил;

— *настройка*: обычно можно настроить выходные данные NMT и обновить модель с помощью терминологических баз данных, глоссариев для конкретных брендов и других источников данных для улучшения результатов.

Нейронный машинный перевод превосходит в переводе больших объемов контента в предельно сжатые сроки, переводе часто повторяющегося контента, переводе, который требует высокой точности нейронной сети, например, руководства пользователя или других типов справочных материалов.

В данной статье заслуживают упоминания интернет-порталы и сервисы для словарей и лингвистических словарей

*Google Translator* — это бесплатная служба перевода, разработанная *Google* в апреле 2006 г. Изначально *Google Translate* был выпущен как *Statistical Machine Translation*. Поскольку *SMT* использует алгоритмы прогнозирования для перевода текста, у него была низкая грамматическая точность. В январе 2010 г. *Google* представила приложение для *Android* и версию для *iOS* в 2011 г., которые служат портативным персональным переводчиком. В мае 2014 г. *Google* приобрела *Word Lens* для улучшения качества визуального и речевого перевода. В 2016 г. *Google* перевел свой метод перевода в систему нейронного машинного перевода. Он использует методы глубокого обучения для перевода всего предложения за раз и обеспечивает большую точность контекста. По состоянию на 2018 г. он переводит более 100 млрд слов в день. Существует около сотни переводимых языков.

*DeepL* родился в 2017 г. и основан на совершенно новом поколении нейронных сетей и инновационном дизайне, благодаря чему эти нейронные сети могут распознавать даже мельчайшие нюансы в значении предложения и правильно переводить его с неизвестной ранее степенью точности. В 2020 и 2021 гг. были запущены новые модели, способные более точно передавать значение переводов даже при работе со специализированными текстами с использованием определенного жаргона. Нейронные сети *DeepL* способны распознавать даже малейшие особенности язы-

ка и воспроизводить их в переводе. В настоящее время доступно 26 языков, бесплатная версия и несколько версий по подписке с более специализированными услугами.

Reverso переводит документы всех типов на более чем 25 языков, оставляя исходное форматирование. Он имеет автоматическую проверку и спрягает глаголы. Можно быстро найти синонимы благодаря интегрированному использованию нескольких словарей. Базовая версия бесплатна, с версией Премиум возможности увеличиваются.

*Wiktionary* — это бесплатный онлайн словарь, который на сегодняшний день содержит более 500 000 слов на широко используемых языках и многих других, от чероки до астурийского. После того, как вы зарегистрировались и приняли к сведению правила этикета, вы можете помочь пополнить словарь.

*Dudge* позволяет переводить примерно на 20 языках, имеет спряжение глаголов, форум, предоставляет несколько одноязычных словарей и совершенно бесплатен.

Несмотря на очевидные успехи в сфере программного обеспечения качества и скорости перевода, до сих пор существует множество проблем, которые нужно преодолеть.

В 2020 г. нейронный машинный перевод достиг точности 60—90 %, а это означает, что еще предстоит приложить усилия в области качественного редактирования и исследований, чтобы пройти старый тест Тьюринга. Существующее программное обеспечение не может обрабатывать нюансы человеческой речи, в том числе слова с несколькими значениями, предложения с несколькими грамматическими структурами. У него нет инструментов для обработки контекста, в котором используется язык. Это связано с тем, что эмоции, невербальное общение и культура влияют на порождение речи и на язык. Программы также продемонстрировали умеренную эффективность при переводе медицинских, юридических и коммерческих документов, в то время как в других случаях они выявили гендерную предвзятость. Кроме того, модели машинного перевода, такие как НМП, нуждаются в большом количестве параллельных данных для обучения переводу. Большинство языков мира не имеют достаточных данных: вот почему мы говорим о языках, которые «бедны ресурсами». По данным *Internet World Stats*, количество пользователей десяти самых популярных языков мира (английский, китайский, испанский, арабский, португальский, индонезийский/малайский, фран-

цузский, японский, русский и немецкий) в Интернете составляет примерно 77 %. Из них на английский и китайский приходится 25,9 % и 19,4 % соответственно, а в сумме все пользователи других языков составляют лишь 23,1 %. Для таких языков как китайский и английский, можно собрать миллиарды пар предложений для обучения модели НМП. Для языковых пар с ограниченными ресурсами, таких как китайский-хинди или китайский-суахили, это намного сложнее [3].

