
ТОЧКА ЗРЕНИЯ

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ТЕХНОЛОГИЯХ МАШИННОГО ПЕРЕВОДА



Константин Константинович Колин

Доктор технических наук, профессор, главный научный сотрудник ФИЦ Институт управления (ИУ) РАН, Москва, Россия



Александр Алексеевич Хорошиллов

Доктор технических наук, профессор МАИ, ведущий научный сотрудник ФИЦ Институт управления (ИУ) РАН, старший научный сотрудник 27 ЦНИИ Минобороны России, Москва, Россия



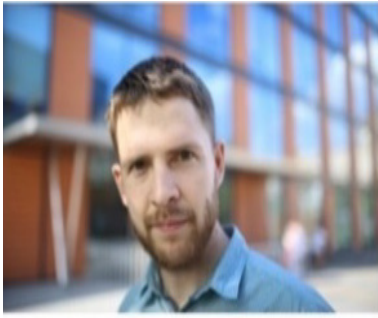
Юрий Викторович Никитин

Научный сотрудник ФИЦ Институт управления (ИУ) РАН, руководитель группы разработки АО Научно-производственная компания «Высокие технологии и системы разработки» (НПК «ВТ и СС»), Москва, Россия



Сергей Игоревич Пшеничный

Кандидат экономических наук, директор программ АО Научно-производственная компания «Высокие технологии и системы разработки» (НПК «ВТ и СС»), Москва, Россия



Алексей Алексеевич Хорошилов

Кандидат технических наук, старший научный сотрудник
27 ЦНИИ Минобороны РФ, Москва, Россия

***Аннотация.** Возможности машинного перевода тесно связаны с совершенствованием моделирования процессов понимания и генерации текстов на естественном языке, что традиционно относится к классу проблем искусственного интеллекта. В статье предпринята попытка проанализировать основные подходы к созданию технологий машинного перевода. Сделан вывод, что они пока не предусматривают формирования и использования динамических моделей мира, а движутся, главным образом, в направлении грамматически согласованного перевода последовательностей слов.*

***Ключевые слова:** машинный перевод; естественный язык; искусственный интеллект; технологии машинного перевода.*

***Для цитирования:** Искусственный интеллект в технологиях машинного перевода / Колин К.К., Хорошилов Ал-др. А., Никитин Ю.В., Пшеничный С.И., Хорошилов Ал-й А. // Социальные новации и социальные науки. – Москва : ИНИОН РАН, 2021. – № 2. – С. 64–80.*

URL: <https://sns-journal.ru/ru/archive/>

DOI: 10.31249/snsn/2021.02.05

© Колин К.К., © Хорошилов Ал-др. А., © Никитин Ю.В., © Пшеничный С.И., © Хорошилов Ал-й А.

Введение

Машинный перевод (МП), наряду с задачами распознавания образов и игрой в шахматы, в середине прошлого столетия был отнесен к классу проблем искусственного интеллекта (ИИ) [Панов, Ляпунов, Мухин, 1956]. Проблема автоматического распознавания образов и проблема игры компьютера в шахматы к настоящему времени успешно решены. Как заявляют некоторые крупнейшие транснациональные IT-компании (Google, IBM и др.), решена и проблема МП [Хобсон, Ханнес, Коул, 2020; Ганегедара, 2019]. Так ли это на самом деле? Заявления на эту тему делались неоднократно, но анализ показывал, что они не вполне соответствуют действительности. И это понятно, поскольку технология МП весьма сложна из-за тесной связи с проблемами моделирования процессов понимания и генерации текстов на естественном языке (ЕЯ). Последний, как известно, является универсальным средством общения между людьми, а также средством восприятия, накопления, хранения и передачи информации. Более того, ЕЯ является инструментом мышления человека [Апресян, 1966; Белоногов, Калинин, Хорошилов, 2004; Мельчук, 1999].

На протяжении длительного времени ЕЯ является объектом изучения ряда научных дисциплин, таких как лингвистика, семиотика, нейролингвистика и др. [Соссюр, 1977; Ветров, 1968; Шрейдер, 1974]. Например, психологи считают, что ЕЯ представляет собой вторую сигнальную систему человека, функционирующую на основе первой сигнальной системы (т.е. системы врожденных безусловных рефлексов, возникающих под воздействием сигналов, получаемых от зрительных, слуховых, тактильных и других рецепторов). Сигналы, поступающие во время разговора или восприятия речи, инициируют мыслительные процессы людей, но не определяют их полностью. Интерпретация речевых сигналов человеком (их понимание) происходит с учетом жизненного опыта и профессиональных знаний, накапливаемых в течение жизни. Этот комплекс знаний представляет собой некоторую динамическую модель мира, которая в процессе речевого общения выполняет роль пресуппозиций [Лурия, 2019; Максименко, 2000].

Известный советский лингвист В.А. Звегинцев констатирует: «Пресуппозиции образуют смысловой каркас, на котором строится текст (дискурс). Здесь... говорилось о существовании наряду с текстом подтекста». Их обязательная согласованность только и может «создать эффект уместности» [Звегинцев, 1976]. Таким образом, пресуппозиции – это определенный подтекст (умолчания), с которым должен согласовываться видимый или слышимый текст. Считается возможным для каждого предложения текста эксплицировать этот подтекст в виде наборов соответствующих «пресуппозиционных» предложений. Без учета «пресуппозиций» в письменных текстах

описание явлений реального или абстрактного мира всегда будет неполным, так как предполагается, что часть информации уже известна и ее нет необходимости излагать в тексте.

