

Горожанов Алексей Иванович
доктор филологических наук,
профессор кафедры грамматики
и истории немецкого языка факультета
немецкого языка
Московский государственный
лингвистический университет
Москва, Россия

Alexey Gorozhanov
Doctor of Science (Philology)
Professor in the Department
of German Language Grammar and History
Moscow State Linguistic University
Moscow, Russian
a_gorozhanov@mail.ru

Гусейнова Иннара Алиевна
доктор филологических наук,
профессор кафедры лексикологии
и стилистики немецкого языка факультета
немецкого языка
Московский государственный
лингвистический университет
Москва, Россия

Innara Guseynova
PhD in Philology
Professor in the Department of German
Language Lexicology and Stylistics
Moscow State Linguistic University
Moscow, Russian Federation
ginnap@mail.ru

ПРИМЕНЕНИЕ ЭЛЕМЕНТОВ ТЕХНОЛОГИИ УПРАВЛЯЕМОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ НАПОЛНЕНИЯ ОНЛАЙН-КУРСА ИНОСТРАННОГО ЯЗЫКА

В исследовании ставится цель экстраполировать возможности программного комплекса «Генератор сбалансированного лингвистического корпуса и корпусный менеджер» (Свидетельство Роспатент № 2023683209), созданного в лаборатории фундаментальных и прикладных проблем виртуального образования Московского государственного лингвистического университета, на предметную область обучения иностранным языкам в части генерации электронных учебных материалов. При этом происходит чередование условно «простых» программных алгоритмов и алгоритмов с использованием технологий искусственного интеллекта. В результате программный код корпусного менеджера был дополнен модулем генерации тестовых заданий для LMS Moodle, которые представляют

собой упражнение на пропуски глаголов в немецком тексте. Таким образом, создана технология, которая позволяет получать в полностью автоматическом режиме большое количество типовых тренировочных заданий на основе текстового материала предварительно загруженного лингвистического корпуса.

Ключевые слова: управляемый искусственный интеллект; корпусная лингвистика, корпусный менеджер; обработка естественного языка; системы управления обучением, тестовые задания.

APPLICATION OF THE ELEMENTS OF CONTROLLED ARTIFICIAL INTELLIGENCE TECHNOLOGY TO COMPLETE AN ONLINE FOREIGN LANGUAGE COURSE

The research aims to extrapolate the capabilities of the software package “Balanced Linguistic Corpus Generator and Corpus Manager” (Rospatent Certificate No. 2023683209), created at the Laboratory for Fundamental and Applied Issues of Virtual Education at Moscow State Linguistic University, to the subject area of teaching foreign languages in terms of generating electronic learning materials. In this case, there is an alternation of so-called “simple” software algorithms and algorithms using artificial intelligence technologies. As a result, the corpus manager programming code was supplemented with a module for generating tests for LMS Moodle, which are exercises on missing verbs in German text. Thus, a technology has been created that makes it possible to obtain, in a fully automatic mode, a large number of typical training tasks based on text material from a preloaded linguistic corpus.

Key words: controlled artificial intelligence; corpus linguistics; corpus manager; natural language processing; learning management systems; tests.

Мы наблюдаем, как технологии искусственного интеллекта активно входят во все сферы нашей жизни, что, с одной стороны, является требованием времени, а, с другой стороны, вызывает ряд опасений по поводу безопасности системы «человек – машина» [Гринин и др., 2023]. Тем не менее, прогресс остановить невозможно, а его отрицание не является лучшим вариантом поведения человека.

В сферу образования технологии искусственного интеллекта входят если не так активно, как в экономику, то однозначно так же неизбежно. В этой связи необходимо не бороться с явлением, а изучить его и направить на создание максимальной пользы для дела.

Исследователи, лингвисты и педагоги, обращаясь к теме искусственного интеллекта, чаще всего рассматривают такие программные продукты, как ChatGPT и ему подобные, формируя запросы и анализируя полученную реакцию, т. е. изучают программный продукт не столько как инструмент, а как объект исследования, не понимая, в сущности, как он работает (что, впрочем, неосуществимо в полной мере, так как алгоритмы работы подобных программ строго защищаются как коммерческая тайна). Заметим, однако, что большинство ученых подходит к вопросу использования технологий искусственного интеллекта в образовании и языкознании с критических и взвешенных позиций [Агальцова, Валькова, 2024; Комарова, 2024; Королева, 2024].

