

Белый Андрей Владимирович
Belyi Andrei Vladimirovich
a.v.belij@yandex.ru
+79119353512

Студент
Филологический факультет, кафедра математической лингвистики, Санкт-Петербургский государственный университет
199034, Санкт-Петербург, Университетская наб., 7-9-11
Student
Department of Mathematical Linguistics, Philological Faculty, Saint-Petersburg State University
199034, Russia, St. Petersburg, Universitetskaya emb., 7-9-11

Митрофанова Ольга Александровна
Mitrofanova Olga Alexandrovna
o.mitrofanova@spbu.ru
+79117181216

Кандидат филологических наук, доцент, кафедра математической лингвистики, филологический факультет, Санкт-Петербургский государственный университет
199034, Санкт-Петербург, Университетская наб., 7-9-11
PhD, Associate Professor, Department of Mathematical Linguistics, Philological Faculty, Saint-Petersburg State University
199034, Russia, St. Petersburg, Universitetskaya emb., 7-9-11

Дубинина Надежда Александровна
Dubinina Nadezhda Alexandrovna
dnadine@yandex.ru
+79213711587

Специалист-тестор, Центр языкового тестирования, Санкт-Петербургский государственный университет
199034, Санкт-Петербург, наб. Лейтенанта Шмидта, 11/2, каб. 210
Language testing specialist, Language Testing Centre, Saint-Petersburg State University
199034, Russia, St. Petersburg, Lieutenant Schmidt emb., 11/2, room 210

Автоматическая генерация лексико-грамматических заданий по РКИ с помощью предсказывающих языковых моделей

Automatic generation of lexico-grammatical tests for RFL using predictive language models

Аннотация

При обучении иностранному языку одной из базовых потребностей участников образовательного процесса является достаточное количество учебных материалов. Среди заданий, способствующих усвоению лексико-грамматических конструкций, задания на заполнение с множественным выбором (multiple-choice gap-filling exercises) получили особую популярность. В настоящее время создание большого количества уникальных заданий вручную оказывается трудозатратным. В отличие от английского, для русского языка разработка алгоритмов автоматической генерации упражнений не ведется столь активно, несмотря на существующую потребность. В связи с этим в статье предлагается метод автоматической генерации заданий данного типа по русскому языку как иностранному (РКИ). Предлагаемый метод основывается на использовании дистрибутивно-семантических моделей типа word2vec и позволяет создавать задания на основе аутентичного текста. Представленный в работе метод не

зависит от жанрово-стилевой принадлежности текста, соотносимого с ним языкового уровня, может быть легко адаптирован для других языков. Для создания модели word2vec был собран корпус детской и учебной литературы, моделирующий языковой опыт обучаемых РКИ. В ходе работы создано веб-приложение для преподавателей. Для оценки состоятельности и релевантности получаемых заданий были проведены два эксперимента: по тестированию наивных носителей русского языка и по опросу экспертов-составителей тестов по РКИ. Их результаты подтверждают эффективность предлагаемого метода, демонстрируют высокую степень корректности заданий и подбираемых дистракторов: 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

Благодарности

Авторы выражают благодарность экспертам в области ТРКИ к.ф.н., доц. Н.В. Семеновой, к.ф.н., доц. К.А. Щукиной, к.ф.н., доц. В.А. Шульгинову, анонимной коллеге и студентам-лингвистам, принявшим участие в тестировании учебных материалов.

Исследование проводится в рамках НИП СПбГУ № 75254082 «Моделирование коммуникативного поведения жителей российского мегаполиса в социально-речевом и прагматическом аспектах с привлечением методов искусственного интеллекта»

Abstract

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, we 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.

Key words

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 TORFL experts Ph.D., associate professor N.V. Semenova; Ph.D., associate professor K.A. Shchukina; Ph.D. associate professor V.A. Shulginov; an anonymous colleague, and students, who participated in the testing of didactic materials.

