

ИСПОЛЬЗОВАНИЕ ИНСТРУМЕНТОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ СОПОСТАВИТЕЛЬНОГО АНАЛИЗА ВИДО- ВРЕМЕННЫХ ГЛАГОЛОВ В РУССКОМ И АНГЛИЙСКОМ ЯЗЫКАХ В ЦЕЛЯХ АВТОМАТИЗИРОВАННОГО ОБУЧЕНИЯ

Муртазина Алина Рамильевна

Студентка 1 курса магистратуры НПУУ имени Низами (Ташкент)

Аннотация: В статье рассматривается инновационный подход к сопоставительному анализу видо-временных систем русского и английского языков с использованием современных инструментов искусственного интеллекта (ИИ) и технологий обработки естественного языка (NLP).

Целью исследования является разработка методики автоматизированного выявления межъязыковых соответствий для создания адаптивных учебных материалов.

Методология базируется на использовании нейросетевых моделей (XLM-RoBERTa, LaBSE), библиотек компьютерной лингвистики (spaCy) и больших языковых моделей (GPT-4) для анализа параллельных корпусов объемом 500 000 словоупотреблений. В работе применяются методы машинного обучения (кластеризация k-means, логистическая регрессия) для систематизации грамматических паттернов и прогнозирования зон межъязыковой интерференции.

Результаты исследования демонстрируют высокую точность автоматической аннотации (до 96,2%) и позволяют количественно описать устойчивые корреляции между категориями вида в русском языке и аспектуально-временными формами английского языка. Авторами разработан и апробирован прототип системы, генерирующей персонализированные учебные задания, релевантность которых подтверждена экспертной оценкой (коэффициент Коэна 0,84).

Практическая значимость работы заключается в возможности интеграции предложенных алгоритмов в интеллектуальные обучающие платформы для автоматизации проверки работ и создания индивидуальных образовательных траекторий.

Ключевые слова: Искусственный интеллект в образовании, обработка естественного языка, сопоставительная лингвистика, видо-временная система глагола, автоматизированное обучение языкам, параллельные корпуса, межъязыковая интерференция, машинное обучение, генерация учебного контента.

Abstract: The article considers an innovative approach to the comparative analysis of the Russian and English language systems using modern artificial intelligence (AI) tools and natural language processing (NLP) technologies.

The purpose of the research is to develop a methodology for the automated identification of interlanguage correspondences for the creation of adaptive educational materials.

The methodology is based on the use of neural network models (XLM- RoBERTa, LaBSE), computational linguistics libraries (spaCy) and large language models (GPT-4) to analyze parallel corpora of 500,000 word usage. The work uses machine learning methods (k-means clustering, logistic regression) to systematize grammatical patterns and predict areas of interlanguage interference.

The results of the study demonstrate high accuracy of automatic annotation (up to 96.2%) and allow us to quantify stable correlations between the categories of the type in the Russian language and the aspectual-temporal forms of the English language. The authors have developed and tested a prototype system that generates personalized learning tasks, the relevance of which is confirmed by an expert assessment (Cohen coefficient 0.84).

The practical significance of the work lies in the possibility of integrating the proposed algorithms into intelligent learning platforms to automate the verification of work and create individual educational trajectories.

Keywords: Artificial intelligence in education, natural language processing, comparative linguistics, verb type-time system, automated language teaching, parallel corpora, interlanguage interference, machine learning, educational content generation.

Annotatsiya: Maqolada zamonaviy sun'iy intellekt (SI) vositalari va tabiiy tilni qayta ishlash (NLP) texnologiyalaridan foydalangan holda rus va ingliz tillarining zamon-mayl tizimlarini qiyosiy tahlil qilishning innovatsion yondashuvi ko'rib chiqiladi.

Tadqiqotning maqsadi moslashuvchan o'quv materiallarini yaratish uchun tillararo mosliklarni avtomatlashtirilgan tarzda aniqlash metodikasini ishlab chiqishdan iborat. Ishda grammatik naqshlarni tizimlashtirish va tillararo aralashuv zonalarini bashorat qilish uchun mashinaviy o'qitish usullari (k-means klasterlash, logistik regressiya) qo'llaniladi. Ishning amaliy ahamiyati taklif etilgan algoritmlarni ishlarni tekshirishni avtomatlashtirish va individual ta'lim yo'llarini yaratish uchun intellektual o'qitish platformalariga integratsiya qilish imkoniyatidadir.

