

А. В. Белый, О. А. Митрофанова,
Н. А. Дубинина

DOI: 10.21638/spbu30.2023.212



**Андрей Владимирович
Белый**

студент

► a.v.belij@yandex.ru

**Ольга Александровна
Митрофанова**

кандидат филологических наук, доцент

► o.mitrofanova@spbu.ru

**Надежда Александровна
Дубинина**

специалист-тестор

► dnadine@yandex.ru

Санкт-Петербургский
государственный университет,
Российская Федерация, 199034, Санкт-
Петербург, Университетская наб., 7–9

**Andrei V. Belyi
Olga A. Mitrofanova
Nadezhda A. Dubinina**

St. Petersburg State University,
7–9, Universitetskaya nab., St. Petersburg,
199034, Russian Federation

АВТОМАТИЧЕСКАЯ ГЕНЕРАЦИЯ ЛЕКСИКО-ГРАММАТИЧЕСКИХ ЗАДАНИЙ ПО РУССКОМУ ЯЗЫКУ КАК ИНОСТРАННОМУ С ПОМОЩЬЮ ПРЕДСКАЗЫВАЮЩИХ ЯЗЫКОВЫХ МОДЕЛЕЙ

ANDREI V. BELYI, OLGA A. MITROFANOVA, NADEZHDA A. DUBININA
AUTOMATIC GENERATION OF LEXICO-GRAMMATICAL TESTS FOR RUSSIAN AS
A FOREIGN LANGUAGE USING PREDICTIVE LANGUAGE MODELS

При обучении иностранному языку одной из базовых потребностей участников образовательного процесса является достаточное количество учебного материала. Среди заданий, способствующих усвоению лексико-грамматических конструкций, особую популярность получили задания на заполнение с множественным выбором (multiple-choice gap-filling exercises). В настоящее время создание вручную большого количества уникальных заданий оказывается трудозатратным. Разработка алгоритмов автоматической генерации упражнений для русского языка, несмотря на существующую потребность, ведется не так активно, как, например, для английского языка. В связи с этим в статье предлагается метод автоматической генерации упомянутого типа заданий по русскому языку как иностранному (РКИ), который основывается на использовании дистрибутивно-семантических моделей типа word2vec и позволяет создавать задания на основе аутентичных текстов. Представленный метод применяется вне зависимости от жанрово-стилевой принадлежности текста, соотносимого с ним языкового уровня, может быть легко адаптирован для других языков. Для обучения модели word2vec был собран корпус детской и учебной литературы, моделирующий языковой опыт обучаемых РКИ. В ходе исследования разработано и протестировано веб-приложение для преподавателей. Для оценки релевантности получаемых заданий были проведены два эксперимента: по тестированию наивных носителей русского языка и по опросу экспертов — составителей тестов по РКИ. Результаты экспериментов подтверждают эффективность предлагаемого метода, демонстрируют высокую степень корректности заданий и подбираемых дистрибутивов: по результатам тестирования точность (precision) составляет 0,8, а полнота (recall) 0,91. Эксперты также отмечают удобство веб-приложения.

Ключевые слова: преподавание русского языка как иностранного, автоматическое создание языковых упражнений, лексико-грамматические упражнения, gap-filling, multiple-choice.

* Исследование проводится в рамках НИП СПбГУ № 75254082 «Моделирование коммуникативного поведения жителей российского мегаполиса в социально-речевом и прагматическом аспектах с привлечением методов искусственного интеллекта» и при поддержке гранта РФФИ № 21-78-10148 «Моделирование значения слова в индивидуальном языковом сознании на основе дистрибутивной семантики».

The research is carried out within the framework of the St. Petersburg State University research project no. 75254082 “Modeling the communicative behavior of residents of a Russian metropolis in socio-speech and pragmatic aspects using artificial intelligence methods” and the Russian Science Foundation grant no. 21-78-10148 “Modeling the meaning of a word in individual linguistic consciousness based on distributive semantics”.

Благодарности

Авторы выражают благодарность исследователям и экспертам в области ТРКИ — И. О. Гурьянову, О. Н. Камшиловой, А. В. Колмогоровой, М. В. Копотеву, Е. В. Курганской, А. Н. Лапошиной, Е. В. Маркасовой, Н. М. Марусенко, М. А. Пономаревой, Н. В. Семеновой, Ш. М. Хамроевой, В. А. Шульгинову, К. А. Щукиной, а также принявшим участие в тестировании учебных материалов студентам-лингвистам.

The authors argue that in teaching a foreign language, one of the basic needs of participants of the educational process is a sufficient number of educational data. Among the tasks that contribute to the acquisition of lexical and grammatical units, gap-filling exercises with multiple-choice have become particularly popular. Nowadays, creating unique tasks manually turns out to be labor-intensive. Unlike English, the development of algorithms for exercise generation for Russian is not so active, despite the existing need. In this regard, the authors propose a method for automatic generation of tasks of this type for Russian as a foreign language (RFL). The proposed method is based on distributive semantic models like word2vec and allows to create tasks based on authentic texts, it does not depend on the genre and style of the text or the corresponding language level, and can be easily adapted for other languages. To train the word2vec model, a corpus of children's and educational literature was developed to emulate the language experience of students. In the course of the work, a web application for teachers was also launched. To assess the consistency and relevance of generated tasks, two experiments were conducted. In the first experiment, naive native speakers of the Russian language were interviewed, while in the second, a survey of experts in RFL was carried out. The high degree of correctness of the tasks and the selected distractors is proved by high scores of precision (0.8) and recall (0.91). The experts have also noted the convenience of the web application.

Keywords: teaching Russian as a foreign language, automatic generation of language exercises, lexical and grammatical exercises, gap-filling, multiple-choice.

Acknowledgements

The authors express their deep gratitude to researches and TORFL experts I. O. Guryanov, O. N. Kamshilova, Sh. M. Khamroeva, A. V. Kolmogorova, M. V. Kopotev, E. V. Kurganskaya, A. N. Laposhina, E. V. Markasova, N. M. Marusenko, M. A. Ponomareva, N. V. Semenova, K. A. Shchukina, V. A. Shulginov, and students, who participated in the testing of didactic materials.