По оценкам групп исследований и разработок (*R&D*) *Lionbridge*, нейронный машинный перевод улучшается на 3—7 % каждый год. Эксперты рассчитывают улучшение с помощью измерения, называемого расстоянием редактирования. Расстояние редактирования рассчитывает количество изменений, которые человек должен внести, чтобы итоговое качество перевода было эквивалентно живому переводу. Нейронный машинный перевод будет продолжать развиваться по мере роста спроса на услуги перевода и улучшения машинного обучения в обучении машин перевода.

Скорость внедрения нейронного машинного перевода в будущем будет увеличиваться, поскольку объем контента, требующего локализации, растет в геометрической прогрессии.

Кризис, вызванный *COVID-19*, ускорил цифровую трансформацию многих компаний, что сделало переводческие услуги все более необходимыми. Такие рыночные условия приведут к использованию машинного перевода для определенного контента с контролем человека или без него. Выполняется это с помощью процесса постредактирования машинного перевода (*МТРЕ*), гибрида машинного перевода и традиционного человеческого перевода. Постредактирование происходит после фазы автоматического перевода для улучшения качества переведенного текста.

Для бизнеса услуги перевода станут более доступными, по крайней мере, для некоторых языков, благодаря нейронному машинному переводу. Такое снижение затрат компании сможет увеличить количество целевых рынков и быстрее доставлять товары на эти рынки. Внедрение нейронного машинного перевода поспособствует цифровой трансформации в международной экономике, появится более конкурентная среда. Конечные пользователи все чаще будут ожидать получения информации о продукте на своем родном языке. Соответствие этим ожиданиям компаний на всех рынках, на которых они работают, станет нормой, а не исключением.

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Гавриленко, Н. Н. Цифровая компетентность — ключевой компонент профессионализма переводчика / Н. Н. Гавриленко // Вестн. ПНИПУ. Проблемы языкознания и педагогики. — 2018. — № 3. — С. 139—150. DOI: 10.15593/2224-9389/2018.3.12
2. Инютин Н. Г. Формирование информационно-технологической компетенции будущего «переводчика в сфере профессиональной коммуникации»: дис. ... канд.пед наук.: 13.00.08/ Н. Г. Инютин. — Н.Новгород, 2006. — 320 л.
3. Internet World Stats [Electronic resource]. — Mode of access: <https://www.internetworldstats.com/>. — Date of access:01.09.2022.

## VERBAL LACUNAE AS PART OF DERIVATIONAL SUBSYSTEMS IN THE ENGLISH LANGUAGE

*V. Malmyha*

*Yanka Kupala State University of Grodno  
Ozheshko st., 22, 230023, Grodno, Belarus, e-mail: veranikak@gmail.com*

*The article considers verbal lacunarity at the morphological and derivational levels of the language with regard to affixation in the English language. The author defines a list of morphological and derivational gaps that admit the absence of a derivative, even if it may exist at the basis of the morphological rules of the language.*

**Keywords:** *lacuna; lacunarity; affixation.*

## ГЛАГОЛЬНЫЕ ЛАКУНЫ КАК ЧАСТЬ ДЕРИВАЦИОННЫХ ПОДСИСТЕМ В АНГЛИЙСКОМ ЯЗЫКЕ

*В. Малмыго*

*Гродненский государственный университет имени Янки Купалы  
ул. Ожешко, 22, 230023, г. Гродно, Беларусь, veranikak@gmail.com*

*В данной статье рассматривается глагольная лакунарность на морфологическом и словообразовательном уровнях английского языка. Определяется список морфологических и деривационных пробелов, допускающих отсутствие производного, даже если оно может существовать в основе морфологических правил языка.*

**Ключевые слова:** *лакуна; лакунарность; аффиксация.*

According to the grammar of the English language, a fairly common model of word formation is the formation of verbal nouns by adding the suffixes -(t)ion- or -al-. For example, the verb "recite" can form verbal nouns such as recitation and recital. However, many other