Kalit so'zlar: Ta'limda sun'iy intellekt, tabiiy tilni qayta ishlash, qiyosiy tilshunoslik, fe'lning zamon-mayl tizimi, tillarni avtomatlashtirilgan o'qitish, parallel korpuslar, tillararo interferentsiya, mashinaviy o'qitish, o'quv mazmunini yaratish. nini beradi. Mualliflar tomonidan shaxsiylashtirilgan o'quv topshiriqlarini yaratuvchi tizim prototipi ishlab chiqilgan va sinovdan o'tkazilgan bo'lib, ularning dolzarbligi ekspert bahosi bilan tasdiqlangan (Koen koeffitsienti 0,84).

Ishning amaliy ahamiyati taklif etilgan algoritmlarni ishlarni tekshirishni avtomatlashtirish va individual ta'lim yo'llarini yaratish uchun intellektual o'qitish platformalariga integratsiya qilish imkoniyatidadir.

ВВЕДЕНИЕ

Актуальность исследования. В эпоху цифровизации образования одной из ключевых задач является разработка интеллектуальных систем, способных адаптироваться к индивидуальным потребностям обучающихся. Особую сложность в преподавании русского и английского языков как иностранных представляет усвоение видо-временной системы глаголов, поскольку в русском языке доминирует категория вида (совершенный / несовершенный вид), а в английском - система из 16 временных форм, передающих тонкие оттенки временной отнесенности [1]. Сопоставительный анализ этих систем традиционно проводится вручную, что трудоемко и не позволяет обрабатывать большие объемы данных для создания адаптивных учебных материалов.

Современные инструменты искусственного интеллекта (ИИ), такие как модели обработки естественного языка (NLP), нейросетевые архитектуры Transformer и корпусные менеджеры, открывают новые возможности для автоматизации сопоставительного лингвистического анализа [2]. Их применение позволяет не только выявлять системные соответствия и расхождения между двумя языками, но и создавать на основе этих данных интеллектуальные обучающие системы, которые генерируют персонализированные упражнения, анализируют ошибки и прогнозируют зоны межъязыковой интерференции.

Цель исследования-разработать и апробировать методику использования инструментов ИИ для сопоставительного анализа видо- временных глаголов в русском и английском языках, ориентированную на создание автоматизированных обучающих материалов.

Задачи исследования:

1. Выявить ключевые параметры сопоставления видо-временных систем, значимые для автоматизированного обучения.
2. Отобрать и адаптировать инструменты ИИ (корпусные технологии, модели машинного обучения) для анализа параллельных текстов.
3. Провести экспериментальный сопоставительный анализ на материале двуязычного корпуса.
4. Разработать прототип автоматизированной системы генерации учебных заданий на основе полученных данных.

МЕТОДЫ

Общая методология. Исследование базируется на принципах корпусной лингвистики, сравнительно-сопоставительного метода и методов машинного обучения. Работа выполнена в междисциплинарном ключе на стыке лингвистики, компьютерных наук и педагогики.

Инструменты ИИ и программное обеспечение. Для реализации поставленных задач был использован следующий инструментарий: библиотека Hugging Face Transformers с моделью XLM-RoBERTa для извлечения признаков и морфологической аннотации параллельных текстов[3]; библиотека spaCy с моделями en_core_web_sm и ru_core_news_sm для токенизации, определения частей речи и маркировки грамматических категорий [4]; язык программирования Python с библиотекой scikit-learn для кластеризации видо-временных форм и выявления корреляций; корпуса COCA (Corpus of Contemporary American English) и Russian National Corpus для сбора репрезентативных примеров, используемых для обучения моделей; а также API OpenAI GPT-4 для генерации учебных заданий и объяснений правил [5].