На наш взгляд, подобный подход не является вполне продуктивным, так как принципы работы программного обеспечения регулируются и мотивируются разработчиками и их заказчиками в рамках целей, которые нам также

до конца непонятны. Таким образом, усилия ученых оказываются направленными не на чистое объективное явление, а на некую искусственную надстройку, которая, говоря формальным языком, зависит от множества переменных параметров, значения которых не могут быть известны.

В настоящем исследовании мы также используем возможности закрытой базы данных обработки естественного языка, однако мы задействуем ее только на начальном этапе работы, вписывая затем полученные результаты в полностью прозрачный и понятный авторский алгоритм. С одной стороны, мы используем технологию искусственного интеллекта на больших потоковых данных, с другой – мы четко «дозуем» ее применение и полностью управляем результатами ее действий. И объектом исследования выступает естественный язык, в то время как программное обеспечение является только лишь инструментом исследования, хотя как таковой и играет важную роль.

Итак, мы поставили перед собой цель экстраполировать возможности программного комплекса «Генератор сбалансированного лингвистического корпуса и корпусный менеджер» (Свидетельство о государственной регистрации программы для ЭВМ № 2023683209, Роспатент), созданного в лаборатории фундаментальных и прикладных проблем виртуального образования Московского государственного лингвистического университета, на предметную область обучения иностранным языкам в части генерации электронных учебных материалов. Ранее этот комплекс был апробирован на материале индоевропейских языков при решении задач интерпретации художественного и публицистического текста [Gorozhanov et al., 2024; Бондарчук, 2024; Степанова, 2023].

До работы над проблемой автоматического составления электронных учебных материалов цепочка шагов от «сырого» текста до корпуса и извлечения из него значимых лингвистических данных выглядела следующим образом:

1. Приведение исходного электронного текста к «нормальному»¹ виду (удаление метаданных, скоплений пробелов и переносов строк).
2. Превращение «нормализованного» текста в лингвистический корпус – базу данных SQLite – с помощью программы-генератора.
3. Формирование запросов к полученной базе данных посредством программы – корпусного менеджера.

Здесь применение технологии искусственного интеллекта ограничено рамками шага № 2, на котором мощности предварительно «натренированных» библиотек обработки естественного языка обеспечивают частеречную разметку текстового массива [Gorozhanov et al., 2024, p. 200]. Далее, на шаге № 3, в работу вступают наши собственные программные алгоритмы, которые уже не задействуют никаких технологий искусственного интеллекта (по крайней мере, на данном этапе исследования).

¹ «Нормализация» может пониматься в предметно-специальной литературе по-разному. В нашем случае мы понимаем под ней приведение электронного текста к такому виду, который позволяет трансформировать его в лингвистический корпус посредством программы-генератора.

Таким образом, чередование условно «простых» программных алгоритмов и алгоритмов с использованием технологий искусственного интеллекта и создает ситуацию, которую мы здесь обозначаем как *управляемый искусственный интеллект*.

Обратимся далее непосредственно к предмету нашего рассуждения, т. е. к тому, как с помощью указанной технологии обеспечить наполняемость онлайн-курса иностранного языка учебными материалами.

Минимальный онлайн-курс, размещенный в системе управления обучением (англ. LMS – learning management system), по нашему мнению состоит из теоретического материала и некоторой формы контроля, например, тестовых заданий с автоматической проверкой. Последние являются максимально формализованными объектами как с точки зрения лингводидактики, так и программирования, поскольку представляют собой типовую программную структуру. Сфокусируем наше внимание на них. В качестве системы управления обучением для проведения нашего эксперимента выберем LMS Moodle, которая имеет встроенный специальный формат хранения объектов (учебных модулей и их частей, онлайн-курсов и пр.) на основе разметки XML.

Мы добавим в корпусный менеджер модуль, который позволит генерировать на материале предварительно загруженного корпуса текстовые задания (файлы XML) для LMS Moodle. Полученный файл должен иметь возможность импорта в онлайн-курс посредством подсистемы резервного копирования.

