

ПЕДАГОГИКА

Научная статья
УДК 372.881.161.1
doi: 10.17223/15617793/512/18

Генерация учебных текстов по русскому языку как иностранному с помощью российских чат-ботов: опыт автоматизированной сравнительной оценки

Николай Вячеславович Головкин¹

¹ Северо-Кавказский федеральный университет, Ставрополь, Россия, ngolovko@ncfu.ru

Аннотация. Представлены результаты исследования учебных текстов по РКИ, сгенерированных российскими чат-ботами GigaChat и «Алиса» на основе 25 тематических групп слов, входящих в лексический минимум элементарного уровня (A1). Путём автоматизированного анализа с помощью программного инструмента «Текстомер» установлено, что чат-боты в таких условиях в среднем генерируют тексты I сертификационного уровня владения РКИ. Сделан вывод о том, что применение чат-ботов может способствовать решению ряда проблем, связанных с отбором и адаптацией учебных текстов.

Ключевые слова: лингводидактика, русский язык как иностранный, генеративный искусственный интеллект, чат-бот, лексический минимум

Для цитирования: Головкин Н.В. Генерация учебных текстов по русскому языку как иностранному с помощью российских чат-ботов: опыт автоматизированной сравнительной оценки // Вестник Томского государственного университета. 2025. № 512. С. 170–177. doi: 10.17223/15617793/512/18

Original article
doi: 10.17223/15617793/512/18

Generation of educational texts in Russian as a foreign language using Russian chatbots: An attempt of automatized comparative evaluation

Nikolay V. Golovko¹

¹ North Caucasus Federal University, Stavropol, Russian Federation, ngolovko@ncfu.ru

Abstract. When didactic materials in foreign languages are created, text becomes a significant means of teaching. This leads to the need to select suitable texts, as well as subject them to qualitative and quantitative adaptation, i.e. transform them in order to make them suitable for solving educational tasks. Currently, methods and scenarios for the effective use of generative artificial intelligence are being actively sought, and the potential usefulness of chatbots based on it is being explored both for the methodology as a whole and for individual subject areas. Certain features of generative AI and chatbots based on large language models could mitigate a number of problems associated with the burden of selecting and adapting or independently creating educational texts in foreign languages. For example, chatbots based on generative AI are capable of generating texts based on sets of keywords, one of such possibly being a part of the lexical minimum in a foreign language. However, when a text is generated, arbitrary lexical and grammatical material can be added to the words of the lexical minimum, making it impossible to use it without adaptation at the target level of language proficiency. In order to establish the magnitude and significance of such deviations, an automated comparative study was carried out on a set of educational texts in Russian as a foreign language, generated through popular Russian public chatbots based on generative AI. For generation purposes, an elementary-level lexical minimum for Russian as a foreign language, divided into topical groups, is used. Based on these topical groups, small texts are created using GigaChat and Alisa (YaGPT) chatbots and then evaluated using a specialized automated analyzer called Textometer. Results were obtained, according to which 50 texts of 25 topical groups have an average size of 1070.4 characters with spaces, an average level of "Early B1" (1st certification level of proficiency in RFL), an average coverage of the elementary-level lexical minimum equal to 71% and an average level of lexical diversity equal to 0.58 on a scale between 0 and 1. The Textometer's assessment, according to which the generated texts would be useful in studying 13 grammatical topics (9 of elementary level, 4 of 1st certification level of proficiency in RFL) was also taken into account. Notably, the average performance of the GigaChat chatbot is biased towards generating more complex texts, and that of the Alisa chatbot is inclined towards simpler texts. It is concluded that popular Russian chatbots intended for general use can be applied without further modifications for the purpose of generating educational texts on RFL of a certain level of complexity; this neutralizes or mitigates a number of problems related to selection and adaptation of educational texts, but at the same time does not solve these problems completely.

Keywords: linguistic didactics, Russian as a foreign language, generative AI, chatbot, lexical minimum

For citation: Golovko, N.V. (2025) Generation of educational texts in Russian as a foreign language using Russian chatbots: An attempt of automatized comparative evaluation. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 512. pp. 170–177. (In Russian). doi: 10.17223/15617793/512/18

Введение

При создании дидактических материалов и пособий для нужд преподавания иностранных языков, в том числе русского языка как иностранного (РКИ), требуется решать ряд практических задач, в число которых входят подбор и обработка текстов. Учебный текст по иностранному языку может рассматриваться как одно из основных средств обучения [1]. В связи с этим к нему предъявляется ряд требований: он должен иметь надлежащий объём, быть посильным для обучающегося, но в то же время репрезентативным с лингводидактической точки зрения, т.е. содержать лексические и грамматические единицы, которые учащемуся требуется освоить. Как следствие, возникает потребность либо в том, чтобы создать подходящий текст самостоятельно, либо в том, чтобы подобрать уже существующий текст и выполнить процедуру его адаптации.

По мнению О.Ю. Редькиной, при обучении русскому языку как иностранному отбор учебных текстов должен производиться с учётом следующих требований: образцовость (быть образцом речи), стимулирующий характер (побуждать студентов к собственным высказываниям), преемственность и перспективность (учитывать пройденный материал и предоставлять доступ к новому), лингвокультурологическая нагруженность (нести информацию о культуре страны изучаемого языка), коммуникативность (служить фундаментом для способности студента участвовать в различных формах коммуникативного взаимодействия) [2].

