

ПЛЕНАРНЫЕ ДОКЛАДЫ

ИСПОЛЬЗОВАНИЕ САЙТА NAVR ДЛЯ ЛИНГВОДИДАКТИЧЕСКОЙ АДАПТАЦИИ ТЕКСТОВ БУДУЩИХ IT-СПЕЦИАЛИСТОВ

Сычёва Е.С.

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

В статье рассматривается сайт Navr как ресурс для обучения и адаптации русскоязычных текстов для иностранных студентов и будущих IT-специалистов. Автор предлагает методику лингводидактической адаптации учебно-научных текстов, направленную на сохранение смысла исходной статьи и её переработку в научный стиль речи. Приведен пример лингводидактической адаптации фрагмента статей сайта Navr.

Занятия по РКИ на старших курсах обычно предполагают более глубокое изучение языка и его использование в профессиональной сфере. Поэтому крайне важно использовать актуальные тексты профессиональной или научно-популярной направленности, так как они помогают студентам лучше понять терминологию и специфику своей области знаний на изучаемом языке. Сайт Navr является отличным ресурсом для обучения и адаптации русскоязычных текстов для иностранных студентов и будущих IT-специалистов. На данном сайте публикуются статьи на различные темы, связанные с информационными технологиями, что делает его идеальным источником для изучения и адаптации текстов.

А. В. Коротышев отмечает, что «целью лингводидактической адаптации текста художественного произведения является его трансформация в соответствии с новыми, учебными, условиями функционирования, которые характеризуются целым набором факторов, связанных с личностью читателя-инофона – прежде всего, с уровнем сформированности его языковой компетенции, кругом интересов, читательским опытом, опытом изучения других иностранных языков и др.» [1, с.5]. Мы считаем, что данный приём подходит для адаптации не только художественного текста, но и учебно-научного. При этом необходимо учитывать следующие аспекты: соответствие уровня языковой подготовки, подходящие тексты и их последующая обработка с целью облегчения понимания и усвоения материала. В контексте IT-специальности это может включать в себя тексты на темы, такие как программирование, системное администрирование, разработка веб-сайтов, искусственный интеллект и т.д.

Статьи на сайте Navr часто написаны профессионалами в своей области, что обеспечивает высокое качество и достоверность. Однако, несмотря на информативность и научную составляющую, стиль речи статей на данном сайте написан простым и доступным языком, ориентированным на широкий круг читателей. В публикациях мы можем наблюдать элементы публицистического и разговорного стиля речи, сленг, эмоциональность и образность. Поэтому мы считаем целесообразным видоизменять текст в научный стиль речи при проведении лингводидактической адаптации для занятий по РКИ. Для обучающихся старших курсов технического университета с высоким уровнем владения языком следует проводить изменение стиля речи статьи, направленное на сохранение смысла исходного текста. При адаптации мы должны учитывать следующие особенности: использование третьего лица, логичность изложения, сохранение узкоспециальных терминов, отбор языковых средств, сложное построение предложений.

Приведем пример лингводидактической адаптации фрагмента статьи «Сравнение нейросетей для перевода».

Оригинальный текст: «Я также пробовал использовать few-shot примеры. Оказалось, что на этой задаче добавление примеров чаще всего ухудшает метрики. Да, можно перебирать наборы примеров и найти те, которые улучшают, и выбрать среди них оптимальный. Однако, это больше похоже на переобучение модели на стиль Tatoeba-предложений. Поэтому я решил не использовать примеры, промт состоит только из системного сообщения и сообщения для перевода. Такое решение, вдобавок, потребляет гораздо меньше токенов и поэтому дешевле.» [2].

Адаптированный вариант текста: «Автору удалось провести эксперименты с использованием небольшого числа примеров для обучения. Результаты показали, что в данном случае добавление дополнительных примеров чаще всего приводит к ухудшению показателей. Однако найти некоторые наборы примеров, которые улучшают результаты, выбор оптимального набора остается сложной задачей. Заметно, что подбор примеров может привести к переобучению модели под стиль предложений из Tatoeba. Поэтому было принято решение не использовать дополнительные примеры, а ограничиться лишь системным сообщением и сообщением для перевода. Этот подход также позволяет сэкономить ресурсы токенов и сделать процесс более эффективным с точки зрения затрат».

Рассмотрим фрагмент текста из статьи «YandexGPT в Алисе: как мы создаём языковую модель нового поколения».

Оригинальный текст: «Новые требования к Pretrain. Всё, что будет знать и уметь модель, закладывается в неё на этапе Pretrain. Раньше мы фокусировались на размере Pretrain-датасета, считая, если положить в него много данных, то и модель многое выучит. Сейчас же мы сфокусировались именно на качестве данных в датасете, это позволило значительно улучшить качество итоговой модели. Под качеством мы понимаем два важных требования к датасету.» [3].

Адаптированный вариант текста: «В процессе Pretrain модель овладевает знаниями и навыками, которые в неё вложены. Ранее в исследовании акцент был сделан на объёме данных в Pretrain-датасете, предполагая, что большое количество информации позволит модели научиться многому. Тем не менее в настоящее время основное внимание уделяется качеству данных в датасете, что существенно повышает качество итоговой модели. Под качеством данных мы подразумеваем два важных аспекта, которыми должен обладать датасет».

Таким образом, чтение статей с Habr поможет студентам улучшить свои навыки владения языком и понимание специфических терминов и технической лексики, которые часто используются в IT-сфере. Поэтому использование сайта Habr для обучения и адаптации текстов на русском языке является эффективным способом для подготовки будущих IT-специалистов к работе в индустрии и расширения их профессиональных навыков.

Список использованных источников:

1. Коротышев А. В. *Технология отбора и лингводидактической адаптации художественных текстов в целях обучения русскому языку как иностранному: автореферат диссертации ... кандидата педагогических наук: 13.00.02 / А. В. Коротышев.* — Москва, 2017. — 27 л.
2. Сравнение нейросетей для перевода [Электронный ресурс]. — Режим доступа: <https://habr.com/ru/articles/737950/jazyk>. — Дата доступа: 10.04.2024.
3. YandexGPT в Алисе: как мы создаём языковую модель нового поколения [Электронный ресурс]. — Режим доступа: <https://habr.com/ru/companies/yandex/articles/739626/>. — Дата доступа: 03.04.2024.