

«УТВЕРЖДАЮ»

Проректор по научной работе
Федеральное государственное
бюджетное образовательное
учреждение высшего образования

«Санкт-Петербургский
государственный университет»

С. В. Микушев

2024 г.



ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

Федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет» на диссертацию **Животовой Алены Анатольевны «Математическая модель, алгоритмы и программный комплекс для повышения качества машинного перевода узкоспециальных технических текстов на английский язык»**, представленную на соискание ученой степени кандидата технических наук по специальности **1.2.2. – Математическое моделирование, численные методы и комплексы программ.**

Актуальность темы диссертационной работы. Язык – это очень гибкая система с неограниченным набором размытых правил. На сегодняшний день нейронные сети уже схватывают семантические и синтаксические связи в предложениях и даже распознают акцент говорящего, однако они не умеют учитывать культурологические, когнитивные, литературные и прочие аспекты перевода. В последние годы машинный перевод получил очень широкое распространение. Наиболее известными сервисами машинного перевода являются Google.Translate и Translate.Yandex. Однако машинный перевод все равно приходится редактировать человеку, а для этого нужно замечать и исправлять ошибки,

сделанные машиной. Это отдельный трудоемкий процесс, который требует специфического навыка и временных ресурсов.

Качество машинного перевода измеряется на основании различных метрик, таких как ROUGE (Recall-Oriented Understudy for Gisting Evaluation), METEOR (Metric for Evaluation of Translation with Explicit ORdering), BLEU (BiLingual Evaluation Understudy) и TER (Translation Edit Rate).

Для части текстов возможно формализованное требование к качеству перевода, однако разработка такого подхода тоже отдельная задача.

Практическая значимость и актуальность исследуемой темы обусловлены развитием контентной экономики, в условиях которой время на генерацию и обновление информации сократилось с месяцев до нескольких дней. В дополнение, необходима локализация информации в режиме реального времени, что создает необходимость в оптимизации временных и финансовых ресурсов для постоянного перевода больших объемов текстовых данных, при этом обеспечивая высокое качество перевода, особенно в специфических технических областях, где в открытых источниках недостаточно обучающих данных. Применение вероятностного анализа сложности перевода и алгоритмов предредактирования текста помогает уменьшить влияние человеческого фактора на качество машинного перевода. Это также дает возможность сформировать правильные критерии для создания стратегии управления рисками, связанными с квалификацией исполнителей и пользователей систем машинного перевода, в процессе реализации крупных переводческих проектов.

Таким образом, тема диссертационной работы Животовой А.А. является актуальной.

Новизна исследования и полученных результатов, выводов и рекомендаций, сформулированных в диссертации.

Идея диссертационной работы заключается в применении характеристик функционирования алгоритмов машинного перевода и основ теории перевода для автоматизации предварительного редактирования

исходных текстов. Это позволит улучшить их структуру, что, в свою очередь, повысит эффективность систем машинного перевода при переводе на нужный язык и уменьшит количество стилистических ошибок. Эти ошибки требуют от пользователя более высокой языковой компетенции для их выявления.

Научную значимость представляют предложенные, и разработанные автором:

1. методика для повышения качества машинного перевода текстов с русского языка на английский язык, отличающаяся от существующих применением обратного перевода для сбора тренировочных данных и оптимизационного предредактирования на основе вероятностной оценки сложности задачи перевода;

2. методика оценки сложности переводческой задачи для переводчика на основе его компетенции и специализации и параметров исходного текста, которая позволяет прогнозировать риски некачественного и/или несвоевременного решения задачи перевода;

3. новый алгоритм, позволяющий расширить область применения оптимизационного метода градиентного спуска путём использования элементов нечеткой логики в выражении функции правдоподобия через функцию принадлежности полученного текста низкой сложности задачи перевода для выбранной системы машинного перевода;

4. новый алгоритм, позволяющий расширить область применения метода наименьших квадратов для поиска весов значимости параметров исходного текста для вероятностной оценки ожидаемого качества его перевода на целевой язык;

5. новая архитектура и программный комплекс для повышения качества машинного перевода текстов с русского языка на английский язык, отличающийся от существующих применением ансамбля моделей для

оптимизационного предредактирования на основе вероятностной оценки сложности задачи перевода с целью повышения качества машинного перевода текстов с русского языка на английский язык.