Введение

Языковые тесты, направленные на оценку компетенций учащихся, играют огромную роль в процессе освоения иностранного языка.

Тесты являются эффективным средством контроля знаний и используются на всех этапах обучения. В отдельную группу можно выделить

лексико-грамматические тесты, представляющие сложный для разработки вид послетекстовых заданий.

Многие экзамены на определение уровня владения языком предусматривают возможность дистанционного пробного и/или итогового тестирования (IELTS, SIELE и др.). Особая востребованность проведения тестирования в онлайн-формате проявилась в силу ограничений, вызванных пандемией. Выросла потребность в большом количестве неповторяющихся заданий. В таких условиях создание тестов вручную становится неэффективным. Решение видится в автоматизации данного процесса, а именно в создании инструмента, способного на основе подаваемого на его входе текста генерировать задания. Можно говорить об упрочняющейся связи между задачами лингводидактики и компьютерной лингвистики, предоставляющей специфические данные и алгоритмы.

В тесте по русскому языку как иностранному (ТРКИ) раздел для проверки лексической компетенции учащегося содержит набор изолированных предложений, тогда как в большинстве других тестов (IELTS, FCE, CAE — по английскому, DELE — испанскому, CILS — итальянскому языкам) чаще используется связный текст.

Перечисленные факты свидетельствуют об актуальности представленного исследования, в котором предлагается метод генерации тестовых лексических заданий по РКИ на заполнение с множественным выбором (gap-filling & multiple-choice questions, далее — **заданий**) на основе связного текста.

Анализ источников показывает, что проблемой автоматической генерации учебных заданий активно занимаются как в России, так и за рубежом, причем охватываются практически все предметные области. Обзор методов генерации заданий дан в работах И. Ю. Балашовой, К. И. Волынской и П. П. Макарычева; В. В. Кручинина и В. В. Кузовикина; В. В. Романенко и И. О. Аксенко [Балашова, Волынская, Макарычев 2016; Кручинин, Кузовикин 2022; Романенко, Аксенко 2017]. В статье А. Ю. Малафеева [Malafeev 2015] приводится сравнительный анализ методов генерации

языковых заданий для обучения иностранному языку. В работах этого автора [Malafeev 2014; 2015; Малафеев 2015] предложены методы генерации лексико-грамматических заданий в форматах open-cloze и word-bank-cloze. В статье Е. Кузьменко и А. Финогеновой [Kuzmenko, Fenogenova 2016] описаны методы генерации заданий на материале английского языка с использованием дистрибутивно-семантических моделей (ДСМ) word2vec.

Однако сегодня отсутствует доступный программный продукт по созданию тестовых лексико-грамматических заданий с множественным выбором на основе аутентичного текста на русском языке. Такой продукт должен включать алгоритмы отбора целевых слов (target words; далее — ЦС) и релевантных дистракторов (distractors; неправильных вариантов ответа). При этом в обоих случаях необходимо учитывать сложность текста и/или языковой уровень.

Тестовые лексические задания по РКИ на заполнение с множественным выбором используются на всех этапах образовательного процесса [Ахола и др. 2017; Балыхина 2004]. Характеристики **заданий** зависят от уровня владения языком и целей тестирования. В данной статье авторы рассматривают генерацию **заданий** для учащихся, владеющих русским языком на уровнях от ТБУ (базовый уровень) (A1/A2 CEFR)¹ до ТРКИ-II (B2 CEFR)². В используемые материалы могут включаться тексты, соответствующие уровню ТРКИ-III (C1 CEFR). При достижении уровней B1–B2 обучающиеся усваивают основную лексику и необходимую для повседневной коммуникации грамматику [Амлинская и др. 2020], а на уровне C1 предусматривается работа со стилем. Лексический минимум уровня B2 составляет около 5000 единиц [Андрюшина и др. 2015], здесь проявляется семантическая неоднозначность, лексические единицы часто используются в переносных значениях. Важно, что на уровне B2 обучающийся выходит на «плато» в освоении языка [Krashen 1982; Richards 2008; Skehan 1998; Xu 2009], это увеличивает потребность в практике и контроле его языковых компетенций.

Сейчас в ТРКИ отсутствует подробное описание выбора ЦС для лексико-грамматических зада-

ний и подбора дистракторов. Поэтому предлагаемые алгоритмы разрабатывались с целью генерации **заданий**, аналогичных представленным в деверсиях различных языковых экзаменов: ЕГЭ по иностранному языку, IELTS, TOEFL, DELE и др.

1. Система генерации лексико-грамматических заданий для РКИ

1.1. Структура задания

В рамках данной работы предметом генерации являются тестовые лексические задания на заполнение с множественным выбором. Из контекста исключается ЦС, а на его месте помещается знак пропуска. Обучающемуся предлагается несколько вариантов ответа, один из которых правильный. Подбор дистракторов к ЦС сводим к задаче лексической замены (lexical substitution). На материале русского языка этому вопросу уделялось особое внимание [Anwar et al. 2020; Arefyev et al. 2020]. Приведем пример таких **заданий** без учета языкового уровня:

Дом был маленький. Он стоял в старом заброшенном саду. Ночью мы иногда просыпались от (1) _____ яблок, падавших с (2) _____ на крышу. В доме мы только ночевали. Все дни, с (3) _____ до темноты, мы проводили на берегу озера, где купались, ловили рыбу, варили на костре уху. Возвращались мы вечером, усталые, обгоревшие на солнце, со (4) _____ серебристой рыбы. И каждый раз нас встречали (5) _____ о том, что и у кого украл рыжий кот, которого в (6) _____ звали Ворюгой.

