

УДК 004+81`32

## ***РОЛЬ ТЕРМИНОЛОГИЧЕСКОГО КОНТРОЛЯ И КОНТЕКСТА ПРИ ИСПОЛЬЗОВАНИИ МАШИННОГО ПЕРЕВОДА***

***Герасимов В.В.***

*магистрант,*

*Калужский государственный университет им. К. Э. Циолковского,*

*Калуга, Россия*

***Новиков И.С.<sup>1</sup>***

*магистрант,*

*Калужский государственный университет им. К. Э. Циолковского,*

*Калуга, Россия,*

### **Аннотация**

В статье рассматриваются технологические основания современного машинного перевода с акцентом на две взаимосвязанные задачи: учет контекста и обеспечение терминологической точности. Показано, что переход от правил и статистических моделей к нейросетевому переводу изменил сам принцип обработки текста: система перестала опираться на жесткие словарные соответствия и начала формировать контекстно-зависимые представления, в которых значение единицы определяется ее окружением. Раскрывается роль механизма внимания и архитектуры «трансформер» в моделировании ближних и дальних зависимостей, а также значение сегментации слов на более мелкие части для работы с редкой лексикой и морфологически сложными формами. Отдельное внимание уделяется документному контексту, необходимому для разрешения местоименной соотнесенности, лексической неоднозначности и поддержания связности перевода.

---

<sup>1</sup> Научный руководитель: к.п.н., доцент Прокофьева О.Н., Калужский государственный университет и. К.Э. Циолковского, Калуга, Россия. PhD, Associate Professor O.N. Prokofieva, Kaluga State University named after K.E. Tsiolkovsky, Kaluga, Russia.

Показано, что даже сильные нейросетевые модели не обеспечивают устойчивую передачу специализированной терминологии без внешнего управления. В связи с этим рассматриваются методы терминологического контроля: глоссарии, терминологические ограничения, аннотирование входных данных и декодирование с ограничениями. Делается вывод, что качество машинного перевода в прикладных сценариях определяется не только мощностью модели, но и способностью системы совместно использовать статистически выявляемый контекст и явно заданные терминологические знания.

**Ключевые слова:** машинный перевод, нейросетевой машинный перевод, архитектура «Трансформер», механизм внимания, контекст, документный контекст, терминологический контроль, декодирование с ограничениями, оценка качества перевода

## ***THE ROLE OF TERMINOLOGY CONTROL AND CONTEXT IN MACHINE TRANSLATION***

***Gerasimov V.V.***

*master's student,*

*Kaluga State University named after K. E. Tsiolkovsky,*

*Kaluga, Russia*

***Novikov I.S.***

*master's student,*

*Kaluga State University named after K. E. Tsiolkovsky,*

*Kaluga, Russia*

### **Abstract**

The article examines the technological foundations of modern machine translation with a focus on two interrelated issues: context modelling and terminological accuracy. It is

shown that the shift from rule-based and statistical systems to neural translation changed the very principle of text processing: instead of relying on rigid lexical correspondences, the system now builds context-dependent representations whose meaning is determined by surrounding units. The paper discusses the role of the attention mechanism and the Transformer architecture in modelling short- and long-range dependencies, as well as the contribution of subword segmentation to processing rare words and morphologically complex forms. Particular attention is paid to document-level context required for pronoun resolution, lexical disambiguation and translation coherence. The article argues that even strong neural models do not guarantee stable rendering of specialised terminology without external control. For this reason, terminology control methods are considered, including glossaries, terminology constraints, source annotation and constrained decoding. It is concluded that machine translation quality in practical settings depends not only on model capacity but also on the ability to combine statistically inferred context with explicitly specified terminological knowledge.

**Key words:** machine translation, neural machine translation, Transformer architecture, attention mechanism, context, document-level context, terminology control, constrained decoding, translation quality evaluation

Современный машинный перевод уже нельзя описывать как простое сопоставление словарных эквивалентов. В центре внимания исследователей и разработчиков сегодня находится вопрос о том, каким образом система устанавливает значение языковой единицы в конкретном окружении и как удерживает это значение при построении перевода. Поэтому две проблемы оказываются особенно важными: учет контекста и терминологическая управляемость. Первая показывает, насколько система способна различать значения в зависимости от синтаксической, семантической и дискурсивной среды.