Материал исследования. В качестве эмпирической базы использован параллельный подкорпус текстов объёмом 500 000 словоупотреблений (250 000 на русском языке, 250 000 на английском языке), включающий художественные тексты с параллельным переводом, учебные тексты уровней A2–C1 по общеевропейской шкале владения иностранным языком, а также синтезированные предложения, сгенерированные для покрытия всех 16 времен английского языка и всех видовременных форм русского языка.

Процедура анализа. На первом этапе тексты обрабатывались с помощью spaCy и XLM-RoBERTa для получения морфологической разметки. Для русского языка выделялись следующие категории: глагол, вид (совершенный / несовершенный), время, лицо, число. Для английского языка выделялись: глагол, время (Present / Past / Future), аспект (Simple / Continuous / Perfect / Perfect Continuous). На втором этапе на уровне предложения устанавливались соответствия между русскими глагольными формами и английскими эквивалентами в параллельных текстах. Для этого использовался алгоритм выравнивания на основе двуязычного эмбединга LaBSE (Language-agnostic BERT Sentence Embedding) [6]. На третьем этапе с помощью метода k-средних (k-means) были выделены устойчивые паттерны соответствий между русскими и английскими видо-временными формами. На четвёртом этапе на основе выявленных паттернов был разработан скрипт на Python, автоматически формирующий упражнения типа «выберите правильную форму», «трансформация предложения», «заполнение пропусков» с персонализацией по уровню сложности.

РЕЗУЛЬТАТЫ

В ходе экспериментального сопоставительного анализа с использованием инструментов ИИ были получены следующие результаты.

1. Систематизация соответствий. На основе кластеризации выявлено 12 устойчивых типов соответствий между русскими и английскими видо- временными формами. Наиболее частотные соответствия распределились следующим образом. Русский глагол несовершенного вида в форме настоящего времени в 54 процентах случаев соответствовал английскому Present Simple и в 41 проценте случаев - Present Continuous. Русский глагол совершенного вида в форме будущего времени в 78 процентах случаев соответствовал английскому Future Simple и в 12 процентах случаев - Present Simple. Русский глагол несовершенного вида в форме прошедшего времени в 52 процентах случаев соответствовал английскому Past Continuous и в 43 процентах случаев - Past Simple. Русский глагол совершенного вида в форме прошедшего времени в 91 проценте случаев соответствовал английскому Past Simple.

2. Оценка точности ИИ-инструментов. Точность автоматической морфологической аннотации составила для английского языка 96,2 процента при использовании библиотеки spaCy, для русского языка - 93,7 процента при использовании spaCy совместно с XLM-RoBERTa с дообучением на учебном корпусе. Основные ошибки были связаны с многозначностью видовых значений в русском языке и сложными временными конструкциями в английском языке, такими как формы Perfect Continuous.

3. Автоматизированная генерация учебных материалов. Разработанный скрипт на Python сгенерировал 1500 учебных заданий, распределенных по уровням сложности от A2 до C1. Проведена экспертная оценка двумя преподавателями русского и английского как иностранных, имеющими стаж работы более 10 лет. Эксперты оценивали релевантность заданий, корректность формулировок и соответствие заявленному уровню сложности. Коэффициент согласия экспертов (Cohen's Kappa) составил 0,84, что свидетельствует о высокой степени согласованности оценок и пригодности сгенерированных материалов для использования в учебном процессе [7].

Прогнозирование зон интерференции. На основе выявленных паттернов была построена модель логистической регрессии, предсказывающая вероятность ошибки носителя русского языка при употреблении английских времен. На тестовой выборке, составившей 20 процентов от общего объёма корпуса, точность прогноза составила 81,3 процента. Наиболее частотные ошибки прогнозируются при использовании Present Perfect вместо Past Simple, Past Continuous вместо Past Simple, а также Present Simple вместо Present Continuous.

ОБСУЖДЕНИЕ

Полученные результаты подтверждают, что инструменты искусственного интеллекта могут эффективно применяться для сопоставительного анализа грамматических систем в целях автоматизированного обучения.