Письменный текст, как и звуковая речь, развертывается последовательно во времени, т.е. имеет линейную структуру, тогда как возникающие в сознании человека мысленные образы «многочисленны». При их словесном описании может быть принят различный порядок линейной развертки. Однако цель описания одна и та же – воссоздание в сознании читателей мысленных образов, адекватных авторским. Такое воссоздание осуществляется постепенно, путем воспроизведения предложения за предложением и «монтажа» возникающих при этом частичных образов в целостный, соответствующий содержанию текста. При этом в каждом предложении элемент его актуального членения – «тема», – выполняет роль «стыковочного узла», служащего для подключения нового, обозначаемого этим предложением образа к ранее построенному. Данная модель восприятия текста позволяет объяснить тот факт, что связи между предложениями выражаются в большинстве случаев с помощью лексических повторов. В «стыковочных узлах» предложений понятия предшествующего текста повторяются или буквально, или в виде синонимических и эллиптических¹ конструкций, или в виде родовых наименований и местоимений.

Таким образом, моделирование процессов восприятия и воспроизводства ЕЯ должно по своей сути отражать работу мозга человека при постижении им окружающей действительности. Поэтому при МП необходимо учитывать важнейшие особенности мыслительной деятельности людей. Кратко рассмотрим основные подходы к разрешению данных проблем на основе технологий ИИ.

Семантико-синтаксический перевод на основе правил

Первые эксперименты по МП, подтвердившие принципиальную возможность его осуществления, были проведены в 1954 г. в Джорджтаунском университете (г. Вашингтон, США). Вскоре после этого в промышленно развитых странах мира были начаты исследования и разработки, направленные на развитие технологий МП. Возникающие при этом проблемы оказались значительно сложнее, чем это представляли себе пионеры МП конца пятидесятых – начала шестидесятых годов прошлого века.

Первоначальный энтузиазм осуществления МП путем прямого пословного перевода (Direct Machine Translation) быстро сменился разочарованием. Тем не менее исследования в области МП продолжались. Достижением периода 1970–1980-х годов стал МП «на основе правил» (Rule-based Machine Translation, RBMT), которые были сформулированы в результате фундаментальных исследований ЕЯ. Одним из итогов таких исследований стала формальная модель языка, предло-

¹ Эллиптические конструкции – это неполные придаточные предложения, которые, как правило, состоят из союза и прилагательного или существительного (с предлогом или без предлога).

женная И.А. Мельчуком и названная им «Смысл-Текст» [Мельчук, 1999]. Эта модель представляет собой описание естественного языка, понимаемого как устройство («система правил»), обеспечивающее человеку переход от СМЫСЛА к ТЕКСТУ (воспроизведение, т.е. «говорение», или построение текста) и от текста к смыслу («понимание», или интерпретация текста).

Теоретическая концепция Мельчука постулирует многоуровневую МОДЕЛЬ языка, в которой построение текста на основе заданного смысла происходит не непосредственно, а с помощью серии переходов от одного уровня представления к другому. При этом выделяются несколько уровней описания языка: фонологический, поверхностно-морфологический, глубинно-морфологический, поверхностно-синтаксический, глубинно-синтаксический уровни и семантический. Каждый уровень характеризуется набором собственных лексических единиц и правил их представления, а также набором лингвистических правил перехода от одного уровня представления к соседним в соответствии с принципами [Мельчук, 1999].

1. Принцип полноты – обеспечение возможно более полного охвата структур описываемого языка. Это возможно при условии, что тестовые предложения будут отбираться не случайно, а на основе тщательной выборки разнообразного текстового материала. Полнота может быть обеспечена, если исходные тесты хорошо представляют генеральную совокупность.

2. Принцип экономичности – создание рационального количества правил во избежание ненужной детализации, повторов и избыточности описания.

3. Принцип непротиворечивости – четкое разграничение подобных и разных структур. Правила грамматики не должны противоречить друг другу. Логичное следование одного из другого, непересечение одних правил с другими позволяет избежать ошибок.

4. Удобство и легкость записи правил – при минимальности условных символов грамматика должна нести максимум информации. Алгоритмы должны также быть достаточно гибкими для облегчения введения в грамматику новых правил для анализа / синтеза неучтенных структур.

5. Оптимизация алгоритмической обработки текста в соответствии с правилами грамматики, которые должны обеспечивать наибольшую скорость анализа и синтеза структуры предложения.

Многоуровневая модель языка послужила основой для ряда систем МП [Кулагина, 1979; Марчук, 1983; Пиотровский, 2002]. Возможность получения грамматически правильного пословного перевода в них достигалась в основном процедурными и декларативными средствами на основе учета синтаксических и семантических признаков слов, включенных в состав двуязычных словарей. В составе этих систем присутствовали процедуры морфологического, синтаксического, семантического анализа и синтеза текстов, а также процедура трансфера – соотнесения слов исходного текста и их переводных соответствий на целевом языке.

Надо отдать должное пионерам МП и их ближайшим последователям. Они немало сделали в области теории и практики МП. Но многие важнейшие проблемы не были решены. Например, выбор переводных эквивалентов для слов и словосочетаний исходного текста. При решении этой проблемы стремились, прежде всего, получить грамматически правильный пословный перевод. Вопросы полисемии¹ слов разрешались в основном процедурными средствами на основе учета их синтаксических и семантических признаков. Поэтому системы МП первых трех десятилетий их развития можно охарактеризовать как системы семантико-синтаксического преимущественно пословного перевода. Хотя словосочетания здесь использовались, но в меньшей степени.