(по К. Паустовскому)

№	a	b	c
1	гудка	оклика	скрежета
2	хвой	лоз	веток
3	утро	кафе	полнолуния
4	паклей	коробкой	связкой
5	разговорами	очерками	отрывками
6	слободе	избе	заимке
№	d	e	f
1	грохота	перезвона	стука
2	порослей	ив	соломинок
3	воскресения	дежурства	вставания
4	бечевой	моткой	отмычкой
5	рассказами	анекдотами	мифами
6	ярмарке	глуши	деревне

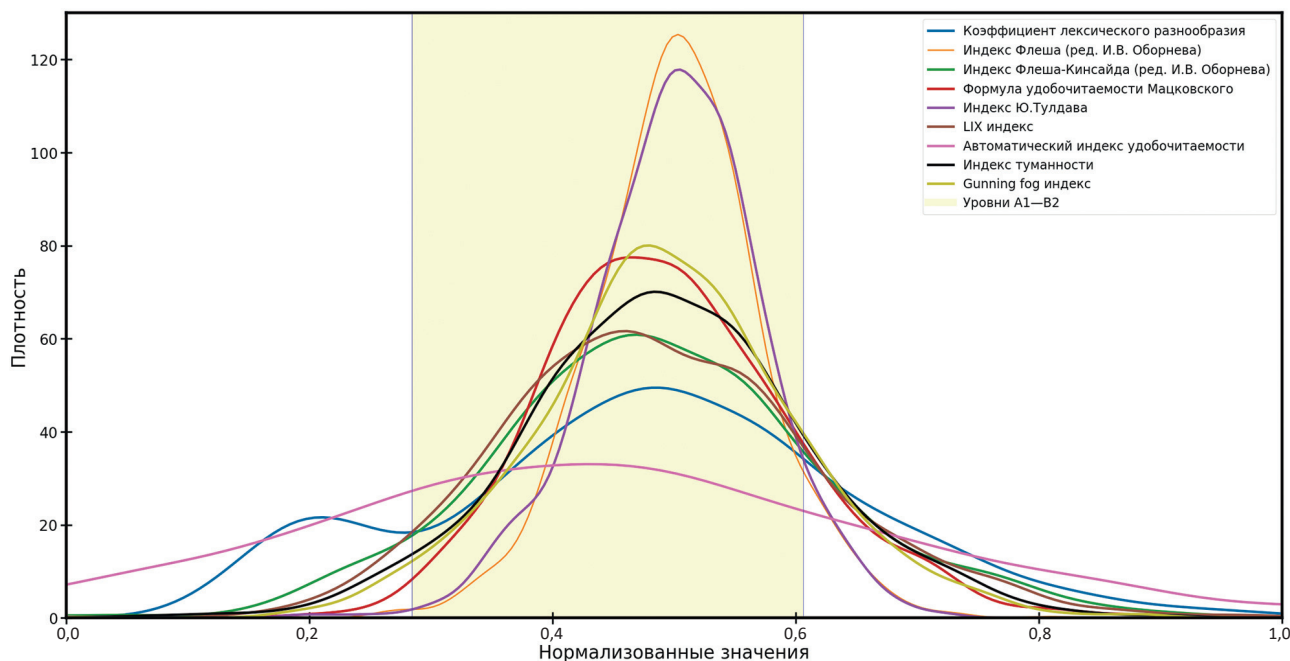
1.2. Корпус текстов для генерации заданий

С учетом опыта коллег [Цзинцин 2005] в качестве источников для корпуса выбрана детская и подростковая художественная литература — адаптированные тексты для изучающих РКИ. Тексты отбирались из «Библиотеки М. Мошкова»³, электронной библиотеки издательства «Златоуст»⁴ и с сайтов о преподавании и обучении РКИ⁵. Также использовались оригинальные и адаптированные тексты из корпуса RuAdapt [Dmitrieva, Tiedemann 2021]; школьные и вузовские учебники по русскому языку, чтению, литературе, культуре речи и др.; учебники РКИ; адаптированные классические произведения с указанным уровнем владения РКИ; публицистические тексты. Из данных источников была сформирована выборка объемом около 50 млн токенов.

Для определения языкового уровня текстов использовались следующие метрики удобочитаемости [Laposhina et al. 2018; Solovyev, Ivanov, Solnyshkina 2018]: индексы Флэша, Флэша — Кинкейда, их адаптации для русского языка [Оборнева 2006], индексы SMOG (Simple Measure of

Gobbledygook), Automated Readability Index, Gunning Fog, формулы Мацковского, Тулдавы и т. д. (рисунок). Был применен такой инструмент оценки удобочитаемости, как «Текстометр»⁶ [Лапошина, Лебедева 2021]. Данные метрики задействуют разные шкалы интерпретации их значений (возраст носителей, год обучения в школе/вузе). При отборе текстов предполагалось, что носитель достигает уровня B2–C1 приблизительно к 8–10 году обучения в школе, или 15–16-летнему возрасту [Дубинина, Птюшкин 2021; Goodier 2018], при этом были учтены различия в обучении РКИ и тестировании школьников и взрослых, связанные с методическими источниками, недостаточной разработанностью лексических минимумов для школьников и стандартов для оценки знаний по шкале CEFR. После вычисления метрик для текстов была произведена нормализация их значений с использованием библиотеки scikit-learn⁷, установлено соответствие границам уровней (нижняя граница — уровень A1 и верхняя — уровень B2) и отобраны соответствующие этим параметрам тексты.

Далее были осуществлены разметка и фильтрация корпуса: тексты приведены к нижнему ре-



Распределение текстов в соответствии со значениями метрик удобочитаемости⁸

гистру, из них удалены стоп-слова и нетекстовые токены, средствами библиотеки `srasu`⁹ произведена частеречная разметка, лемматизация и фрагментационный анализ. После преобразований объем корпуса составил 25 млн токенов.

При разметке омонимия не снималась. Выбор семантически близких лексических единиц, связанных с другим членом омонимической пары, представляет один из способов подбора дистракторов, наряду с подбором дистракторов к разным значениям многозначного слова. Так повышается сложность задания.