Лингводидактическая адаптация текста, по мнению А.В. Петрухиной [3], это упрощение, приспособление, облегчение или усложнение текста в соответствии с уровнем языковой компетенции учащихся [3]. Согласно диссертационному исследованию А.В. Коротышева, адаптация преследует цель трансформировать текст в соответствии с новыми – учебными – условиями его функционирования, обеспечить его соответствие таким требованиям, как информативность, смысловая завершенность, делимость, связность планов содержания и выражения, фиксированная доля неизвестной лексики и объём [4]. Адаптация может быть количественной или качественной; в первом случае из текста изымаются менее существенные элементы с целью акцентировать его суть, во втором случае лексика и грамматика подвергаются изменению, чтобы текст достиг уровня языковой подготовки учащегося [5].

Значимую роль при адаптации текстов играют лексические минимумы – списки лексических единиц, подлежащих обязательному усвоению на соответствующих уровнях владения русским языком, согласно Н.П. Андриюшиной и соавт. [6]. Лексический минимум играет роль эталона, с которым сверяются учебные и тестовые материалы.

Н.П. Андриюшина и соавт. упоминают в цитируемом выше фрагменте об уровнях владения русским языком. В российской системе тестирования выделя-

ются такие уровни владения, как элементарный, базовый и I–IV сертификационные; они соответствуют европейским уровням системы ALTE A1, A2, B1, B2, C1 и C2 [7].

Выполнение упомянутых ранее процедур и обеспечение соответствия текстов приведённым критериям требуют, как представляется, существенных затрат времени и других ресурсов. Кроме того, при использовании текстов, которые охраняются авторским правом, составителям дидактических материалов или пособий необходимо учитывать ряд юридических ограничений. Эти проблемы могли бы быть смягчены, например, посредством использования возможностей генеративного искусственного интеллекта (ИИ). Данная тема ранее обсуждалась, в частности, на II Международной научно-методической конференции «Актуальные проблемы преподавания русского языка как неродного / иностранного: российские и зарубежные практики» (Северо-Кавказский федеральный университет, 18–22 сентября 2023 г.) [8].

Применение генеративного ИИ и диалоговых приложений на его основе (чат-ботов) в последнее время обсуждается как потенциально полезное для методики обучения в целом и для отдельных предметных областей в частности [9]. В лингводидактике чат-боты часто рассматриваются как эффективные, хотя и отчасти ограниченные, партнёры по диалогам, с которыми можно в комфортной обстановке отрабатывать языковые навыки [10, 11]. В то же время, на наш взгляд, ещё не в полной мере оценены возможности генеративного ИИ по созданию учебных текстов, в том числе на основе лексических минимумов определённого уровня. Представляется вполне вероятным, что чат-бот на основе генеративного ИИ мог бы порождать учебные тексты, которые сразу соответствовали бы определённым критериям, состояли бы из лексики строго заданного уровня сложности и не создавали бы юридических рисков. Это позволило бы составителям дидактических материалов и пособий уменьшать трудозатраты, минимизируя объём работы по отбору и адаптации текстов. При этом лингводидактика, по нашему мнению, в определённой мере «защищена» от возможной фактологической некомпетентности генеративного ИИ, которая могла бы иметь значение в другой предметной области: будучи в первую очередь материалом обучения языку и речи, лингводидактический учебный текст может в общем случае иметь произвольное смысловое наполнение при условии его соответствия языковой норме, т.е. его план выражения более значим, чем план содержания.

Руководствуясь вышеизложенным, мы поставили следующую цель: использовать один из существующих стандартных лексических минимумов по русскому языку как иностранному для генерации набора учебных текстов с помощью популярных российских чат-ботов общего назначения на основе генеративного

ИИ, а затем выполнить автоматизированный сравнительный анализ полученных текстов и определить степень их пригодности для использования в учебном процессе. Под «пригодностью для использования» в контексте настоящей статьи мы понимаем соответствие сгенерированных текстов целевому уровню владения русским языком как иностранным; целевой уровень определяется лексическим минимумом, на основе которого выполнялась генерация текстов. Задача апробации полученных текстов в ходе учебной работы с иностранными студентами в рамках данного отдельно взятого исследования не ставилась и не решалась.

Исследование, таким образом, позволит понять, насколько успешно составители дидактических материалов и пособий по русскому языку как иностранному могли бы в настоящее время использовать российские чат-боты общего назначения для генерации учебных текстов, соответствующих определённому уровню владения РКИ. Очевидно, что при генерации текста чат-бот будет использовать не только предложенный ему набор слов из лексического минимума, но и другие лексические единицы, а также грамматические конструкции различного характера; эти включения могут существенно увеличить степень сложности текста и сделать его малоприспособленным для использования на определённом уровне владения языком без той же самой предварительной адаптации, которой пришлось бы подвергнуть обычный текст, созданный человеком. Автоматизированная оценка даст возможность судить о том, насколько в среднем велики подобные отклонения.