На наш взгляд, семантико-синтаксический пословный и преимущественно пословный МП текстов не имеет перспективы. В ЕЯ смысл предложений и словосочетаний, как правило, не сводится или не полностью сводится к смыслу составляющих их слов, и при переводе он не может быть «вычислен» на основе синтаксических и семантических признаков этих слов. Поэтому эта концепция была заменена на концепцию семантико-синтаксического преимущественно фразеологического перевода. Такой подход в большей мере соответствует природе ЕЯ, и благодаря ему специалисты существенно продвинулись в повышении качества перевода. Но эти улучшения были получены путем колоссальных трудозатрат при составлении правил и ручного создания двуязычных словарей. Тем не менее задача достижения уровня МП, незначительно уступающего ручному переводу переводчика средней квалификации, была далека от завершения, и перспектив ее успешного решения с помощью этого подхода не предвиделось.

В качестве свидетельства неблагоприятного состояния исследований и разработок в области МП этого периода можно рассматривать заявление руководителя японской государственной программы профессора Макото Нагао из университета Киото о тупиковом развитии систем RBMT, сделанное им в одном из своих докладов в 1982 г. В 1984 г. он предложил новую концепцию МП [Nagao, 1984]. Согласно этой концепции, автоматический перевод текста должен осуществляться по аналогии с текстами, ранее переведенными вручную. Подход Макото Нагао дал толчок новому направлению, получившему название статистический МП – Statistical Machine Translation (SMT).

Статистический машинный перевод

Впервые система на основе этой концепции была реализована на рубеже 1990-х годов в исследовательском центре IBM. Словари и лингвистические правила в данном случае были полностью заменены большими корпусами параллельных текстов (билингв). На их базе строилась так называемая модель перехода (transition model), в которой вычислялась вероятность того, что целе-

¹ Полисемия (от греч. Πολυσημεία – «многозначность») – многозначность, многовариантность, т.е. наличие у слова (единицы языка, термина) двух и более значений, исторически обусловленных или взаимосвязанных по смыслу и происхождению.

вое слово или фраза будет правильным переводом соответствующего слова или фразы. Дополнением к модели перехода являлась модель выравнивания слов (word alignment model), устанавливающая взаимное соответствие между словами предложения или фразы исходного и целевого языков. Алгоритмы перевода обучались с использованием кортежей исходного предложения и соответствующих ему целевых предложений. Предполагалось, что такой перевод будет более качественным, поскольку данные в виде параллельных предложений богаче, чем двуязычные словари слов или словосочетаний [Денисова, 2018].

Хотя первые версии системы SMT компании IBM выполняли перевод на основе слов, но затем разработчики стали экспериментировать с SMT на основе фраз. Этот подход к переводу позволил реализовать ряд моделей отношений: «один-к-многим», «многие-к-многим» и «многие-к-одному». Основной целью SMT на основе фраз является построение модели фразового перевода (phrase translation model), которая содержит распределение вероятностей различных целевых фраз-кандидатов исходной фразы. Отсутствие словарей компенсируется ведением огромных баз параллельных предложений на двух языках. В случае отсутствия прямого соответствия между грамматическим строем исходного и целевого языков фразовая модель перехода дополняется операциями перестановки слов [Денисова, 2018].

Одним из существенных недостатков SMT является значительная ресурсозатратность процесса декодирования (поиска наиболее вероятной целевой фразы для данной исходной). Это связано с вычислительной обработкой большого массива выявленных фраз в различных комбинаторных сочетаниях слов и огромного числа соответствующих им целевых фраз. Для сокращения числа рассматриваемых комбинаций в модель перехода включались синтаксические модели представления исходного предложения. Последние обеспечивали возможность перестановки синтаксических конструкций предложения и вставки слов в состав конструкций целевого языка.

С 2006 г. языковую модель SMT начали использовать онлайн-переводчики таких компаний, как Google, «Яндекс», Bing и др. Несмотря на то что этот подход обеспечил возможность более качественных МП, он все же не лишен серьезных недостатков. Основным из них является неполная грамматическая согласованность перевода, обусловленная тем, что не используются лингвистические механизмы такого согласования. Кроме того, переводные соответствия, полученные статистическим методом, «выдергивались» из разных текстов, не связанных общей темой. Поэтому в рамках SMT не всегда удается обеспечить смысловую связанность¹ текста.

¹ Под смысловой связанностью текста или его фрагмента понимается совокупность наименований или понятий, расположенных в тексте в определенном порядке и отражающих основное смысловое содержание текста (его фрагмента).

Нейросетевой машинный перевод

Достижения в области нейробиологии и NLP¹ в конце прошлого столетия привели к развитию теории сознания, в которой моделирование мыслительных или поведенческих процессов осуществляется с использованием нейросетей (коннекционизма). В рамках этой концепции возникло предположение, что МП также может быть реализован на основе нейросетевых моделей. Основная идея нейронного МП (Neural Machine Translation, NMT) заключается в том, что он может быть осуществлен на основе предварительного (глубокого) обучения сети (программы, модели) на большом корпусе параллельных предложений (на исходном и целевом языках) путем их последовательной обработки конечным набором логических правил. Сам процесс NMT заключается в обучении модели на параллельных текстах, проверке обученной модели и выполнении на этой основе переводов текстов на целевой язык [Google's multilingual neural ..., 2017].