1.3. Обучение модели *word2vec*

На материале корпуса была обучена предсказывающая статическая модель *word2vec* [Mikolov et al. 2013], используемая для подбора дистракторов и дающая для ЦС кластер сходных с ним замен [Pérez, Cuadros 2017; Ren, Zhu 2021]. Мера семантического сходства двух слов определяется на основе косинусного расстояния между их векторами:

$$similarity = \cos \theta = \frac{\vec{d} + \vec{t}}{\|\vec{d}\| \times \|\vec{t}\|}.$$

Для предсказания замен слова в контексте более результативна архитектура CBOW (Continuous Bag of Words, ‘непрерывный мешок слов’). Модель обучалась с параметрами размерности вектора $d = 300$ и ширины контекстного окна $w = 10$ в пять эпох до сходимости с минимальным значением функции потерь.

1.4. Алгоритм выбора предложений-кандидатов

При генерации заданий применимы алгоритмы экстрактивной суммаризации [Miller 2017; Pílan, Volodina, Borin 2017]. Принято решение использовать экстрактивную суммаризацию на основе нейросетевой модели BERT в библиотеке `bert-extractive-summarization`¹⁰. Большинство источников по изучению иностранных языков предлагают 10, 15 или 20 заданий по связному тексту на уровнях A2–B2, размеры текстов варьируются

от 10 до 50 предложений и от 150 до 1200 слов. В нашем случае пользователь выбирает текст для генерации заданий. После предобработки текст сокращается до 10, 15 или 20 предложений.

1.5. Алгоритм выбора целевого слова в предложении

При выборе ЦС для замены в предложениях авторы статьи руководствовались работами М. Агарвала и П. Маннем, а также И. Пилан [Agarwal, Mannem 2011; Pílan 2016]. ЦС должны быть восстанавливаемы по контексту и отвечать ряду требований. В базовой версии алгоритм считает целевыми словами все токены, кроме стоп-слов, именованных существностей, имен собственных, служебных частей речи, числительных. Дальнейший выбор ЦС случаен, предпочтение отдается словам с большим числом зависимых. Не исключается выбор нескольких ЦС в одном предложении, однако они не должны находиться рядом и/или быть связанными отношениями зависимости. Очевидно, что ЦС уникальны для всего текста: одно и то же слово (и/или его формы) не может служить ЦС более одного раза, а также не может выступать в роли дистрактора.

1.6. Алгоритм подбора дистракторов

При подборе дистракторов ЦС лемматизируется, после чего производится запрос к модели. Из выдачи отбираются потенциальные дистракторы, входящие в выбранный пользователем лексический минимум, они относятся к той же части речи, что и ЦС, имеют сопоставимую с ними длину, не встречаются в тексте. При этом значение сходства (*likelihood*) дистракторов с ЦС ниже заранее установленной планки¹¹, леммы дистракторов и ЦС не являются орфографически близкими, что обеспечивается с помощью расстояния Левенштейна. Затем на основе полученного списка лемм дистракторов с помощью библиотеки `rumorphy`¹² образуются формы с теми же грамматическими характеристиками, что и ЦС. Наконец, случайным образом отбирается нужное число дистракторов.

1.7. Программная реализация разработки

Нами было создано веб-приложение¹³, демонстрирующее применение метода генерации заданий. Веб-интерфейс реализован с помощью библиотеки streamlit¹⁴ для Python. Материалом для генерации заданий является пользовательский текст, который загружается в поле веб-приложения напрямую или в виде текстового файла. Для одного текста можно получить задания различной сложности в зависимости от выбранного пользователем лексического минимума. Количество дистракторов варьируется от двух до девяти. Такой широкий диапазон предусмотрен с учетом того, что некоторые из дистракторов могут отсутствовать в минимумах или в ДСМ.

Возможные типы выдачи:

- версия «для ученика»: файл с заданиями и невыделенными правильными ответами;
- версия «для учителя»: файл с заданиями и выделенными правильными ответами, файл с ключами;
- вывод заданий на экран;
- вывод на экран онлайн-теста.

2. Эксперименты по генерации лексико-грамматических заданий для РКИ

Для оценки качества генерируемых заданий проведены два эксперимента с совпадающими тестовыми данными и отличающимися параметрами целевой аудитории и запрашиваемыми у респондентов ответами.

Для генерации заданий были отобраны тексты из разделов «Чтение» пособий по РКИ. Для каждого текста с помощью инструмента «Текстомер» был определен уровень для выбора лексических минимумов. С помощью метода, предложенного авторами статьи, были получены тестовые задания с шестью вариантами ответа: верный ответ и пять дистракторов. Стандартное число дистракторов в заданиях такого типа равно трем. Однако в экспериментах их число было расширено до пяти, поскольку необходимо было оценить наибольшее возможное число заданий и дистрак-

торов и предусмотреть возможность фильтрации дистракторов преподавателями.

Каждый тест был создан в двух версиях: с маркированными и немаркированными правильными ответами. В первом эксперименте участниками были наивные носители русского языка и студенты-лингвисты, не специализирующиеся на РКИ. Во втором эксперименте профессиональную оценку проводили пять экспертов — преподавателей РКИ, в том числе из Центра языкового тестирования СПбГУ. Такой выбор позволил получить значимые данные и от обучающихся, и от составителей тестов.

В ходе первого эксперимента респондентам предлагалось пройти тесты, указывая свою степень уверенности в выбранном ответе: «Укажите, насколько Вы уверены в выбранном Вами варианте. Считаете ли Вы его однозначным? Нет ли другого подходящего по смыслу варианта?» по шкале от 1 до 3: от 1 — «Совсем не уверен(-а)» до 3 — «Полностью уверен(-а)».

Формулировки определяются спецификой формы проведения эксперимента (на платформе GoogleForms) и соответствуют рекомендациям по проведению психолингвистических экспериментов [Ягунова 2005].

Результаты первого эксперимента (57 ответов респондентов) представлены в таблице. Релевантными дистракторами считались только выбранные абсолютным большинством респондентов, «полностью уверенных» в своем выборе. Были вычислены значения полноты (recall), точности (precision), аккуратности (accuracy) в видоизменной форме:

$$recall = \frac{УПО}{УПО+НОО}; \quad precision = \frac{УПО}{УПО+НПО};$$

$$accuracy = \frac{УПО+УОО}{O},$$

где УПО — количество уверенных правильных ответов; УОО — количество уверенных ошибочных ответов; НОО — количество неуверенных ошибочных ответов; НПО — количество неуверенных правильных ответов; O — общее количество ответов.