Вторая связана со стабильностью перевода в специализированных областях, где ошибка в одном термине может изменить смысл целого сообщения [1; 2].

История развития машинного перевода хорошо объясняет, почему именно эти задачи стали центральными. В системах, основанных на правилах, основная надежда возлагалась на грамматические описания и словарные базы. Такой подход позволял видеть, по каким правилам строится перевод, но плохо переносился на новые языки и предметные области. Статистический этап сделал опорой корпуса параллельных текстов и вероятностные соответствия между фразами, однако и в этом случае контекст чаще всего ограничивался коротким отрезком текста, а терминологическая устойчивость достигалась дополнительными таблицами соответствий и наборами эвристик. Переход к нейросетевому переводу изменил саму логику обработки: модель стала обучаться не отдельным соответствиям, а преобразованию одной последовательности в другую на основе больших массивов данных [1; 2; 5; 15].

В работе D. Bahdanau, K. Cho и Y. Bengio было показано, что одного фиксированного представления для всего предложения недостаточно. Если система должна последовательно строить перевод, она должна иметь возможность в каждый момент опираться на наиболее значимые фрагменты исходного предложения [5]. Из этого положения выросла архитектура кодировщика и декодировщика с механизмом внимания. Ее смысл состоит не в символическом «понимании» текста, а в вычислении весов значимости между частями входной и выходной последовательностей. Когда система выбирает очередное слово перевода, она перераспределяет внимание между элементами исходного предложения и тем самым уточняет, какие связи важны в данный момент.

Следующим крупным этапом стала работа A. Vaswani и соавторов «Attention Is All You Need», в которой была предложена архитектура «Трансформер» [15]. В ней отказались от последовательного рекуррентного просмотра слов и сделали

механизм внимания базовым способом вычисления зависимостей между единицами текста. Каждое слово соотносится со всеми остальными словами предложения, поэтому его представление формируется не изолированно, а в системе отношений со всем входом. Для машинного перевода это принципиально, поскольку значение единицы определяется не словарем как таковым, а ее положением в структуре фразы, тематикой высказывания, соседними словами и грамматическими связями. Именно поэтому одна и та же лексема в разных окружениях получает разные векторные представления и может переводиться по-разному.

Важную роль в современных системах играет и сегментация слов на более мелкие части. Переход от цельнословного словаря к более мелким сегментам позволяет заметно повысить устойчивость к редкой лексике, сложным словоформам, именам собственным и новым терминам [2; 12]. Для языков с богатой морфологией это особенно важно: система работает не только с готовыми словоформами, но и с повторяющимися частями слова, из которых может собирать новые комбинации. Однако сама по себе такая сегментация не решает проблему выбора значения. Она лишь делает входное представление более гибким; окончательное решение по-прежнему зависит от того, какие связи система обнаружит внутри предложения или документа.

С вычислительной точки зрения контекст в машинном переводе имеет несколько уровней. Во-первых, это локальный контекст внутри одного предложения, который нужен для разрешения многозначности, выбора порядка слов, грамматического времени, вида и согласования. Во-вторых, это межфразовый, или документный, контекст, включающий предшествующие и последующие предложения. Он необходим для правильной передачи местоимений, эллипсиса, тематической связности и согласованности терминологии. В-третьих, следует выделять доменный контекст, то есть принадлежность текста к определенной

предметной области, где привычные слова получают специализированную интерпретацию [6; 8; 9].

Долгое время большинство систем переводили предложения изолированно, потому что это упрощало обучение и подготовку данных. Однако в работе R. Bawden, R. Sennrich, A. Birch и В. Haddow показано, что именно на границе предложений возникают типичные ошибки, связанные с местоименной соотнесенностью, тематической связностью и выбором варианта перевода для многозначного слова [6; 8; 9]. Близкий вывод содержится в работе Y. Kim, D.T. Tran и H. Ney: дополнительный контекст действительно помогает, но его полезность оказывается избирательной и зависит от конкретного языкового явления [8; 9]. В более позднем исследовании L. Jin, Y. Wang, X. Du и M. Carpuat подчеркивается, что современные системы нередко демонстрируют хороший общий результат, но все еще слабо используют межфразовые связи там, где требуется точная дискурсивная интерпретация [8].