Первые попытки реализации NMT в компании Google в 2004 г. выявили целый ряд недостатков. Одним из них была низкая скорость обучения программы, так как большое число выявленных признаков требовали много времени на формирование целевой информации. Кроме того, NMT оказался неэффективным при работе с редкими словами, а в отдельных случаях не удавалось перевести все слова входного предложения.

Дальнейшее совершенствование Google's NMT (GNMT) было направлено на повышение быстродействия и качества переводимых текстов. Например, для эффективного использования всей информации, полученной на предыдущих этапах, были задействованы рекурсивные нейросети (RNS) с долгой краткосрочной памятью (long short-term memory, LSTM) [Hochreiter, Schmidhuber, 1997; Gers, Schmidhuber, Cummins, 2000]. Чтобы уменьшить время вывода, использовались вычисления с низкой точностью (low-precision arithmetic), а в целях дополнительного ускорения – специальное оборудование (Google's Tensor Processing Unit – тензорный процессор Google). Перевод редких слов осуществлялся путем их представления на входе и выходе в форме набора составных элементов (частей слова – wordpieces) [Schuster, Nakajima, 2016]. Для решения проблемы неполного перевода применялся метод лучевого поиска (beam search) и был реализован штраф за пропуск слов [Ганегедара, 2019]. Привлекались также другие разнообразные механизмы и приемы, способствующие значительному улучшению результатов NMT.

Разработчики стремятся к созданию моделей с большей глубиной нейросети (большим числом слоев), так как они характеризуются более высокой точностью по сравнению с другими моделями машинного обучения. Дополнение их, например, современными методами инициализации

¹ Нейролингвистическое программирование (от англ. Neuro-linguistic programming) – подход к межличностному общению, развитию личности и психотерапии.

Завьера (Xavier) обеспечивает более короткий период обучения [Gradient flow in recurrent nets ..., 2001].

В настоящее время GNMT является одним из самых востребованных в мире автоматических переводчиков. Ежедневно сервис обрабатывает около 143 млрд слов более чем на 100 языках [Блеск и нищета ..., 2020]. Компания Google перешла на GNMT вместо ранее используемой SMT [Ганегедара, 2019]. «Майкрософт» использует похожую технологию для перевода речи (в том числе в «Майкрософт-переводчике» и «Skype-переводчике»). Гарвардской группой по обработке естественного языка была выпущена OpenNMT – система нейронного МП с открытым исходным кодом. «Яндекс-переводчик» базируется на гибридной модели, использующей технологии SMT и NMT. Выбор лучшего перевода из полученных результатов обеспечивает технология CatBoost, в основе которой также лежит машинное обучение [Машинный перевод, 2021].

Фразеологический машинный перевод

Впервые в СССР идею создания переводчика, обеспечивающего высококачественный автоматический перевод текстов, высказал Г.Г. Белоногов¹ (1975). Свое предложение он сформулировал в виде концепции фразеологического МП (FMT) [Каким быть машинному переводу ..., 2002]. В краткой форме ее основные положения представляют следующее.

По мнению Г.Г. Белоногова, для реализации адекватного МП необходимо учитывать объективные законы функционирования языка и мышления, богатый опыт межнационального общения, а также опыт переводческой деятельности, накопленный человечеством. Этот опыт свидетельствует о том, что в процессе перевода текстов в качестве основных единиц смысла выступают, прежде всего, фразеологические словосочетания, выражающие определенные понятия. Поэтому системы МП должны переводить не слова и их последовательности, а мысленные образы в форме слов и словосочетаний.

Кроме того, в языке объективно существует иерархия смысловых единиц. Причем смысловое содержание вышестоящих единиц не всегда сводимо к смыслу составляющих их нижестоящих единиц. Наиболее устойчивыми единицами смысла являются фразеологические понятия. Они также служат теми базовыми «строительными блоками», которые формируют смысловые единицы более высоких уровней – предложения, сверхфразовые единства и тексты [Белоногов, Хорошилов, Хорошилов, 2005].

¹ Белоногов Г.Г. (1925–2018) – один из основоположников отечественной информатики, признанный как в России, так и за рубежом, известный специалист в области компьютерной лингвистики и автоматической обработки текстов. Работал в 27 ЦНИИ МО (1961–1980), затем в ВИНТИ РАН (1980–2001) и лингвистической фирме МетаФраз (2003–2005). В частности, им была разработана уникальная машинная грамматика, базирующаяся на системе флексивных классов русского языка при реализации процедур морфологического анализа [Каким быть машинному переводу ..., 2002; Средства машинной грамматики русского языка ..., 2018].

Второй не менее важной единицей смысла является предложение. Основной чертой предложения выступает предикативность – т.е. свойство утверждать наличие у объектов определенных признаков и их отношений. Свойством предикативности обладают также высказывания на формализованных языках. Это позволяет сделать вывод, что в основе предложений на ЕЯ и формализованных логических высказываний лежит предикатно-актантная структура, компоненты которой представляют понятия-предикаты (признаки и отношения) и понятия-актанты, выступающие в роли описываемых объектов. Предикатно-актантные структуры являются теми смысловыми инвариантами, которые позволяют осуществлять перевод текстов с одного ЕЯ на другие. Одновременно они обеспечивают возможность решения основной задачи МП – передачу смыслового содержания исходного текста на целевой язык.