Материалы и методы

В качестве опорного лексического минимума используется «Лексический минимум по русскому языку как иностранному. Элементарный уровень. Общее владение» (сост.: Н.П. Андрушина, Т.В. Козлова) [7]. Согласно предисловию к электронному изданию, его объём составляет 780 единиц.

Для нужд исследования применяется раздел «Тематические группы слов» части II. Единицы лексического минимума, сгруппированные по темам, последовательно включаются в состав стимулирующих запросов («промптов») для чат-ботов. Например, для тематической группы слов «1.6. Одежда, обувь» в чат-бот вводится запрос следующего вида:

Напиши небольшой текст, в котором будут слова: одежда, костюм, платье, рубашка, шапка, шарф, обувь, покупать, дарить, современный, красивый, лёгкий.

Запрос намеренно примитивизирован с целью упростить его обработку чат-ботом.

Если тематическая группа велика по размеру, то она делится на части по 20–30 слов. Каждая часть используется для генерации отдельного текста. Малые группы объединяются.

Используются те тематические группы слов, для которых возможна генерация связного текста. Характерным контрпримером может выступать группа «5.6.

Вопросительные слова». В редких случаях исключаются тематические группы или их части, для которых чат-бот отказывается сгенерировать текст из-за встроенных в него ограничений.

В качестве популярных российских чат-ботов общего назначения на основе генеративного ИИ используются GigaChat и «Алиса» (YaGPT). Диалоговое взаимодействие ведётся с веб-версиями чат-ботов, доступными по адресам <https://developers.sber.ru/gigachat/> и <https://a.ya.ru> соответственно. В каждый чат-бот вводится один и тот же стимулирующий запрос. Соответственно, в таблице результатов (см. далее) нечётные тексты сгенерированы GigaChat, чётные – YaGPT.

В охват исследования включаются только связные тексты. В ходе генерации были отмечены, например, случаи, когда чат-бот не мог установить смысловых связей между предложенными ему словами и заменял текст последовательным перечислением словарных определений каждого слова. В таких случаях вводился запрос на повторную генерацию. В отдельных ситуациях после нескольких неудачных попыток применялось изменение стимулирующего запроса: вместо «напиши небольшой текст» указывалось «напиши небольшой рассказ».

Специальные инженерные команды для управления чат-ботами (так называемый промпт-инжиниринг) в ходе генерации текстов не использовались.

Автоматизированная оценка полученных текстов выполняется с помощью онлайн-инструмента «Текстометр» [12]. Основным предметом оценки является уровень текста, определяемый «Текстометром» по количествам и соотношениям лексических единиц, входящих в те или иные лексические минимумы по русскому языку как иностранному. В качестве сопутствующих показателей нами учитываются размер текста в знаках с пробелами, степень охвата лексического минимума элементарного уровня (т.е. какова доля лексических единиц элементарного уровня в тексте), степень лексического разнообразия текста в размере от 0 до 1, а также мнение «Текстометра» о том, при изучении каких грамматических тем этот текст был бы наиболее полезен. Совокупность этих показателей позволяет, на наш взгляд, в достаточно полной мере оценить степень соответствия полученных текстов ожидаемому уровню владения РКИ. Иными словами, по этим показателям можно судить о том, насколько сильно текст отклоняется от целевого элементарного уровня (A1 по европейской системе ALTE).

Результаты

В охват исследования включены 50 сгенерированных чат-ботами текстов по 25 тематическим группам слов лексического минимума элементарного уровня по русскому языку как иностранному.

Размер полученных текстов варьируется от 711 до 1 638 знаков с пробелами, среднее значение – 1 070,4.

Уровень сгенерированных текстов по оценке «Текстометра» также различается во вполне широких пределах, от элементарного до II сертификационного.

Средняя величина составила 1,9, что соответствует началу I сертификационного уровня.

Наименьшее значение охвата целевого лексического минимума отмечено в тексте № 9 и составляет 48% (т.е. менее половины слов в тексте относятся к лексическому минимуму элементарного уровня, все остальные имеют более высокий уровень). Максимальное значение зафиксировано в тексте № 7 и составляет 93%. Средняя доля – 71%.

Степень лексического разнообразия текстов, выражаемая «Текстометром» в виде числа от 0 до 1, находится в пределах диапазона от 0,35 до 0,75. Среднее значение – 0,58.

Возможные грамматические темы, при изучении которых могли бы использоваться сгенерированные

тексты согласно оценке «Текстометра», – «Вид глагола», «Модальные глаголы», «Прилагательные», «Деепричастия», «Родительный падеж», «Порядковые числительные», «Творительный падеж», «Краткие формы прилагательных и причастий», «Причастия», «Предложный падеж», «Числительные», «Наречия» и «Пассивные формы». Последняя из перечисленных тем является наиболее редкой и встречается 1 раз. Наиболее частотная грамматическая тема – «Причастия» (14 упоминаний).

Все результаты представлены в таблице. Уровень текста в таблице указан в соответствии с оценочной шкалой инструмента «Текстометр», где А1 – элементарный уровень, А2 – базовый уровень и т.д.