В отличие от традиционных методов, основанных на интроспекции и ограниченных выборках, корпусный подход с использованием технологий обработки естественного языка позволяет обрабатывать большие объёмы аутентичных текстов, получать статистически значимые данные о частотности соответствий и выявлять неочевидные паттерны, такие как использование Present Simple для передачи русского совершенного вида будущего времени в определённых контекстах [8].

Ограничения исследования. Несмотря на высокую точность автоматической аннотации, остаются проблемные зоны. К ним относятся омонимия видовых форм в русском языке, в частности двувидовые глаголы типа женить, казнить, организовать, грамматическая категория которых определяется только контекстом. Также сложность представляют модальные и фазовые глаголы, меняющие интерпретацию временной отнесенности, и стилистическая вариативность между разговорной и письменной речью. Эти ограничения требуют дальнейшего совершенствования моделей, в частности дообучения на специализированных подкорпусах с ручной разметкой.

Практическая значимость. Разработанная методика может быть внедрена в интеллектуальные обучающие платформы с адаптивной траекторией обучения, в системы автоматической проверки письменных работ с обратной связью по ошибкам, связанным с видо-временными формами, а также в генераторы учебных материалов для преподавателей, позволяющие быстро создавать персонализированные упражнения.

Перспективы дальнейших исследований. Видится целесообразным расширить языковую базу, включив узбекский язык как язык с агглютинативной структурой и отсутствием категории грамматического вида, для создания трехязычной модели сопоставительного анализа. Также перспективным представляется использование более мощных мультимодальных моделей, таких как GPT-4V и Gemini, для анализа контекстуальных факторов, влияющих на выбор видо-временной формы, включая визуальный контекст и прагматические особенности коммуникативной ситуации. Кроме того, необходима организация педагогического эксперимента для оценки эффективности сгенерированных искусственным интеллектом материалов в реальном учебном процессе с участием контрольной и экспериментальной групп.

ЗАКЛЮЧЕНИЕ

В ходе исследования была разработана и апробирована методика использования инструментов искусственного интеллекта для сопоставительного анализа видо-временных глаголов в русском и английском языках. Применение моделей обработки естественного языка, корпусных технологий и методов машинного обучения позволило выявить количественные паттерны соответствий между видо-временными системами двух языков,

автоматизировать процесс генерации учебных заданий с высокой степенью релевантности и создать модель прогнозирования типичных ошибок, обусловленных межъязыковой интерференцией.

Полученные результаты демонстрируют потенциал искусственного интеллекта как эффективного инструмента в области автоматизированного обучения языкам. Предложенный подход может быть масштабирован на другие языковые пары и грамматические явления, что открывает перспективы для создания полноценных интеллектуальных систем поддержки языкового образования.

СПИСОК ЛИТЕРАТУРЫ

- [1] Плунгян В.А. Введение в грамматическую семантику: грамматические значения и грамматические системы языков мира. М.: РГГУ, 2011. 672 с.
- [2] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I. Attention Is All You Need // Advances in Neural Information Processing Systems. 2017. Vol. 30. P. 5998–6008.
- [3] Conneau A., Khandelwal K., Goyal N., Chaudhary V., Wenzek G., Guzmán F., Grave E., Ott M., Zettlemoyer L., Stoyanov V. Unsupervised Cross-lingual Representation Learning at Scale // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020. P. 8440–8451.
- [4] Honnibal M., Montani I. spaCy 2: Natural Language Understanding with Bloom Embeddings, Convolutional Neural Networks and Incremental Parsing. 2017. URL: <https://spacy.io> (дата обращения: 15.03.2026).
- [5] OpenAI. GPT-4 Technical Report. 2023. 100 p.
- [6] Feng F., Yang Y., Cer D., Arivazhagan N., Wang W. Language-agnostic BERT Sentence Embedding // Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. 2022. P.878–891.
- [7] Cohen J.A Coefficient of Agreement for Nominal Scales // Educational and Psychological Measurement.1960. Vol.20.No.1. P.37–46.
- [8] Biber D.,Reppen R. The Cambridge Handbook of English Corpus Linguistics. Cambridge: Cambridge University Press, 2015. 623 p.