Выдающейся заслугой Г.Г. Белоногова является то, что он разработал модель FMT, обосновал и предложил базовый состав ее процедурных и программных средств, а также сформулировал принципы функционирования. В соответствии с этими принципами система FMT должна включать понятийную базу, содержащую переводные эквиваленты часто встречающихся терминологических словосочетаний, а также фрагментов фраз, служебных конструкций и отдельных слов. В процессе перевода текстов система использует хранящиеся в этой базе переводные эквиваленты в следующем порядке: а) сначала для очередного предложения исходного текста делается попытка перевести его как целостную фразеологическую единицу; б) в случае неудачи – переводятся входящие в его состав наиболее длинные синтаксические конструкции, а при их отсутствии – более короткие словосочетания; в) наконец, осуществляется пословный перевод тех фрагментов предложения, которые не удалось перевести первыми тремя способами. Фрагменты выходного текста, полученные всеми рассмотренными способами, должны грамматически согласовываться друг с другом (с помощью процедур морфологического и синтаксического синтеза) [Хорошилов, 2006; Хорошилов, Кан, Хорошилов, 2019].

Архитектура системы FMT, ориентированная на реализацию вышеуказанных принципов, состоит из трех модулей: модуль анализа исходного текста, модуль трансфера и модуль генерации переведенного текста. Это позволяет осуществлять сквозную обработку всего текста в оптимальном режиме [Хорошилов, Кан, Хорошилов, 2019]. В свою очередь, сквозная обработка текстов (а не отдельных предложений, как в SMT и NTM) предоставляет возможность смыслового связывания системы понятий и их отношений в единое целое в пределах всего текста.

Реализация подхода FMT, также как и подхода NMT, базируется на предварительно обученных языковых моделях. Отличием является только то, что NMT основываются на упрощенной модели языка и на его поверхностном представлении (см. [Мельчук, 1999]). Между тем в основе ЕЯ лежит, прежде всего, понятийная система, не всегда полностью отражающаяся на его поверхностном уровне.

Развитие системы FMT происходило в том числе путем реализации трансформационной двуязычной модели и создания четырехслойного комплекса двуязычных словарей по широкому спектру тематических областей [Технологии создания новых направлений перевода ..., 2017]. Отличительной особенностью словарей FMT является простая структура словарных статей. Входом служат любые фрагменты исходных текстов (но не более 16 слов), а понятия и их переводные эквиваленты на целевом языке могут быть представлены в любой грамматической форме. Единственное требование заключается в том, что не должна быть разрушена грамматическая согласованность слов внутри этих конструкций.

Разработанные технологии FMT с начала 2000-х использовались в ряде высокотехнологичных отраслей, учебных и научно-исследовательских организациях России.

Опыт практического использование технологий МП

Когда речь заходит о технологиях МП, то в первую очередь возникает вопрос о качестве перевода. Ряд используемых метрик качества (например, BLEU¹), несмотря на то что базируется на вероятностных и статистических методах вычислений, при ближайшем рассмотрении оказывается ориентированным на те же технологии, на которых построены анализируемые системы [Хобсон, Ханнес, Коул, 2020; Ганегедара, 2019]. А схожесть или отличие от лексики эталонного перевода лексики выполненных переводов не может служить критерием качества последних. На практике основными критериями качества перевода остаются те, которые устанавливает заказчик (в техническом задании и т.д.) [Технологии создания новых направлений перевода ..., 2017].

В настоящее время агрегаторы новостных информационных сообщений в полной мере используют технические возможности автоматических переводчиков. Например, авиационное происшествие может описываться следующей фразой: «Сам самолет был поврежден без ремонта и его списали и демонтировали там, где он приземлился». При чтении этой «корявой» фразы сразу становится понятно, что она, скорее всего, получена путем МП. Человек тем не менее все же поймет, что означает словосочетание «поврежден без ремонта». С научно-техническим переводом даже относительно простых предложений ситуация гораздо сложнее.

В статье [Блеск и нищета ..., 2020] приводятся следующие варианты перевода предложения «*Less fuel is consumed in the use of V shape loading*», выполненные инструментами различных сервисов:

– вариант № 1 (DeepL): «*при использовании нагрузки V-образной формы расходуется меньше топлива*»;

– вариант № 2 (GT): «*меньше топлива расходуется при использовании V-образной загрузки*»;

¹ Сопоставляет число n-грамм (n последовательных слов) МП с n-граммами эталонного ручного перевода. Разработана сотрудниками компании IBM в начале 2000-х годов.

– вариант № 3 («Яндекс-переводчик»): «меньше топлива потребляется при использовании V-образной загрузки».

Все три варианта – гладкие, грамматически верные, – но абсолютно бессмысленные. Синтаксической конструкции «*V shape loading*» в узкой предметной области исходного предложения соответствовало словосочетание «загрузка (самосвала) с движением колесного погрузчика по V-образной траектории».

Данный пример иллюстрирует явление пресуппозиции, без учета которой перевод становится бессмысленным. Особенно слабо улавливают современные технологии МП профессиональную специфику. Но даже в этом случае нельзя согласиться с категоричным заявлением автора той же публикации: «МП по любой технологии никогда не сможет правильно переводить специализированные тексты, это принципиальная невозможность – такая же, как невозможность постройки вечного двигателя». На самом деле для различных задач МП необходимо использовать тот набор инструментов, который способен их решить. Как следует из рассмотрения подходов SMT и NMT, в этих системах таких инструментов нет. В то же время методика FMT был изначально ориентирована на перевод подобных текстов.