Результаты автоматизированного исследования сгенерированных текстов

№	Раздел	Размер	Уровень	Охват А1, %	Разнообразие	Возможные грамматические темы
1	1.1. Фазы жизни, возраст	897	Начало В2	62	0,60	Вид глагола
2		825	Конец В1	59	0,64	Модальные глаголы, Прилагательные, Деепричастия
3	1.4. Питание, продукты	760	А1	76	0,35	Родительный падеж, Порядковые числительные, Творительный падеж
4		989	А1	69	0,55	Модальные глаголы, Порядковые числительные, Творительный падеж
5		725	А1	75	0,61	Краткие формы прилагательных и причастий, Деепричастия
6		808	Начало А2	72	0,70	Причастия, Краткие формы прилагательных и причастий, Родительный падеж, Деепричастия, Творительный падеж
7	1.5. Жильё, мебель	1 140	А1	93	0,42	Причастия, Прилагательные
8		1 246	Середина В1	67	0,68	Порядковые числительные
9	1.6. Одежда, обувь	1 063	Начало В2	48	0,58	Деепричастия
10		964	Конец В1	58	0,68	Деепричастия
11		1 108	Середина В1	70	0,52	Модальные глаголы
12	2.1. Интеллект, воля, желание	1 082	Середина В1	76	0,53	Предложный падеж, Модальные глаголы
13	2.2. Черты характера, эмоции, состояние	1 288	Середина В1	73	0,51	Нет
14		1 226	Начало В2	65	0,70	Прилагательные
15	2.3. Действия, направленные на кого-либо, на что-либо	1 425	Середина В1	77	0,45	Краткие формы прилагательных и причастий, Причастия, Прилагательные
16		1 198	Конец В1	73	0,56	Краткие формы прилагательных и причастий, Порядковые числительные
17	2.4. Оценка чего-либо, кого-либо, значимость	1 278	Середина В1	74	0,43	Нет
18		1 089	Середина А2	79	0,54	Краткие формы прилагательных и причастий
19	3.1. Государственное устройство; 3.2. Наименование человека, его характеристика	972	А1	84	0,41	Предложный падеж, Родительный падеж, Причастия, Прилагательные
20		929	Начало А2	85	0,41	Прилагательные, Творительный падеж
21	3.3. Семья, родственники	1 638	Середина А2	77	0,59	Причастия, Числительные
22		1 084	Начало А2	75	0,53	Краткие формы прилагательных и причастий, Причастия, Прилагательные
23	3.4. Профессия, специальность; 3.5. Обозначение человека по статусу и роду деятельности	960	Начало В2	64	0,67	Вид глагола, Прилагательные, Деепричастия, Творительный падеж
24		939	Начало А2	73	0,69	Краткие формы прилагательных и причастий
25	3.7. Сферы деятельности	1 314	Начало В2	55	0,70	Модальные глаголы, Числительные, Деепричастия
26		1 043	Середина А2	77	0,59	Порядковые числительные, Деепричастия, Творительный падеж
27	3.8. Образование, наука	1 092	Середина В1	62	0,65	Вид глагола, Причастия, Деепричастия
28		1 109	Конец А2	57	0,67	Причастия
29	3.9. Искусство	1 219	Конец В1	67	0,69	Прилагательные, Порядковые числительные
30		1 108	Середина В1	73	0,57	Прилагательные
31		765	Середина А2	63	0,75	Деепричастия, Творительный падеж
32		991	А1	83	0,56	Причастия, Прилагательные, Творительный падеж

№	Раздел	Размер	Уровень	Охват A1, %	Разнообразие	Возможные грамматические темы
33	3.10. Спорт	1063	Конец B1	57	0,69	Нет
34		1178	Середина B1	68	0,49	Модальные глаголы, Творительный падеж
35	3.11. Отдых и праздники	1348	Середина B1	62	0,60	Причастия, Порядковые числительные, Творительный падеж
36		736	A1	76	0,69	Краткие формы прилагательных и причастий, Причастия, Творительный падеж
37	3.12. Путешествия, туризм	995	Середина B1	67	0,65	Деепричастия
38		1196	Конец A2	76	0,53	Нет
39	3.13. Транспорт, поездки	1026	Середина A2	70	0,57	Предложный падеж
40		1017	Середина A2	66	0,61	Родительный падеж
41	3.14. Место проживания и работы (городские учреждения)	1069	Середина A2	71	0,58	Родительный падеж, Причастия, Порядковые числительные
42		711	Начало A2	83	0,61	Творительный падеж
43		1142	Середина B1	71	0,56	Предложный падеж, Наречия
44		1161	A1	85	0,57	Предложный падеж, Родительный падеж
45	3.15. Коммуникация	936	Начало B1	76	0,57	Вид глагола, Порядковые числительные
46		1221	Середина B1	75	0,50	Наречия
47	3.17. Документы, печатная продукция	849	Середина B1	75	0,54	Причастия, Пассивные формы, Краткие формы прилагательных и причастий, Прилагательные, Порядковые числительные
48		1202	Середина A2	80	0,53	Нет
49	4.2. Климат	1551	Середина B1	61	0,51	Творительный падеж
50		845	Середина A2	70	0,61	Причастия, Порядковые числительные

Сгенерированные тексты также в определённой мере могут рассматриваться как результаты исследования, поэтому приведём примеры. В частности, как видно из таблицы, по тематической группе слов «Питание, продукты» оба чат-бота сгенерировали тексты, которые оценены «Текстометром» как относящиеся к элементарному уровню (A1). В стимулирующем запросе были указаны слова *завтрак, обед, ужин, суп, хлеб, рыба, яйцо, рис, салат, овощи, картофель, соль, фрукты, яблоко, мороженое, шоколад, сахар, молоко, чай, кофе, сок, вода, пиво, вино, колбаса, сыр, масло*.