Среди многообразия тестовых заданий² выберем в рамках настоящего исследования упражнения на заполнение пропусков готовыми вариантами ответов. Выберем также для демонстрации немецкий язык, а в качестве языкового явления глаголы. Тестовое задание будет представлять собой отрезок связного текста (например, 10 предложений), в котором будут пропущены глаголы. При наведении курсора на пропуск обучающийся получает список из всех пропущенных глаголов из этого отрывка, причем дублирование исключаются.

На стадии программной реализации модуля были предприняты следующие шаги:

1. Проанализирована структура XML-документа тестового вопроса LMS Moodle; при этом выделены те структурные элементы, которые должны подлежать заполнению в ходе генерации: формулировка задания, текст задания, правильные ответы и др.

2. Построен шаблон с одинаковыми для всех тестовых заданий на заполнение пропусков элементами; варьирующиеся элементы обозначены переменными для последующего автоматического замещения.

3. Создана функция генерации тестового задания в коде корпусного менеджера, которая:

- 3а) выбирает из всего корпуса $n=10$ предложений (и далее $+10$, пока не закончатся предложения или не будет задано иное);

- 3б) в текущем отрывке фиксируются глаголы;

- 3в) в списке глаголов исключается дублирование словоформ;

² URL: https://docs.moodle.org/404/en/Question_types.

3г) происходит перебор всех токенов отрывка, при котором глагольные формы заменяются на код пропусков, соотнесенных с ключами;

3д) формируется список ключей;

3е) происходит запись XML-файла, к наименованию которого добавляется временная метка с целью исключения «затирания» предыдущих файлов.

4. Внедрено оповещение о безошибочной генерации файла.

Код функции занял ок. 45 строк на языке программирования Python.

Далее был проведен эксперимент по загрузке полученного файла в банк вопросов онлайн-курса LMS Moodle, который показал, что полученный файл обладает параметрами интегративности и соответствия формату Moodle-XML.

В экспериментальном отрезке дважды встретились словоформы ‘entscheiden’ и ‘erfüllt’, однако в ключах они не были продублированы, что отвечает поставленному техническому заданию. Всего программа идентифицировала 18 глагольных форм, причем формы от одного и того же глагола были обозначены как различные (‘erhoben’ и ‘erhob’). Некоторые глаголы программа не идентифицировала, например, ‘werden’. Это связано с тем, что в разметке корпуса есть отдельная «часть речи» – «вспомогательный глагол», что не совсем отвечает общепринятому лингвистическому канону, но, тем не менее, вполне объяснимо в русле традиции алгоритмов обработки естественного языка³.

Визуализация полученного результата показана ниже (см. Рисунок 1):