Подводя итоги анализа существующих способов МП, можно констатировать, что в рамках современных технологий SMT и NMT достаточно хорошо решаются задачи автоматизированного перевода текстов, лексический состав которых доступен в большом количестве исходных и переводных вариантов. К ним можно отнести новостные, гуманитарные и общетехнические тексты. Системы выдают неплохие результаты при условии, что лексический, а главное, понятийный состав переводимых текстов в незначительной степени отличается от того, по которому они обучались. В случае значительного изменения этих параметров необходимо «доучивать» систему.

Как часто это необходимо выполнять и какова величина требуемых трудозатрат – вопрос открытый. Видимо, они «по силам» транснациональным IT-компаниям и крупным агрегаторам. Тем более что информационный перевод не требует точной передачи смыслового содержания текстов – достаточно выстроить грамматически правильные предложения. В тех случаях, когда необходимо обеспечить точный и качественный перевод, эту задачу предполагается возложить на профессиональных переводчиков. Как видится заказчикам таких переводов, последним нужно только немного подправить результаты МП.

То, что это далеко не так, поняли разработчики сервисов постредактирования МП (РМЕТ), целью которых являлось обеспечение переводчиков удобными эргономичными инструментами для коррекции выполненного МП. Для редактирования уже выполненного перевода переводчику необходимо было обратиться к исходному тексту, понять, как был произведен перевод самого текста и его синтаксических конструкций, а по сути – выполнить ручной перевод, – и только после этого произвести коррекцию МП.

Возникает вопрос, а не лучше ли обеспечить переводчику возможность самому контролировать процесс перевода, т.е. наблюдать, как выполняется перевод синтаксических конструкций исходного предложения, и вмешиваться только в тех случаях, когда перевод является неприемлемым.

Такая технология диалогового МП, ориентированная на перевод научно-технических текстов, существует и называется интерактивный фразеологический перевод. Она предоставляет переводчику возможность не только вмешиваться в процесс перевода, но и обеспечивает обучение системы для последующего перевода аналогичных по лексическому составу и понятийной структуре научно-технических текстов.

В частности, на этих принципах была разработана технология перевода русскоязычного сайта для портала федерального ведомства, с целью обеспечения его англоязычного «зеркала» [Хорошилов, Кан, Хорошилов, 2019]. Причем заказчик выставил жесткие требования к скорости и качеству перевода. Время доступа к информации, содержащейся на англоязычной странице сайта, должно было не более чем на 30% превышать время доступа к аналогичной информации русскоязычного сайта. За этот отрезок времени встроенный переводчик должен был разобрать русскоязычную страницу, выделить ее текстовую составляющую, автоматически перевести и вставить в каркас страницы. При всем этом вычислительные ресурсы, выделенные заказчиком для реализации этой задачи, были весьма скромными. Требуемое качество перевода сайта обеспечивалось предварительным автоматизированным созданием словарей понятий тематической области, а также обеспечением возможности оперативного контроля и динамического пополнения словарей переводной лексикой вновь поступающей на сайт информации [Хорошилов, Кан, Хорошилов, 2019].

И все же – насколько существующие технологии МП приблизились к изначальному представлению об ИИ [Панов, Ляпунов, Мухин, 1956; Апресян, 1966; Белоногов, Калинин, Хорошилов, 2004; Мельчук, 1999; Соссюр, 1977]. Здесь необходимо обратиться к опыту профессиональных переводчиков. Переводчик, прежде всего, должен быть грамотным лингвистом, хорошо знающим все грамматические тонкости построения как исходного, так и целевого языков, а главное, понимать, какие мысленные образы заложены в исходном тексте, и как они представляются на целевом языке. Он также должен хорошо ориентироваться в понятийной системе предметной области переводимых текстов, хотя в случаях необходимости всегда можно обратиться к различной справочной информации или к консультациям специалистов. Качество перевода переводчик должен уметь оценивать как с точки зрения носителя целевого языка, так и специалиста в данной предметной области.

Всего этого лишены современные системы МП. Если на начальном этапе были попытки с помощью профессиональных лингвистов построить модели ЕЯ и на их основе реализовать МП, то с ориентацией на статистический перевод они отошли на второй план. В основу новых моделей

был заложен принцип перевода последовательностей слов исходного языка «предсказанными» последовательностями целевого языка, установленными на основе аналогии с ранее выполненными переводами. Другими словами – переводятся часто встречающиеся и относительно несложные синтаксические конструкции.

В современных системах МП также полностью игнорируется явление пресуппозиции. Хотя это те самые «базовые знания», которые должны «незримо» присутствовать в профессионально выполненном переводе. Нужно сказать, что апологеты NLP это понимают [Ганегедара, 2019]. Однако изначально положенная в основу NMT достаточно примитивная коннекционистская теория сознания не предполагает иного решения проблемы МП.

В большей степени базовые знания учитываются в подходе FMT. Именно в его рамках предпринимаются попытки формализации и автоматического формирования тематических баз данных, пока только в виде тематических онтологий [Хорошилов, Кан, Хорошилов, 2019]. Однако превалирование технологий NMT замедлило развитие других подходов в МП.