Текст № 3:

Завтрак был легким и состоял из чашки кофе и кусочка хлеба с маслом. Обед был более плотным: суп из овощей, картофель с рыбой и салат из свежих овощей. Ужин был еще более сытным: рис с курицей, салат из помидоров и огурцов, а также яблоко на десерт.

На следующий день завтрак был более разнообразным: яйцо всмятку, тосты с джемом и чашка чая. Обед был легким и состоял из салата из свежих овощей и фруктов, а также стакана сока. Ужин был более плотным: суп из рыбы, картофель с мясом и салат из свежих овощей.

На третий день завтрак был более плотным: каша из риса с молоком и чашка кофе. Обед был легким и состоял из салата из свежих овощей и фруктов, а также стакана сока. Ужин был более плотным: суп из овощей, картофель с рыбой и салат из свежих овощей.

Текст № 4:

Завтрак – это самый важный прием пищи за день. Он должен быть сытным и вкусным, чтобы зарядить энергией на весь день. Обычно на завтрак едят яйца, рис, салат, овощи, картофель, рыбу, курицу или мясо. Также не стоит забывать о хлебе, соли и масле, которые являются основой любого завтрака.

Обед – это второй по важности прием пищи за день. Обед обычно состоит из супа, который может быть мясным или рыбным, а также из второго блюда, например, картофеля с овощами или мяса с рисом. Также на обед едят хлеб, фрукты, яблоки, мороженое и шоколад.

Ужин – это последний прием пищи за день, который должен быть легким и не слишком калорийным. На ужин едят фрукты, овощи, курицу, рыбу или мясо, а также пьют чай, кофе, сок, молоко или воду.

В кафе и ресторанах люди обычно заказывают блюда из меню, которые могут включать в себя различные салаты, супы, вторые блюда, десерты и напитки. В столовых же люди могут купить готовые блюда и напитки, а также воспользоваться столовыми приборами и посудой.

Обсуждение

Как видно из таблицы результатов, размер текстов варьируется во вполне широких пределах, однако представляется приемлемым для применения на элементарном уровне владения русским языком как иностранным. Напомним, что среднее значение составляет 1070,4 знака с пробелами. Чат-бот GigaChat в среднем генерирует более объёмные тексты, чем «Алиса»: 1104,9 и 1035,9 знака с пробелами соответственно.

Полученные результаты автоматизированной оценки уровня сгенерированных текстов могут указывать на то, что чат-боты общего пользования на основе генеративного ИИ без дополнительных настроек и доработок порождают, как правило, тексты более высокого уровня по сравнению с целевым. Так, только в 8 случаях из 50 (16%) уровень текста был оценён «Текстометром» как элементарный (A1).

Чат-боты генерировали тексты этого уровня по тематическим группам слов «Питание, продукты», «Жильё, мебель», «Государственное устройство; Наименование человека, его характеристика», «Искусство», «Отдых и праздники», «Место проживания и работы (городские учреждения)». Ещё 7 текстов были оценены как относящиеся к базовому уровню (A2). Большинство текстов (21 из 50, 42%) инструмент «Текстометр» отнёс к I сертификационному уровню. Тексты уровня A2 можно, на наш взгляд, считать не требующими дополнительной адаптации, однако даже при этом допущении 70% сгенерированных текстов будут нуждаться в качественном преобразовании с целью обеспечить возможность их применения на элементарном уровне. В то же время на основании этих данных можно утверждать, что сгенерированные чат-ботами общего пользования на основе генеративного ИИ тексты могли бы без адаптации использоваться на I и II сертификационных уровнях владения русским языком как иностранным.

Следует также отметить, что чат-бот «Алиса» в среднем генерировал более простые тексты, чем GigaChat: для «Алисы» средний уровень текста по оценке «Текстометра» составляет 1,7, что эквивалентно значению «Конец A2», в то время как для GigaChat средний уровень равен 2,1, или «Середина B1». Эти показатели могут свидетельствовать о том, что чат-бот «Алиса» более эффективен для генерации учебных текстов элементарного и базового уровней, а GigaChat – для порождения текстов I и II сертификационных уровней.

Определённая инструментом «Текстометр» доля слов, относящихся к лексическому минимуму элементарного уровня, в сгенерированных текстах в целом вполне велика. Е.А. Мирошникова указывает, что максимальная приемлемая доля незнакомой лексики в иноязычном тексте составляет в общем случае 30% [5]; как следует из таблицы результатов, этому требованию соответствует 31 текст из 50 (62%), и ещё в 13 текстах доля слов, которые не входят в лексический минимум элементарного уровня, не превышает 40%. Эти результаты могут указывать на то, что в определённых сценариях, таких как просмотровое чтение, сгенерированные чат-ботами общего пользования на основе генеративного ИИ тексты могли бы использоваться без лексической адаптации или с незначительной лексической трансформацией даже на элементарном уровне владения русским языком как иностранным. Отметим при этом, что чат-бот «Алиса» в среднем использовал больше слов элементарного уровня, чем GigaChat: 72,8 и 69,2% соответственно. Это также может свидетельствовать о том, что «Алиса» в большей степени эффективна для генерации текстов элементарного и базового уровней.