Поэтому было бы некорректно утверждать, что машина «понимает контекст» в человеческом смысле. Корректнее говорить о статистическом приближении зависимостей между фрагментами текста. Чем богаче архитектура, обучающие данные и процедура обучения, тем выше вероятность того, что система уловит важные контекстные соотношения. Но эта зависимость остается вероятностной. На практике это означает, что даже сильная модель может правильно передавать общий смысл абзаца и одновременно ошибаться в выборе конкретного термина, если соответствующее различие редко встречалось в обучающем материале [8; 9].

В специализированных текстах проблема терминологической точности становится самостоятельной инженерной задачей. Пользователь ожидает и смысловой близости, и воспроизведения именно того термина, который принят в организации, отрасли, нормативном документе или программном продукте. Нейросетевая система выбирает наиболее вероятное продолжение, тогда как

прикладной сценарий требует нормативно закреплённого варианта. В работе E. Hasler, A. de Gispert, G. Iglesias и В. Byrne показано, что одних словарных подсказок недостаточно: система должна не только вставить предписанный термин, но и согласовать его с грамматикой и порядком слов целевого языка [2; 7].

В современных исследованиях выделяют несколько путей терминологического контроля. Первый связан с использованием глоссариев и терминологических баз, которые подаются системе до начала перевода. Второй предусматривает аннотирование исходного текста или добавление специальных меток, сообщающих модели, какой вариант должен быть выбран. Третий путь опирается на декодирование с ограничениями, когда при построении перевода система обязана включить заданный термин или исключить нежелательный вариант [7; 14]. Подобные ограничения могут быть встроены в сам процесс генерации, а не только применяться после него [14].

Особенно важным является соотношение контекста и терминологии. На первый взгляд между ними может возникать конфликт: система, ориентированная на плавность и частотность, выбирает наиболее естественный в среднем вариант, а терминологическое ограничение требует менее частотного, но нормативно правильного решения. Однако в хорошо построенной системе эти механизмы не должны противостоять друг другу. Напротив, доменный контекст должен помогать выбрать корректный термин, а терминологическое ограничение должно сужать круг допустимых интерпретаций. Именно поэтому наиболее перспективными считаются подходы, в которых терминологические знания подключаются либо на этапе обучения, либо на этапе управляемого декодирования, а затем результат дополнительно проверяется средствами автоматической верификации [2; 7; 14].

Проблемы контекста и терминологии напрямую влияют на способы оценки качества перевода. В работе K. Papineni, S. Roukos, T. Ward и W.-J. Zhu предложен показатель совпадения фрагментов машинного перевода с эталоном, который

долгое время оставался основным средством сравнения систем [10]. Позднее в работе M. Snover, B. Dorr, R. Schwartz, L. Micciulla и J. Makhoul был предложен другой подход, оценивающий количество правок, необходимых для доведения результата до эталона [13]. Эти методы полезны для общего сопоставления систем, но недостаточно чувствительны к тем явлениям, ради которых и привлекается дополнительный контекст: местоименной соотнесенности, межфразовой связности, последовательности терминологии и выбору между близкими по смыслу вариантами [3; 4; 10; 13].

По этой причине в последние годы усилился интерес к более сложным метрикам, использующим представления исходного текста, гипотезы перевода и эталона. Такие методы лучше согласуются с человеческими оценками, чем многие традиционные показатели [11]. Тем не менее даже они не отменяют необходимости специальных тестов. Если исследователь хочет понять, научилась ли система использовать документный контекст, общий балл может скрыть улучшение на редких, но принципиально важных явлениях. Аналогично и терминологическая точность требует самостоятельной проверки: система может демонстрировать хороший интегральный результат и при этом регулярно нарушать утвержденный терминологический стандарт [3; 11].