Заключение

В интервью одного из высокопоставленных отечественных чиновников было сказано: «...Цифровизация очень скоро освободит нас от переводов. Благодаря использованию нейронно-сетевых технологий, качество переводов буквально от месяца к месяцу существенно улучшается. Функция запоминания позволяет машине выбрать из большого числа вариантов тот перевод, который наиболее близок к правильному. Сейчас это становится реальным. Естественно-научные статьи уже можно не переводить. Нажали кнопку – автомат выдает перевод...» [Блеск и нищета ..., 2020]. Это высказывание как бы подтверждает факт решения всех проблем МП. Больше нет необходимости что-либо делать в этой области, кроме как использовать технологии NMT, разработанные в западных странах.

Однако это абсолютно не соответствует реальному положению дел, особенно в области естественно-научного перевода. В естественно-научных статьях формулируются новые представления о реальном мире, вводятся новые понятия, излагаются новые теории и концепции. Незначительное число таких публикаций не может обеспечить информационную базу для систем SMT и NMT, требующих миллионы (для систем SMT) и миллиарды (для систем NMT) параллельных предложений, или обеспечить их полноценное обучение. При этом значительная часть информации в профессиональных сообществах находится в зоне пресуппозиций. Несмотря на то что в названии подхода NMT содержится термин Neural, ассоциируемый с нервной системой человека, ничего даже отдаленно связанного с моделированием мыслительной деятельности человека в области МП в этом подходе нет.

Каким будет будущее МП, если технология NMT окажется тупиковым направлением НТП? Не лучше ли именно сейчас, когда имеют место некоторые технологические прорывы в ряде высокотехнологичных отраслей, начать развивать отечественные технологии МП? В России есть потенциал и понимание того, как можно построить оригинальные системы ИИ, а не «плестись в хвосте» зарубежных разработок. Но пока мы вынуждены констатировать, что в области МП технологий ИИ в его традиционном понимании нет.

Список литературы

1. Апресян Ю.Д. Идеи и методы современной структурной лингвистики (краткий очерк). – Москва : Просвещение, 1966. – 304 с. URL: <https://www.booksite.ru/fulltext/apresyn/text.pdf> (дата обращения 04.04.2021).
2. Белоногов Г.Г., Хорошилов Ал-др А., Хорошилов Ал-сей А. Единицы языка и речи в системах автоматической обработки текстовой информации // Научно-техническая информация. Сер. 2. – 2005. – № 11. – С. 21–29.
3. Белоногов Г.Г., Калинин Ю.П., Хорошилов А.А. Компьютерная лингвистика и перспективные информационные технологии. Теория и практика построения систем автоматической обработки текстовой информации. – Москва : Русский мир, 2004. – 246 с.
4. Блеск и нищета машинного перевода // TOPTR. – 2020. – 11.02. – URL: <https://www.toptr.ru/library/translation-as-service/blesk-i-nishheta-mashinnogo-perevoda.html> (дата обращения 02.03.2021).
5. Ветров А.А. Семиотика и ее основные проблемы. – Москва : Издательство политической литературы, 1968. – 264 с.
6. Ганегедара Т. Обработка естественного языка с TensorFlow. – Москва : ДМК Пресс, 2019. – 382 с.
7. Денисова Д.С. Современные системы машинного перевода: статический машинный перевод // Синергия наук. – 2018. – № 19. – С. 1425–1434.
8. Звегинцев В.А. Предложение и его отношение к языку и речи. – Москва : Изд-во МГУ им. М.В. Ломоносова, 1976. – 308 с.
9. Каким быть машинному переводу в XXI веке / Белоногов Г.Г., Хорошилов Ал-др А., Хорошилов Ал-сей А., Козачук М.В., Рыжова Е.Ю., Гуськова Л.Ю. // Перевод: традиции и современные технологии. – Москва : ВЦП, 2002. – С. 56–69.
10. Кулагина О.С. Исследования по машинному переводу. – Москва : Наука, 1979. – 320 с.
11. Лурия А.Р. Язык и сознание. – Санкт-Петербург : Питер, 2019. – 336 с.
12. Максименко С.Д. Общая психология. – Москва : Рефл-бук : Ваклер, 2000. – 528 с.
13. Марчук Ю.Н. Проблемы машинного перевода. – Москва : Наука, 1983. – 233 с.
14. Машинный перевод // Яндекс. – 2021. – URL: <https://yandex.ru/company/technologies/translation> (дата обращения 12.03.2021).
15. Мельчук И.А. Опыт теории лингвистических моделей «Смысл \Leftrightarrow текст». – 2-е изд. – Москва : Школа «Языки русской культуры», 1999. – 346 с. – URL: https://www.studmed.ru/melchuk-ia-opyt-teorii-lingvisticheskikh-modeley-smysl-tekst_f059f362eba.html (дата обращения 11.06.2021).
16. Панов Д.Ю., Ляпунов А.А., Мухин И.С. Автоматизация перевода с одного языка на другой // Труды сессии по научным проблемам автоматизации производства. – Москва : Изд-во АН СССР, 1956. – С. 181–214.
17. Пиотровский Р.Г. Новые горизонты машинного перевода // Научно-техническая информация. Сер. 2. – 2002. – № 1. – С. 17–29.
18. Соссюр Ф. де. Труды по языкознанию. – Москва : Прогресс, 1977. – 696 с.
19. Средства машинной грамматики русского языка (по Г.Г. Белоногову) / Аблов И.В., Козичев В.Н., Ширманов А.В., Хорошилов Ал-др А., Хорошилов Ал-сей А. // Научно-техническая информация. Сер. 2. – 2018. – № 6. – С. 32–46.
20. Технологии создания новых направлений перевода для системы МетаФраз (на примере казахско-русского перевода) / Захаров В.Н., Никитин Ю.В., Хорошилов Ал-др А., Хорошилов Ал-ей А. // Научно-техническая информация. Сер. 2. – 2017. – № 9. – С. 29–39.
21. Хобсон Л., Ханнес Х., Коул Х. Обработка естественного языка в действии. – Санкт-Петербург : Питер, 2020. – 576 с.
22. Хорошилов А.А. Архитектура систем фразеологического машинного перевода // Научно-техническая информация. Сер. 2. – 2006. – № 5. – С. 15–19.
23. Хорошилов Ал-др А., Кан А.В., Хорошилов А.А. Фразеологический машинный перевод. – Москва : Директ-Медиа, 2019. – 467 с.
24. Шрейдер Ю.А. Семиотические основы информатики. – Москва : ИПКИР, 1974. – 81 с.
25. Gers F.A., Schmidhuber J., Cummins F. Learning to forget: Continual prediction with LSTM // Neural Computation. – 2000. – Vol. 12, №. 10. – P. 2451–2471.