В главе 1 представлено исследование текущих работ по автоматизированной предобработке исходных текстов для машинного перевода, которое выявило несколько важных аспектов, соответствующих ключевым направлениям развития систем машинного перевода. В частности, обнаружено следующее:

- отсутствуют исследования, посвящённые анализу ошибок в нейронном машинном переводе с русского языка, которые могли бы быть связаны с особенностями определённых предметных областей или с языковыми нюансами;

- существующие исследования зачастую не имеют системного подхода, несмотря на то, что для русского языка разработаны синтаксические анализаторы и возможно создание правил для автоматического предредактирования с целью повышения качества перевода;

- в научной литературе лишь в малой степени рассматриваются правила и алгоритмы автоматического предредактирования, нацеленные на улучшение качества перевода, и имеющиеся прикладные исследования преимущественно ориентированы на адаптацию текстов для получателей с различной квалификацией;

- существующие методы автоматизации предредактирования ограничены в своих возможностях из-за проблем со сбором и подготовкой обучающих данных.

В главе 2 был обоснован подход к оценке сложности переводческой задачи и разработана соответствующая методология. Результаты проведенного моделирования демонстрируют, что уже на этапе анализа исходного текста можно заранее оценить ожидаемое качество перевода,

учитывая параметры текста, а также знания и специализацию переводчика.

Впервые представлена математическая модель для расчета сложности переводческой задачи, и предложено использовать полученные данные в качестве критерия для оптимизации на этапе предредактирования текста перед машинным переводом. Подход к задаче оптимизационного предредактирования расширяет применение метода градиентного спуска, внедряя элементы нечеткой логики в формализацию функции потерь.

В главе 3 предложен алгоритм, позволяющий оценить сложность переводческой задачи исходя из компетенции и специализации переводчика, а также характеристик оригинального текста в рамках разработки методов обучения моделей для оптимизационного редактирования текстов на русском языке. Этот алгоритм помогает предсказывать риски, связанные с низким качеством или задержкой выполнения перевода. Также был разработан новый метод оценки русскоязычного текста, основывающийся на лексических, синтаксических и морфологических характеристиках. Кроме того, представлена новая методика обучения модели редактирования русскоязычных текстов, которая отличается от существующих благодаря использованию обратного перевода для формирования обучающих данных и критерия оптимизации для улучшения качества машинного перевода на английский язык.

В главе 4 представлен программный комплекс оптимизационного предредактирования узкоспециализированных русскоязычных текстов для их перевода на английский язык. Результаты тестирования программного комплекса подтверждают работоспособность программного комплекса в части повышения качества машинного перевода узкоспециальных технических текстов путем автоматического оптимизационного предредактирования и использования оценки сложности задачи перевода в качестве критерия оптимизации. Интеграция созданной системы оптимизационного предредактирования в контур автоматизации процессов

перевода технической документации позволит снизить затраты на постредактирование машинного перевода и организацию переводческих процессов. Отражено внедрение программного комплекса в работу Агентства переводов, что позволило повысить эффективность использования машинного перевода и производительность труда редакторов переводов, оптимизировав тем самым затраты на оказание услуг перевода узкоспециальной технической документации.

Автор имеет 6 публикаций в изданиях, рекомендуемых ВАК, 10 свидетельств о государственной регистрации программ для ЭВМ и БД.