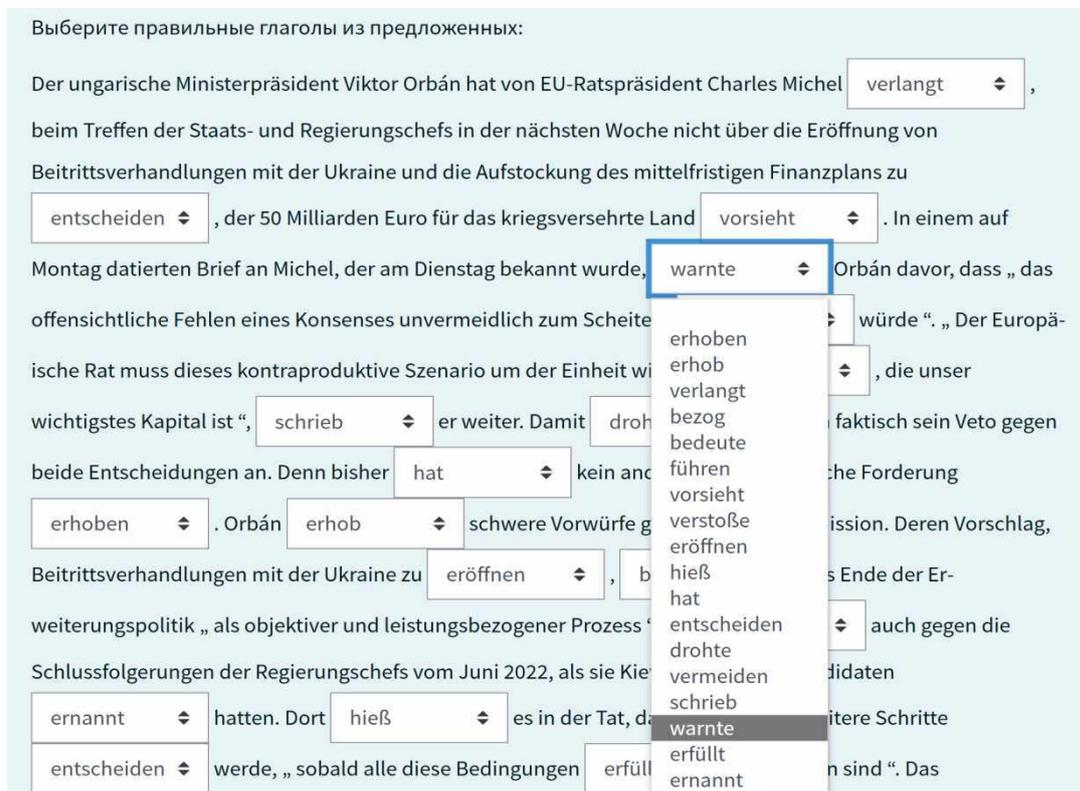


Рис. 1. Сгенерированное тестовое задание в режиме предпросмотра правильных ответов

³ См., например, URL: <https://spacy.io/usage/linguistic-features#pos-tagging>.

Мы можем заключить, что наше практико-ориентированное исследование завершилось успешно. Нам удалось модифицировать корпусный менеджер, добавив в него специальную функцию генерации тестовых заданий на выбор правильной лексики-глагола в когерентном тексте. Программа автоматически выбирает отрезок текста из корпуса, формирует файл специализированного формата Moodle-XML, который без ошибок имплементируется в онлайн-курсы LMS Moodle.

У настоящего исследования видятся значительные перспективы. Прежде всего, постулируя примат методики обучения иностранным языкам над какими бы ни было технологиями, необходимо разработать лингводидактическую составляющую, согласно которой будут «управляться» технологии искусственного интеллекта. Отечественная школа методики обучения иностранным языкам является устойчивой системой, которая стойко сохраняет баланс традиционного и нового, прогрессивного в свете достижений мировой науки и формирует концепцию положительного знания, в которую органично вписывается и технологическая компонента.

В качестве следующих шагов планируется разработать комплексы упражнений (тестов-тренажеров) на различные грамматические и лексические явления и продумать работу над лексическим минимумом к текстовому материалу.

ЛИТЕРАТУРА

1. Агальцова Д. В. Вызов искусственного интеллекта традиционной системе образования // Мир науки, культуры, образования. 2024. № 2 (105). С. 169–172. DOI 10.24412/1991-5497-2024-2105-169-172. – EDN QLVGNB.

2. Бондарчук Г. Г. Семиотические функции английских наименований одежды в публицистическом тексте (корпусное исследование) // Вестник Московского государственного лингвистического университета. Сер. Гуманитарные науки. 2024. № 4 (885). С. 23–29. – EDN BXILCR.

3. Гринин Л. Е. Искусственный интеллект: развитие и тревоги. Взгляд в будущее статья вторая. Искусственный интеллект: terra incognita или управляемая сила? // Философия и общество. 2023. № 4 (109). С. 5–32. DOI: 10.30884/jfio/2023.04.01. – EDN RVZOFX.

4. Комарова Е. В. Вовлечение чат-ботов в процесс обучения иностранным языкам // Филология и культура. 2024. № 1 (75). С. 149–158. DOI: 10.26907/2782-4756-2024-75-1-149-158. – EDN NQVTZM.

5. Королева А. В. Технология ознакомления студентов с возможностями ChatGPT и Midjourney как образовательных инструментов // Психолого-педагогический журнал Гаудеамус. 2024. Т. 23, № 2. С. 35–41. DOI: 10.20310/1810-231X-2024-23-2-35-41. – EDN CFXIKK.

6. Степанова Д. В. Программный комплекс для генерации динамического корпуса текстов СМИ // Вестник Минского государственного лингвистического университета. Сер. 1: Филология. 2023. № 6 (127). С. 123–130. – EDN FMBTKO.

7. Gorozhanov A. I. Natural Language Processing and Fiction Text: Basis for Corpus Research // RUDN Journal of Language Studies, Semiotics and Semantics. 2024. Vol. 15, № 1. P. 195–210. DOI: 10.22363/2313-2299-2024-15-1-195-210. – EDN FKVAOI.