Таким образом, современный машинный перевод строится не на прямом словарном сопоставлении, а на формировании контекстно-зависимых представлений, в которых значение единицы определяется ее связями с другими элементами текста. Механизм внимания и архитектура «Трансформер» сделали возможным более гибкое моделирование ближних и дальних зависимостей, а сегментация слов на части повысила устойчивость систем к редкой и морфологически сложной лексике. Однако способность модели извлекать статистический контекст не равна гарантии точного перевода в специализированной коммуникации. Именно поэтому терминологический

контроль остается не вспомогательной, а системообразующей частью прикладного машинного перевода. Он превращает перевод из вероятностной генерации в управляемый процесс, где внешние знания ограничивают и направляют работу модели. Перспективы дальнейшего развития связаны с сочетанием документного контекста, управляемого декодирования и более адресных методов оценки, способных фиксировать не только общую гладкость текста, но и корректность тех языковых решений, которые критичны для профессиональной среды.

### **Библиографический список:**

1. Беляева Л. Н., Камшилова О. Н. Машинный перевод в системе обучения: процедуры и ресурсы // Известия РГПУ им. А. И. Герцена. - 2023. - № 208. - URL: <https://cyberleninka.ru/article/n/mashinnyy-perevod-v-sisteme-obucheniya-protsedury-i-resursy> (дата обращения: 15.04.2026).
2. Беляева Л. Н., Камшилова О. Н. Лексикографические проблемы систем машинного перевода: на пути от буквального до нейронного // Вестник Волгоградского государственного университета. Серия 2, Языкознание. - 2024. - Т. 23. - № 5. - С. 6-19. - URL: <https://cyberleninka.ru/article/n/leksikograficheskie-problemy-sistem-mashinnogo-perevoda-na-puti-ot-bukvalnogo-do-neyronnogo> (дата обращения: 15.04.2026).
3. Нуриев В. А., Егорова А. Ю. Методы оценки качества машинного перевода: современное состояние // Информатика и ее применения. - 2021. - Т. 15. - № 2.- С. 104-111.
4. Рычихин А. К. О методах оценки качества машинного перевода // Системы и средства информатики. - 2019. - Т. 29. - № 4. - С. 106-118.
5. Bahdanau D., Cho K., Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate // arXiv preprint arXiv:1409.0473. - 2014.

6. Bawden R., Sennrich R., Birch A., Haddow B. Evaluating Discourse Phenomena in Neural Machine Translation // Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. - 2018. - P. 1304-1313.
7. Hasler E., de Gispert A., Iglesias G., Byrne B. Neural Machine Translation Decoding with Terminology Constraints // Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers). - 2018. - P. 506-512.
8. Jin L., He J., May J., Ma X. Challenges in Context-Aware Neural Machine Translation // Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. - 2023.- P. 15246-15263.
9. Kim Y., Tran D. T., Ney H. When and Why is Document-Level Context Useful in Neural Machine Translation? // Proceedings of the Fourth Workshop on Discourse in Machine Translation. - 2019. - P. 24-34.
10. Papineni K., Roukos S., Ward T., Zhu W.-J. BLEU: a Method for Automatic Evaluation of Machine Translation // Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. - 2002. - P. 311-318.
11. Rei R., Stewart C., Farinha A. C., Lavie A. COMET: A Neural Framework for MT Evaluation // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). - 2020. - P. 2685-2702.
12. Sennrich R., Haddow B., Birch A. Neural Machine Translation of Rare Words with Subword Units // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. - 2016. - P. 1715-1725.
13. Snover M., Dorr B., Schwartz R., Micciulla L., Makhoul J. A Study of Translation Edit Rate with Targeted Human Annotation // Proceedings of the 7th Conference of the Association for Machine Translation in the Americas. -2006. - P. 223-231.

14. Susanto R. H., Chollampatt S., Tan L. Lexically Constrained Neural Machine Translation with Levenshtein Transformer // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. - 2020. - P. 3536-3543.

15. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser L., Polosukhin I. Attention Is All You Need // Advances in Neural Information Processing Systems. - 2017. - Vol. 30.