Результаты первого эксперимента

Показатель	Языковой уровень				Суммарно
	A1	A2	B1	B2	
Правильных ответов, %	90,91	97,32	100	80,07	87,88
Ошибочных ответов, %	9,09	2,68	0,00	19,93	12,12
Уверенных ответов, %	87,70	76,79	95,00	63,12	75,30
Уверенных правильных ответов, %	82,89	75,00	95,00	58,14	71,36
Уверенных ошибочных ответов, %	4,81	1,79	0,00	4,98	3,94
Неуверенных ответов, %	12,30	23,21	5,00	36,88	24,70
Неуверенных правильных ответов, %	8,02	22,32	5,00	21,93	16,52
Неуверенных ошибочных ответов, %	4,28	0,89	0,00	14,95	8,18
Recall	0,95	0,99	1,00	0,80	0,90
Precision	0,91	0,77	0,95	0,73	0,81
Accuracy	0,87	0,77	0,95	0,63	0,75

Для уровней A1–B1 испытуемые показывали наилучшие результаты. Подтверждается центральное положение уровня B1 в системе уровней. Для уровня B2 и выше подготовка данных требует дополнительных процедур в связи с особенностями текстов (высокая доля многозначных слов, средства художественной выразительности и т. д.). Малое итоговое число уверенно-ошибочных ответов значимо в данном эксперименте, так как выявляет отсутствие неоднозначности в заданиях. В то же время не утрачена их сложность, о чем говорит близкая к 25 % доля неуверенных ответов. Можно говорить об удовлетворительных значениях обобщенных показателей и сделать вывод о пригодности предложенного метода с учетом калибровки.

В ходе второго эксперимента эксперты оценивали задания на пригодность по сложности и целям тестирования в целом, однозначность решения и релевантность дистракторов.

Все эксперты отметили значимость работы. По результатам экспертизы было принято решение о внесении изменений, направленных на большую соотносимость заданий с компетенциями, которыми владеет учащийся на каждом уровне, так как в рамках проведенных экспериментов

ЦС и дистракторы подбирались со стереотипными морфосинтаксическими ограничениями.

Относительно генерируемых дистракторов было сделано два существенных наблюдения.

Во-первых, в заданиях встречаются группы нерелевантных дистракторов:

- близкие с ЦС синонимы и/или соотносимые понятия (*молчаливый/безмолвный, контроль/надзор, безмолвие/молчание, решительный/энергичный, идеология/мораль*);
- устаревшая лексика, встречающаяся в художественной литературе (*орда, делец, барыш*);
- согипонимы (*гитара/бубен, ожерелье/кольцо*);
- однокоренные слова.

Во-вторых, нерелевантные дистракторы обнаруживаются приблизительно в трети заданий, где их количество не превышает одного-двух, что позволяет использовать метод с учетом отбора релевантных дистракторов.

Выбор целевых слов с точки зрения экспертов также может быть улучшен: на низких уровнях необходимо исключить возможность попадания в набор ЦС абстрактных существительных, обо-

значений видов растений, животных, единиц измерения, переносных употреблений слов.

Таким образом, авторы приходят к выводу, что разработанный метод генерации **заданий** пригоден для применения в учебном процессе.

Заключение

Авторами разработан метод генерации лексико-грамматических заданий на заполнение с множественным выбором. Проанализированы технологии существующих программных продуктов, совмещающих принципы преподавания РКИ и других языков с алгоритмами и инструментами компьютерной лингвистики. Проведены эксперименты по оценке тестовых заданий, полученных в результате применения разработанного метода. В статье обоснована пригодность метода для использования в учебном процессе, предложены способы повышения качества заданий.

Определены перспективы развития исследования:

- расширение корпусного материала и обучение предсказывающих моделей с разными параметрами;
- управляемый выбор ЦС на основе интеграции модулей генерации заданий и модуля вероятностного тематического моделирования;
- использование предсказывающих контекстуализированных моделей типа BERT и словарных баз данных WordNet для генерации **заданий**;
- модернизация веб-приложения, например внедрение возможности самостоятельного выбора ЦС пользователем;
- реализация описанного метода для других языков.

ПРИМЕЧАНИЯ

¹ Уровни знания в системе CEFR (Common European Framework of Reference, Общеευропейские компетенции владения иностранным языком).

² Российская государственная система тестирования граждан зарубежных стран по русскому языку. (2021) *Государственный стандарт по русскому языку как иностранному*. URL: https://gct.msu.ru/docs/A1_standart.pdf; https://gct.msu.ru/docs/A2_standart.pdf; https://gct.msu.ru/docs/B1_standart.pdf; https://gct.msu.ru/docs/B2_standart.pdf; https://gct.msu.ru/docs/C1_standart.pdf (дата обращения: 15.11.2023).

³ Библиотека Максима Мошкова. URL: <http://lib.ru/> (дата обращения: 15.11.2023).

⁴ Златоуст — все для изучения и преподавания русского языка как иностранного или второго. URL: <https://zlatoust.store/> (дата обращения: 15.11.2023).

⁵ Сайты о РКИ: <https://www.ros-edu.ru/>; <https://mgu-russian.com/ru/teach/biblioteka/>; <https://rucont.ru/collections/2714>; <https://www.rus-lang.ru/books/category/8>.

⁶ Текстометр. Анализ сложности текста. URL: <https://textometr.ru/> (дата обращения: 15.11.2023).

⁷ Scikit-learn. URL: <https://scikit-learn.org/stable/> (дата обращения: 15.11.2023).

⁸ Для визуализации использовалась библиотека matplotlib (<https://matplotlib.org/>).

⁹ SpaCy. URL: <https://spacy.io/> (дата обращения: 15.11.2023).

¹⁰ Bert Extractive Summarizer. URL: <https://pypi.org/project/bert-extractive-summarizer/> (дата обращения: 15.11.2023).

¹¹ Для определения пограничных планок были проанализированы аутентичные задания из демоверсий и сборников по подготовке к ТРКИ с точки зрения меры семантического сходства ЦС и дистракторов к ним. В результате установлены средние значения для каждого уровня, которые составляют около 0,8.