Параметр «Лексическое разнообразие», оцениваемый «Текстометром» по шкале от 0 до 1, приближается к 0, если в тексте велико количество одинаковых слов. Как следствие, по этой величине можно косвенно судить о степени трудоёмкости текста при работе с ним: повторяющиеся лексические единицы способствуют более простому и быстрому извлечению информации. Большинство текстов, как следует из таблицы резуль-

татов, относятся к диапазону от 0,49 до 0,61, т.е. типичный сгенерированный текст располагается приблизительно посередине шкалы. Средние величины по этому показателю почти одинаковы: общий – 0,58, для GigaChat – 0,57, для «Алисы» – 0,59. Эти показатели могут свидетельствовать о том, что в большинстве случаев при генерации учебных текстов по русскому языку как иностранному чат-ботами порождается текстовый материал умеренной словарной трудоёмкости.

Грамматическая сложность материала, для оценки которой мы используем мнение инструмента «Текстометр» о том, при изучении каких грамматических тем тот или иной текст мог бы быть наиболее полезен, в целом видится нам невысокой. Девять тем из 13 (69%) относятся к числу тех, которые рассматриваются и изучаются уже на элементарном уровне. Следует, однако, отметить, что две грамматические темы I сертификационного уровня – «Причастия» и «Деепричастия» – входят в пятёрку наиболее часто встречающихся (14 и 11 случаев соответственно). В то же время частота встречаемости в целом показывает почти такое же распределение, как и тематика: 61 упоминание из 96 (63,5%) приходится на те грамматические темы, которые рассматриваются на элементарном уровне. При этом GigaChat чаще генерирует тексты, которые «Текстометр» оценивает как полезные для изучения грамматических тем I сертификационного уровня: 22 случая против 13 у чат-бота «Алиса». При оценке данных результатов следует, на наш взгляд, учитывать, что грамматические темы изучаются циклично и к ним возвращаются повторно на последующих этапах изучения языка; соответственно, нет оснований утверждать, что наличие грамматики, которая осваивается уже на элементарном уровне, непосредственно свидетельствует об отсутствии потребности в грамматической адаптации полученного текста. К примеру, падежи изучаются как на элементарном уровне владения РКИ, так и на базовом; в последнем случае к уже известной информации добавляются сведения о склонении имён прилагательных. Соответственно, текст, в котором широко представлены прилагательные в косвенных падежах, может потребовать грамматической адаптации, хотя косвенные падежи как грамматическая тема изучаются уже на элементарном уровне.

Заключение

На основании полученных результатов и их обсуждения может быть сделан ряд выводов.

Российские чат-боты общего пользования на основе генеративного ИИ могут без дополнительных доработок и модификаций применяться для генерации учебных текстов по русскому языку как иностранному с опорой на лексические минимумы, однако тексты, порождаемые ими на основе лексических минимумов элементарного уровня, в большинстве случаев (42%) оцениваются профильным автоматизированным анализатором «Текстометр» как соответствующие I сертификационному уровню владения РКИ, поэтому получаемый таким образом учебный материал потребует

качественной адаптации при необходимости применять его на элементарном или базовом уровне владения русским языком как иностранным.

Учебные тексты по русскому языку как иностранному, порождаемые российскими чат-ботами общего пользования на основе генеративного ИИ с опорой на лексические минимумы элементарного уровня, в большинстве случаев (62%) соответствуют типовому требованию, согласно которому учебный текст по иностранному языку должен содержать не более 30% незнакомой лексики (в данном случае – не входящей в лексический минимум элементарного уровня, согласно оценке анализатора «Текстомер»). Как следствие, в отдельных сценариях, таких как просмотровое чтение, сгенерированные таким образом тексты могут использоваться на любом уровне владения РКИ без лексической адаптации или при незначительной лексической адаптации. Уровень лексического разнообразия таких текстов – средний (около 0,58 по шкале от 0 до 1).

Учебные тексты по русскому языку как иностранному, порождаемые российскими чат-ботами общего пользования на основе генеративного ИИ с опорой на лексические минимумы элементарного уровня, в большинстве случаев (63,5%) полезны, согласно оценке профильного автоматизированного анализатора «Текстомер», для изучения грамматических тем, которые осваиваются уже на элементарном и базовом уровнях владения РКИ («Вид глагола», «Модальные глаголы», «Прилагательные», «Родительный падеж», «Порядковые числительные», «Творительный падеж», «Предложный падеж», «Числительные», «Наречия»). Тем не менее сгенерированные таким образом тексты могут требовать грамматической адаптации, так как одни и те же грамматические темы изучаются на разных уровнях владения РКИ.