Значимость для науки и практики полученных результатов

Предложенные методики, алгоритмы и программные комплексы могут быть адаптированы для работы с различными языковыми парами и методами перевода, включая ручной перевод. Они предлагают стратегии управления рисками, связанными с качеством перевода, которые зависят от квалификации выбранных специалистов. Кроме того, эти решения предоставляют отрасли инструменты для объективной оценки исполнителей в контексте заданной переводческой задачи, автоматизации подготовки текстов к переводу и улучшения качества перевода. Это особенно полезно для редакторов, которые не владеют языком оригинала.

Конкретные рекомендации по использованию результатов и выводов диссертационной работы

Результаты диссертационного исследования рекомендуются к использованию в научных организациях и в организациях, занимающихся переводами узкоспециальных текстов, для оптимизации процесса перевода текста и проведения научно-исследовательских работ, направленных на внедрение цифровых технологий в перевод текстов.

Замечания по работе.

- Научный обзор не содержит некоторые из последних работ в области text simplification (для machine translation):

1. Tong Wang, Ping Chen, John Rochford, Jipeng Qiang, Text Simplification Using Neural Machine Translation, Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16).
 2. Sanja STAJNER, Maja POPOVIC, Can Text Simplification Help Machine Translation?, Baltic J. Modern Computing, Vol. 4 (2016), No. 2, pp. 230–242.
 3. Wei Xu, Ellie Pavlick, Quanze Chen, Chris Callison-Burch, Optimizing Statistical Machine Translation for Text Simplification, Transactions of the Association for Computational Linguistics, Volume 4, 2016.
 4. Suha S. Al-Thanyyan, Aqil M. Azmi, Automated Text Simplification: A Survey, ACM Computing Surveys, Vol. 54, No. 2.
 5. Yanbin Zhao, Lu Chen, Zhi Chen, Kai Yu, Semi-Supervised Text Simplification with Back-Translation and Asymmetric Denoising Autoencoders, Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34 No. 05: AAAI-20 Technical Tracks 5, 2020.
- В работе не представлено сравнение предложенного подхода с существующими (необязательно на основе перевода с русского на английский язык) для упрощения исходного текста для последующего перевода.

Заключение. Диссертационное исследование **Животовой Алены Анатольевны «Математическая модель, алгоритмы и программный комплекс для повышения качества машинного перевода узкоспециальных технических текстов на английский язык»** является законченной научно-квалификационной работой, в которой содержится новое решение научной задачи повышения качества машинного перевода узкоспециальных технических текстов путем автоматического оптимизационного предредактирования, имеющей важное значение для перевода узкоспециализированных текстов и разработки языковых моделей искусственного интеллекта. Работа соответствует требованиям п. 9 «Положения о порядке присуждения ученых степеней», утвержденного постановлением Правительства Российской Федерации от 24.09.2013 г. №

842 (в редакции Правительства Российской Федерации от 21.04.2016 г. № 335; от 02.08.2016 г. № 748; от 01.10.2018 г. № 1168), предъявляемым к диссертациям на соискание ученой степени кандидата технических наук, а её автор, **Животова Алена Анатольевна**, заслуживает присуждения ученой степени **кандидата технических наук** по специальности **1.2.2. – Математическое моделирование, численные методы и комплексы программ.**

Отзыв подготовлен д. ф.-м. наук, профессором Петросян Ованесом Леоновичем.

Отзыв обсужден и одобрен на заседании кафедры математического моделирования энергетических систем 16.09.2024, протокол № 44/4/6-02-3.

И.о. зав. каф. математического моделирования
энергетических систем,
профессор, д. ф.-м. н.

Парилина Елена Михайловна

Профессор, д. ф.-м. н.

Петросян Ованес Леонович

Подписи заверяю:

И.о. начальника
отдела кадров № 3
И.И. Константинова



Федеральное государственное бюджетное образовательное
учреждение высшего образования «Санкт-Петербургский государственный
университет»

Адрес: 199034, Санкт-Петербург, Университетская набережная, д. 7-9.

Телефон (812) 328-97-01

E-mail: spbu@spbu.ru