¹² Морфологический анализатор rumorphy2. URL: <https://rumorphy2.readthedocs.io/en/stable/> (дата обращения: 15.11.2023).

¹³ URL: <https://huggingface.co/spaces/a-v-bely/russian-task-generator> (дата обращения: 15.11.2023).

¹⁴ Streamlit. URL: <https://streamlit.io> (дата обращения: 15.11.2023).

ЛИТЕРАТУРА

Амлинская и др. 2020 — Амлинская Ю. Р., Дубинина Н. А., Гельфрейх П. Г., Ильчева И. Ю. *Тестирование школьников по русскому языку в СПбГУ: о чем спрашивают преподаватели и родители?* 2020. URL: https://testingcenter.spbu.ru/images/webinars/webinar_6_bi.pdf (дата обращения: 08.02.2023).

Андрюшина и др. 2015 — Андрюшина Н. П., Афанасьева И. Н., Битехтина Г. А., Клобукова Л. П., Яценко И. И. *Лексический минимум по русскому языку как иностранному. Второй сертификационный уровень. Общее владение*. Андрюшина Н. П. (ред.). 5-е изд. СПб., 2015. URL: <https://www.ipr-bookshop.ru/81260.html> (дата обращения: 15.11.2023).

Ахола и др. 2017 — Ахола С., Башарин А. А., Башмакова Н. И. [и др.] *Актуальные вопросы языкового тестирования*. Павловская И. Ю. (ред.). Вып. 2. СПб.: Изд-во С.-Петербург. гос. ун-та, 2017. 684 с.

Балашова, Волынская, Макарычев 2016 — Балашова И. Ю., Волынская К. И., Макарычев П. П. *Методы и средства генерации тестовых заданий из текстов на естественном*

языке. *Модели, системы, сети в экономике, технике, природе и обществе*. 2016, № 1(17): 195–202.

Балыхина 2004 — Балыхина Т. М. *Основы теории тестов и практика тестирования (в аспекте русского языка как иностранного)*. М.: Русский язык. Курсы, 2004. 240 с.

Дубинина, Птюшкин 2021 — Дубинина Н. А., Птюшкин Д. В. Уровни тестирования по русскому языку как иностранному в аспекте возрастной специфики школьников. *Русистика*. 2021, т. 19, № 2: 222–234.

Кручинин, Кузовикин 2022 — Кручинин В. В., Кузовикин В. В. Обзор существующих методов автоматической генерации задач с условиями на естественном языке. *Компьютерные инструменты в образовании*. 2022, № 1: 85–96.

Лапошина, Лебедева 2021 — Лапошина А. Н., Лебедева М. Ю. Текстомер: онлайн-инструмент определения уровня сложности текста по русскому языку как иностранному. *Русистика*. 2021, т. 19, № 3: 331–345.

Малафеев 2015 — Малафеев А. Ю. Метод автоматического создания лексико-грамматических упражнений в формате wordbank cloze. *Иностранные языки в высшей школе*. 2015, № 2 (33): 88–95.

Оборнева 2006 — Оборнева И. В. *Автоматизированная оценка сложности учебных текстов на основе статистических параметров*. Дис. ... канд. пед. наук. М., 2006. 165 с.

Романенко, Аксенко 2017 — Романенко В. В., Аксенко И. О. Обзор технологий автоматизированного создания и публикации тестовых заданий в системах дистанционного обучения. В сб.: *Информационные технологии в науке, управлении, социальной сфере и медицине: сборник научных трудов IV Международной научной конференции*. Берестнева О. Г., Мицель А. А., Гладкова Т. А. (ред.). Томск: Изд-во ТПУ, 2017, ч. 1: 317–321.

Цзинцзин 2005 — Цзинцзин Л. Система принципов отбора учебных текстов для формирования межкультурной компетенции иностранных студентов-филологов (уровень В2). *Вестник Томского государственного педагогического университета*. 2005, № 7 (184): 128–133.

Ягунова 2005 — Ягунова Е. В. *Эксперимент в психолингвистике: конспекты лекций и методические рекомендации*. СПб.: Остров, 2005. 51 с.

Agarwal, Mannem 2011 — Agarwal M., Mannem P. Automatic Gap-fill Question Generation from Text Books. In: *Proceedings of the Sixth Workshop on Innovative Use of NLP for Building Educational Applications*. Portland, Oregon: Association for Computational Linguistics, 2011. P. 56–64.

Anwar et al. 2020 — Anwar S., Shelmanov A., Panchenko A., Biemann C. Generating Lexical Representations of Frames using Lexical Substitution. In: *Proceedings of the Probability and Meaning Conference (PaM 2020)*. Gothenburg: Association for Computational Linguistics, 2020. P. 95–103.

Arefyev et al. 2020 — Arefyev N., Sheludko B., Podolskiy A. V., Panchenko A. Always Keep your Target in Mind: Studying Semantics and Improving Performance of Neural Lexical Substitution. In: *Proceedings of the 28th International Conference on Computational Linguistics*. Barcelona: International Committee on Computational Linguistics, 2020. P. 1242–1255.

Dmitrieva, Tiedemann 2021 — Dmitrieva A., Tiedemann J. Creating an Aligned Russian Text Simplification Dataset from Language Learner Data. In: *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing*. Babych B. [et al.] (eds). Kiyv: Association for Computational Linguistics, 2021. P. 73–79.

Goodier 2018 — Goodier T. (ed.) *Collated Representative Samples of Descriptors of Language Competences Developed for Young Learners — resource for educators, Vol. 2: Ages 11–15*, Education Policy Division, Council of Europe. 2018.

Krashen 1982 — Krashen S. D. *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon Press, 1982. 202 p.

Kuzmenko, Fenogenova 2016 — Kuzmenko E., Fenogenova A. Automatic generation of lexical exercises. In: *CLLS 2016. Computational Linguistics and Language Science. Proceedings of the Workshop on Computational Linguistics and Language Science*. Aachen: CEUR Workshop Proceedings, 2016, vol. 1886. P. 20–27.