Чат-бот «Алиса» (YaGPT) в среднем порождает в описанных условиях тексты уровня «Конец А2» (базовый уровень владения РКИ), где 72,8% слов входят в

лексический минимум элементарного уровня, а чат-бот GigaChat – тексты уровня «Середина В1» (I сертификационный уровень), где 69,2% слов входят в лексический минимум элементарного уровня. Кроме того, чат-бот GigaChat чаще генерирует тексты, которые анализатор «Текстомер» оценивает как полезные для изучения грамматических тем I сертификационного уровня. Это позволяет рекомендовать в настоящее время применять «Алису» в тех случаях, когда требуется сгенерировать учебные тексты по русскому языку как иностранному более низкого уровня, а GigaChat – в тех случаях, когда необходим текст более высокого уровня сложности. Следует, однако, учитывать, что чат-боты общего пользования на основе генеративного ИИ являются самообучающимися системами и могут выдавать различные результаты с течением времени.

Полученные результаты дают, на наш взгляд, основания утверждать, что применение российских чат-ботов общего пользования на основе генеративного ИИ при создании дидактических материалов и пособий по русскому языку как иностранному с опорой на лексические минимумы позволяет быстро создавать тексты определённого уровня сложности, содержащие заранее заданный объём целевой лексики, упрощать и ускорять процедуру отбора текстов, снимать или существенно уменьшать потребность в лексической адаптации текста, смягчать юридические риски, связанные с использованием текстов, которые охраняются авторским правом. В то же время следует учитывать широкий разброс общих итоговых уровней сложности генерируемых текстов, который не позволяет в настоящее время говорить о полном и всестороннем решении проблем, связанных с необходимостью адаптации текстов для их применения в учебном процессе. Перспективы дальнейшей работы по этому направлению могли бы включать в себя, на наш взгляд, создание специализированных больших языковых моделей и приложений на их основе, изначально нацеленных на решение лингводидактических задач.

Список источников

1. Одинцова Е.А. Текст и текстовые категории в преподавании РКИ. Проект учебного пособия для филологов-иностранцев : магистерская диссертация. Екатеринбург : УрФУ, 2021. 124 с.
2. Редькина О.Ю. Требования к учебному тексту в сфере обучения русскому языку как иностранному // Вестник Челябинского государственного университета. 2011. № 33 (248). Серия «Филология. Искусствоведение». Вып. 60. С. 258–260.
3. Петрухина А.В. Лингводидактическая адаптация учебных медицинских текстов для иностранных студентов // Научный диалог. 2022. Т. 11, № 9. С. 124–141.
4. Коротышев А.В. Технология отбора и лингводидактической адаптации художественных текстов в целях обучения русскому языку как иностранному : автореф. дис. ... канд. пед. наук. М. : МГУ, 2017.
5. Мирошникова Е.А. Адаптация текстового учебного материала при дифференцированном обучении иностранному языку // Вестник Брянского государственного университета. 2016. № 3 (29). С. 229–234.
6. Андрияшина Н.П., Афанасьева И.Н., Дунаева Л.А. [и др.] Структура лексического минимума как компонента российской государственной системы тестирования // Язык, сознание, коммуникация : сб. ст. / ред. кол. М.Л. Ремнёва, Е.Л. Бархударова, А.И. Изотов, В.В. Красных, Ф.И. Панков. М. : МАКС Пресс, 2013. Вып. 47. 590 с.
7. Лексический минимум по русскому языку как иностранному. Элементарный уровень. Общее владение / Н.П. Андрияшина, Т.В. Козлова. 5-е изд. СПб. : Златоуст, 2015. 80 с.
8. Головкин Н.В. Рассуждение о возможном способе применения генеративного искусственного интеллекта при создании пособий по русскому языку как иностранному // Актуальные проблемы преподавания русского языка как неродного / иностранного: российские и зарубежные практики : материалы II Междунар. науч.-практ. конф., Ставрополь, 18–22 сентября 2023 г. Ставрополь : Северо-Кавказский федеральный университет, 2023. С. 126–129.
9. Константинова Л.В., Ворожихин В.В., Петров А.М. [и др.] Генеративный искусственный интеллект в образовании: дискуссии и прогнозы // Открытое образование. 2023. Т. 27, № 2. С. 36–48.
10. Dauletbayeva A.E. Chatbots as an auxiliary element in learning foreign languages: advantages and disadvantages // Педагогический опыт: идеи и инновации : сб. ст. Междунар. профессионально-исследовательского конкурса, Пенза, 25 января 2023 года. Пенза : Наука и Просвещение (ИП Гуляев Г.Ю.), 2023. С. 11–14.
11. Сысоев П.В., Филатов Е.М. Чат-боты в обучении иностранному языку: преимущества и спорные вопросы // Вестник Тамбовского университета. Серия: Гуманитарные науки. 2023. Т. 28, № 1. С. 66–72.

12. Лапошина А.Н., Лебедева М.Ю. Текстометр: онлайн-инструмент определения уровня сложности текста по русскому языку как иностранному // Русистика. 2021. Т. 19, № 3. С. 331–345.