26. Google's multilingual neural machine translation system: Enabling zero-shot translation / Johnson M., Schuster M., Le Q.V., et al. // Computational Linguistics. – 2017. – Vol. 5. – P. 339–352. – URL: <https://www.aclweb.org/anthology/Q17-1024.pdf> (дата обращения 01.04.2021).
27. Gradient flow in recurrent nets: The difficulty of learning long-term dependencies / Hochreiter S., Bengio Y., Frasconi P., Schmidhuber J. // A field guide to dynamical recurrent networks / Eds. J.F. Kolen, S. Kremer. – Los Alamitos, CA, USA : IEEE Press, 2001. – P. 1–15. – URL: <http://www.bioinf.jku.at/publications/older/ch7.pdf> (дата обращения 04.04.2021).
28. Hochreiter S., Schmidhuber J. Long short-term memory // Neural Computation. – 1997. – Vol. 9, N 8. – P. 1–32. – URL: <https://www.bioinf.jku.at/publications/older/2604.pdf> (дата обращения 01.03.2021).
29. Machine Translation: the ALPAC report // Pangeanic. – 2013. – 04.07. – URL: <https://pangeanic.com/knowledge/machine-translation-the-alpac-report/> (дата обращения 01.03.2021).
30. Nagao M. A framework of a mechanical translation between Japanese and English by analogy principle, in Artificial and Human Intelligence / ed. A. Elithorn and R. Banerji. – North Holland, 1984. – P. 173–180. – URL: <https://www.bibsonomy.org/bibtex/1aa3312962635f792f88c336a66e2e97e/idsia> (дата обращения 11.06.2021).
31. Schuster M., Nakajima K. Japanese and Korean voice search // IEEE Conference (International) on Acoustics. – Las Vegas, NV, USA : IEEE, 2016. – Vol. 11. – P. 5149–5152. – URL: <https://static.googleusercontent.com/media/research.google.com/ru/pubs/archive/37842.pdf> (дата обращения 02.03.2021).
32. Schuster M., Paliwal K. Bidirectional recurrent neural networks // IEEE on Signal Processing. – 1997. – Vol. 45, N 11. – P. 2673–2681.
33. Sutskever I., Vinyals O., Le Q.V. Sequence to sequence learning with neural networks // 27th Conference (International) on Neural Information Processing Systems Proceedings. – Cambridge, MA, USA : MIT Press, 2014. – Vol. 2. – P. 1–9. – URL: <https://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf> (дата обращения 02.04.2021).

ARTIFICIAL INTELLIGENCE IN MACHINE TRANSLATION TECHNOLOGIES

Colin Konstantin

DrS (Tech. Sci.), Professor, Chief Researcher of the Federal Research Center “Computer Science and Control”, Russian Academy of Sciences, Moscow, Russia

Khoroshilov Alexander

DrS (Tech. Sci.), Professor of the Moscow Aviation Institute (National Research University), Leading Researcher of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, Senior Researcher of the 27 Central Research Institute of the Ministry of Defense of the Russian Federation, Moscow, Russia

Nikitin Yuri

Researcher of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, Development Team Leader of the Scientific and Industrial Company “High Technologies and Strategic Systems”, Moscow, Russia

Pshenichny Sergey

PhD (Econ. Sci.), Program Director of the Scientific and Industrial Company “High Technologies and Strategic Systems”, Moscow, Russia

Khoroshilov Alexey

Ph.D (Engineering Sci.), Senior Researcher of the 27 Central Research Institute of the Ministry of Defense of the Russian Federation, Moscow, Russia

***Abstract.** The capabilities of machine translation are closely related to the improvement of modeling the processes of understanding and generating texts in natural language, which traditionally belongs to the class of artificial intelligence problems. The article attempts to analyze the main approaches to the creation of machine translation technologies. It is concluded that these approaches have not yet provide for the formation and use of dynamic models of the world, but are moving mainly in the direction of a grammatically consistent translation of word sequences.*

***Keywords:** machine translation; natural language; artificial Intelligence; machine translation technologies.*

***For citation:** Artificial intelligence in machine translation technologies / Kolin K.K., Khoroshilov Al-dr A., Nikitin Yu.V., Pshenichny S.I., Khoroshilov Al-ei A. // Social Novelties and Social Sciences. – Moscow : INION RAN, 2021. – N 2. – Pp. 64–80.*

URL: <https://sns-journal.ru/ru/archive/>

DOI: 10.31249/snsn/2021.02.05