Laposhina et al. 2018 — Laposhina A., Veselovskaya T., Lebedeva M., Kupreshchenko O. Automated Text Readability Assessment for Russian Second Language Learners. In: *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue 2018”*. 2018. Issue 17 (24). P. 403–413.

Malafeev 2014 — Malafeev A. Automatic Generation of Text-Based Open Cloze Exercises. In: *Communications in Computer and Information Science, Ignatov D., Khachay M., Panchenko A., Konstantinova N., Yavorsky R. (eds) Analysis of Images, Social Networks and Texts AIST 2014*. Springer, 2014, vol. 436. P. 140–151.

Malafeev 2015 — Malafeev A. Exercise Maker: Automatic Language Exercise Generation. In: *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference “Dialogue” (2015)*. Moscow: RSUH, 2015. Issue 14 (21). P. 441–452.

Mikolov et al. 2013 — Mikolov T., Chen K., Corrado G. S., Dean J. Efficient Estimation of Word Representations in Vector Space. *International Conference on Learning Representations ICLR*. 2013. URL: https://dscomp2019.github.io/papers/Mikolov_et_al-2013-Efficient.pdf (дата обращения: 15.11.2023).

Miller 2017 — Miller D. Leveraging BERT for Extractive Text Summarization on Lectures. 2019. In: *arXiv*. URL: <https://arxiv.org/abs/1906.04165> (дата обращения: 08.02.2023).

Perez, Cuadros 2017 — Pérez N., Cuadros M. Multilingual CALL Framework for Automatic Language Exercise Generation from Free Text. In: *Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. Valencia: Association for Computational Linguistics, 2017. P. 45–52.

Pilán 2016 — Pilán I. Detecting Context Dependence in Exercise Item Candidates Selected from Corpora. In: *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*. San Diego: Association for Computational Linguistics, 2016. P. 151–161.

Pilán, Volodina, Borin 2017 — Pilán I., Volodina E., Borin L. Candidate sentence selection for language learning exercises: from a comprehensive framework to an empirical evaluation. *Traitement Automatique des Langues*. 2017, vol. 57 (3): 67–91.

Ren, Zhu 2021 — Ren S., Zhu K. Q. Knowledge-Driven Distractor Generation for Cloze-style Multiple Choice Questions. In:

The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21). Issue 5 (35). 2021. P. 4339–4347.

Richards 2008 — Richards J. *Moving Beyond the Plateau From Intermediate to Advanced Levels in Language Learning*. New York, 2008. 28 p.

Skehan 1998 — Skehan P. *A Cognitive Approach to Language Learning*. Hong Kong; Oxford, 1998. 332 p.

Solovyev, Ivanov, Solnyshkina 2018 — Solovyev V., Ivanov V., Solnyshkina M. Assessment of reading difficulty levels in Russian academic texts: Approaches and metrics. *Journal of Intelligent and Fuzzy Systems*. 2018, vol. 34, issue 5: 3049–3058.

Xu 2009 — Xu Q. Moving beyond the Intermediate EFL Learning Plateau. *Asian Social Science*. 2009, vol. 5 (2): 66–68.

REFERENCES

Амлинская и др. 2020 — Amlinskaya Yu. R., Dubinina N. A., Gelfreikh P. G., Ilcheva I. Yu. *Testing schoolchildren in the Russian language at St Petersburg State University: what do teachers and parents ask about?* 2020. Available at: https://testingcenter.spbu.ru/images/webinars/webinar_6_bi.pdf (accessed: 08.02.2023) (In Russian)

Андрюшина, Афанасьева, Битехтина, Клобукова, Яценко 2015 — Andryushina N. P., Afanasyeva I. N., Bitekhtina G. A., Klobukova L. P., Yatsenko I. I. *Lexical minimum in Russian as a foreign language. Second certification level. General language proficiency*. Andryushina N. P. (ed.). 5th ed., St. Petersburg, 2015. Available at: <https://www.iprbookshop.ru/81260.html> (accessed: 15.11.2023). (In Russian)

Ахола и др. 2017 — Ahola S., Basharin A. A., Bashmakova N. I. [et al.] *Actual issues of language testing*. Pavlovskaya I. Yu. (ed.). Issue 2. St Petersburg: St Petersburg University Press, 2017. 684 p. (In Russian)

Балашова, Волынская, Макарычев 2016 — Balashova I. Yu., Volynskaya K. I., Makarychev P. P. Methods and tools for generating test items from natural language texts. *Modeli, sistemy, seti v ekonomike, tekhnike, prirode i obshchestve*. 2016, no. 1 (17): 195–202. (In Russian)

Балыхина 2004 — Balykhina T. M. *Fundamentals of the theory of tests and the practice of testing (in the aspect of Russian as a foreign language)*. Moscow: Russkii iazyk. Kursy Publ., 2004. 240 p. (In Russian)

Дубинина, Птюшкин 2021 — Dubinina N. A., Ptyushkin D. V. Levels of testing in Russian as a foreign language in terms of age specificity of schoolchildren. *Russian Language Studies*. 2021, vol. 19, no. 2: 222–234. (In Russian)

Кручинин, Кузовикин 2022 — Kruchinin V. V., Kuzovikin V. V. Review of existing methods for automatic generation of problems with conditions in natural language. *Kompiuternye instrumenty v obrazovanii*. 2022, no. 1: 85–96. (In Russian)

Лапошина, Лебедева 2021 — Laposhina A. N., Lebedeva M. Yu. Textometer: An online tool for determining the level of complexity of a text in Russian as a foreign language. *Russian Language Studies*. 2021, vol. 19, no. 3: 331–345. (In Russian)

Малафеев 2015 — Malafeev A. Yu. Method of automatic creation of lexico-grammatical exercises in wordbank cloze for-

mat. *Inostrannyye yazyki v vysshei shkole*. 2015, no. 2 (33): 88–95. (In Russian)

Оборнева 2006 — Osborne I. V. *Automated assessment of the complexity of educational texts based on statistical parameters*. PhD thesis. Moscow, 2006. 165 p. (In Russian)