References

1. Odintsova, E.A. (2021) *Tekst i tekstovye kategorii v prepodavanii RFL. Proekt uchebnogo posobiya dlya filologov-inostrantsev* [Text and Text Categories in Teaching RFL. Draft Textbook for Foreign Philologists]. Master's Thesis. Yekaterinburg: UrFU.
2. Red'kina, O.Yu. (2011) *Trebovaniya k uchebnomu tekstu v sfere obucheniya russkomu yazyku kak inostrannomu* [Requirements for Educational Text in the Field of Teaching Russian as a Foreign Language]. *Vestnik Chelyabinskogo gosudarstvennogo universiteta*. 33 (248). pp. 258–260.
3. Petrukhina, A.V. (2022) *Lingvodidakticheskaya adaptatsiya uchebnykh meditsinskikh tekstov dlya inostrannykh studentov* [Linguodidactic Adaptation of Educational Medical Texts for Foreign Students]. *Nauchnyy dialog*. 11 (9). pp. 124–141.
4. Korotyshev, A.V. (2017) *Tekhnologiya otbora i lingvodidakticheskoy adaptatsii khudozhestvennykh tekstov v tselyakh obucheniya russkomu yazyku kak inostrannomu* [Technology of Selection and Linguodidactic Adaptation of Literary Texts for the Purpose of Teaching Russian as a Foreign Language]. Abstract of Pedagogics Cand. Diss. Moscow: MSU.
5. Miroshnikova, E.A. (2016) *Adaptatsiya tekstovogo uchebnogo materiala pri differentsirovannom obuchenii inostrannomu yazyku* [Adaptation of Textual Educational Material in Differentiated Foreign Language Teaching]. *Vestnik Bryanskogo gosudarstvennogo universiteta*. 3 (29). pp. 229–234.
6. Andryushina, N.P., Afanas'eva, I.N. & Dunaeva, L.A. et al. (2013) *Struktura leksicheskogo minimuma kak komponenta rossiyskoy gosudarstvennoy sistemy testirovaniya* [The Structure of the Lexical Minimum as a Component of the Russian State Testing System]. In: Remnyova, M.L. et al. (eds) *Yazyk, soznanie, kommunikatsiya* [Language, Consciousness, Communication]. Collection of Articles. Issue 47. Moscow: MAKSS Press.
7. Andryushina, N.P. & Kozlova, T.V. (2015) *Leksicheskiy minimum po russkomu yazyku kak inostrannomu. Elementarnyy uroven'. Obshchee vladenie* [Lexical Minimum for Russian as a Foreign Language. Elementary Level. General Proficiency]. 5th Ed. Saint Petersburg: Zlatoust.
8. Golovko, N.V. (2023) [Reasoning on a Possible Way of Applying Generative Artificial Intelligence in Creating Textbooks for Russian as a Foreign Language]. *Aktual'nye problemy prepodavaniya russkogo yazyka kak nerodnogo / inostrannogo: rossiyskie i zarubezhnye praktiki* [Current Problems of Teaching Russian as a Non-Native/Foreign Language: Russian and Foreign Practices]. Proceedings of the 2nd International Conference. Stavropol. 18–22 September 2023. Stavropol: NCFU. pp. 126–129. (In Russian).
9. Konstantinova, L.V. et al. (2023) *Generativnyy iskusstvennyy intellekt v obrazovanii: diskussii i prognozy* [Generative Artificial Intelligence in Education: Discussions and Forecasts]. *Otkrytoe obrazovanie*. 27 (2). pp. 36–48.
10. Dauletbayeva, A.E. (2023) Chatbots as an auxiliary element in learning foreign languages: advantages and disadvantages. In: *Pedagogicheskiy opyt: idei i innovatsii* [Pedagogical Experience: Ideas and Innovations]. Collection of Articles of the International Professional Research Competition. Penza. 25 January 2023. Penza: Nauka i Prosveshchenie (IP Gulyaev G.Yu.). pp. 11–14.
11. Sysoev, P.V. & Filatov, E.M. (2023) *Chat-boty v obuchenii inostrannomu yazyku: preimushchestva i spornye voprosy* [Chatbots in Foreign Language Teaching: Advantages and Controversial Issues]. *Vestnik Tambovskogo universiteta. Seriya: Gumanitarnye nauki*. 28 (1). pp. 66–72.
12. Laposhina, A.N. & Lebedeva, M.Yu. (2021) *Tekstometr: onlayn-instrument opredeleniya urovnya slozhnosti teksta po russkomu yazyku kak inostrannomu* [Textometer: An Online Tool for Determining the Complexity Level of a Text in Russian as a Foreign Language]. *Rusistika*. 19 (3). pp. 331–345.

Информация об авторе:

Головко Н.В. – канд. филол. наук, доцент кафедры русского языка как иностранного Северо-Кавказского федерального университета (Ставрополь, Россия). E-mail: ngolovko@ncfu.ru

Автор заявляет об отсутствии конфликта интересов.

Information about the author:

N.V. Golovko, Cand. Sci. (Philology), associate professor, North Caucasus Federal University (Stavropol, Russian Federation). E-mail: ngolovko@ncfu.ru

The author declares no conflicts of interests.

Статья поступила в редакцию 02.12.2024;
одобрена после рецензирования 20.03.2025; принята к публикации 31.03.2025.

The article was submitted 02.12.2024;
approved after reviewing 20.03.2025; accepted for publication 31.03.2025.