Романенко, Аксенко 2017 — Romanenko V. V., Aksenko I. O. Review of technologies for automated creation and publication of test tasks in distance learning systems. In: *Informatsionnyye tekhnologii v nauke, upravlenii, sotsial'noy sfere i meditsine: sbornik nauchnykh trudov IV Mezhdunarodnoy nauchnoy konferentsii*. Berestneva O. G., Mitsel' A. A., Gladkova T. A. (eds). Tomsk: Izd-vo TPU Publ., 2017. Part 1. P. 317–321. (In Russian)

Цзинцзин 2005 — Jingjing L. The system of principles for the selection of educational texts for the formation of intercultural competence of foreign students-philologists (level B2). *Tomsk State Pedagogical University Bulletin*. 2017, no. 7 (184): 128–133. (In Russian)

Ягунова 2005 — Yagunova E. V. *Experiment in psycholinguistics: Lecture notes and guidelines: Textbook for universities*. St Petersburg: Ostrov Publ., 2005. 51 p. (In Russian)

Agarwal, Mannem 2011 — Agarwal M., Mannem P. Automatic Gap-fill Question Generation from Text Books. In: *Proceedings of the Sixth Workshop on Innovative Use of NLP for Building Educational Applications*. Portland, Oregon: Association for Computational Linguistics, 2011. P. 56–64.

Anwar et al. 2020 — Anwar S., Shelmanov A., Panchenko A., Biemann C. Generating Lexical Representations of Frames using Lexical Substitution. In: *Proceedings of the Probability and Meaning Conference (PaM 2020)*. Gothenburg: Association for Computational Linguistics, 2020. P. 95–103.

Arefyev et al. 2020 — Arefyev N., Sheludko B., Podolskiy A. V., Panchenko A. Always Keep your Target in Mind: Studying Semantics and Improving Performance of Neural Lexical Substitution. In: *Proceedings of the 28th International Conference on Computational Linguistics*. Barcelona: International Committee on Computational Linguistics, 2020. P. 1242–1255.

Dmitrieva, Tiedemann 2021 — Dmitrieva A., Tiedemann J. Creating an Aligned Russian Text Simplification Dataset from Language Learner Data. In: *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing*. Babych B. [et al.] (eds). Kyiv: Association for Computational Linguistics, 2021. P. 73–79.

Goodier 2018 — Goodier T. (ed.) *Collated Representative Samples of Descriptors of Language Competences Developed for Young Learners — resource for educators, Vol. 2: Ages 11–15*, Education Policy Division, Council of Europe. 2018.

Krashen 1982 — Krashen S. D. *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon Press, 1982. 202 p.

Kuzmenko, Fenogenova 2016 — Kuzmenko E., Fenogenova A. Automatic generation of lexical exercises. In: *CLLS 2016. Computational Linguistics and Language Science. Proceedings of the Workshop on Computational Linguistics and Language Science*. Aachen: CEUR Workshop Proceedings, 2016, vol. 1886. P. 20–27.

Laposhina et al. 2018 — Laposhina A., Veselovskaya T., Lebedeva M., Kupreshchenko O. Automated Text Readability Assessment for Russian Second Language Learners. In: *Computational Linguistics and Intellectual Technologies: Proceedings of the*

International Conference "Dialogue 2018". 2018. Issue 17 (24). P.403–413.

Malafeev 2014 — Malafeev A. Automatic Generation of Text-Based Open Cloze Exercises. In: *Communications in Computer and Information Science, Ignatov D., Khachay M., Panchenko A., Konstantinova N., Yavorsky R. (eds) Analysis of Images, Social Networks and Texts AIST 2014*. Springer, 2014, vol. 436. P.140–151.

Malafeev 2015 — Malafeev A. Exercise Maker: Automatic Language Exercise Generation. In: *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue" (2015)*. Moscow: RSUH Publ., 2015. Issue 14 (21). P.441–452.

Mikolov et al. 2013 — Mikolov T., Chen K., Corrado G.S., Dean J. Efficient Estimation of Word Representations in Vector Space. *International Conference on Learning Representations ICLR*. 2013. URL: https://dscomp2019.github.io/papers/Mikolov_et_al-2013-Efficient.pdf (дата обращения: 15.11.2023).

Miller 2017 — Miller D. Leveraging BERT for Extractive Text Summarization on Lectures. 2019. In: *arXiv*. URL: <https://arxiv.org/abs/1906.04165> (дата обращения: 08.02.2023).

Perez, Cuadros 2017 — Pérez N., Cuadros M. Multilingual CALL Framework for Automatic Language Exercise Generation from Free Text. In: *Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. Valencia: Association for Computational Linguistics, 2017. P.45–52.

Pilán 2016 — Pilán I. Detecting Context Dependence in Exercise Item Candidates Selected from Corpora. In: *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*. San Diego: Association for Computational Linguistics, 2016. P.151–161.

Pilán, Volodina, Borin 2017 — Pilán I., Volodina E., Borin L. Candidate sentence selection for language learning exercises: from a comprehensive framework to an empirical evaluation. *Traitement Automatique des Langues*. 2017, vol. 57 (3): 67–91.

Ren, Zhu 2021 — Ren S., Zhu K. Q. Knowledge-Driven Distractor Generation for Cloze-style Multiple Choice Questions. In: *The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21)*. Issue 5 (35). 2021. P.4339–4347.

Richards 2008 — Richards J. *Moving Beyond the Plateau From Intermediate to Advanced Levels in Language Learning*. New York, 2008. 28 p.

Skehan 1998 — Skehan P. *A Cognitive Approach to Language Learning*. Hong Kong; Oxford, 1998. 332 p.

Solovyev, Ivanov, Solnyshkina 2018 — Solovyev V., Ivanov V., Solnyshkina M. Assessment of reading difficulty levels in Russian academic texts: Approaches and metrics. *Journal of Intelligent and Fuzzy Systems*. 2018, vol. 34, issue 5: 3049–3058.

Xu 2009 — Xu Q. Moving beyond the Intermediate EFL Learning Plateau. *Asian Social Science*. 2009, vol. 5 (2): 66–68.