1. Введение

Языковые тесты, направленные на оценку компетенций учащихся, играют огромную роль в процессе освоения иностранного языка. Они являются эффективным средством контроля и используются на всех этапах обучения. В отдельную группу можно выделить лексико-грамматические тесты, представляющие сложный для разработки вид послетекстовых заданий.

В последнее время наблюдается тенденция к автоматизации языкового тестирования. Многие экзамены на определение уровня владения языком предусматривают возможность дистанционного пробного и/или итогового тестирования (например, IELTS, SIELE и др.). В силу ограничений, вызванных пандемией, проявляется особая востребованность проведения тестирования в онлайн формате. Это повышает потребность в большом количестве неповторяющихся заданий. В таких условиях создание тестов вручную становится неэффективным. Решение видится в автоматизации данного процесса, а именно в создании генератора, способного на основе подаваемого на его вход текста создавать задания данного типа. Можно говорить об упрочняющейся связи между задачами лингводидактики и компьютерной лингвистики, предоставляющей специфические данные и алгоритмы.

Специалисты отмечают, что в тесте по русскому языку как иностранному (ТРКИ) раздел для проверки лексической компетенции учащегося представляет собой набор изолированных предложений, тогда как в большинстве других экзаменов (например, IELTS, FCE, CAE – по английскому; DELE – по испанскому, CILS – по итальянскому) чаще используется связный текст.

Вышеперечисленные факты свидетельствуют об актуальности данной работы, в которой предлагается метод генерации тестовых лексических заданий по РКИ на заполнение с множественным выбором (*gap filling & multiple choice questions*, далее **заданий**) на основе связного текста.

Проведенный 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 [Балашова и др. 2016; Кручинин, Кузовикин 2022; Романенко, Аксенко 2017]. В статье [Malafeev 2015] приводится сравнительный анализ методов генерации языковых заданий для обучения иностранному языку и использующих их программных решений. В работах [Malafeev 2014; Malafeev 2015; Малафеев 2015] предложены методы генерации лексико-грамматических заданий в форматах *open-cloze* и *word-bank-cloze*. В статье [Kuzmenko, Fenogenova 2016] описаны методы генерации заданий на материале английского языка с использованием ДСМ (дистрибутивно-семантических моделей) *word2vec*.

Однако сейчас отсутствует доступный и состоятельный 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 должна учитываться 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. Мы рассматриваем генерацию **заданий** для учащихся, владеющих русским языком на уровнях от ТБУ (A1/A2 CEFR¹) до ТРКИ-II (B2 CEFR)². В используемые материалы могут включаться тексты, соответствующие уровню ТРКИ-III (C1 CEFR). Наш выбор связан с тем, что 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 5.000 лексических единиц [Андрюшина и др. 2015], где 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 его потребность в практике и в контроле языковых компетенций.

Сейчас в ТРКИ отсутствует подробное описание выбора ЦС для лексико-грамматических заданий и подбора дистракторов. Поэтому предлагаемые нами алгоритмы разрабатывались с целью получения заданий, аналогичных представленным в демоверсиях различных языковых экзаменов: ЕГЭ по иностранным языкам, IELTS, TOEFL, DELE, и т. д.

2. Реализация системы генерации лексико-грамматических заданий для РКИ

2.1. Структура заданий

В рамках данной работы предметом генерации являются тестовые лексические задания на заполнение с множественным выбором. В них из контекстов исключается ЦС, а на его месте помещается знак пропуска. Обучающемуся предлагается несколько вариантов ответа, только один из которых правильный. Подбор дистракторов к ЦС сводим к задаче лексической замены (lexical substitution). На материале русского языка особое внимание этому вопросу уделено в работах [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	d	e	f
1	гудка	оклика	скрежета	грохота	перезвона	стука
2	хвой	лоз	веток	порослей	ив	соломинок
3	утро	кафе	полнолуния	воскресения	дежурства	вставания
4	наклей	коробкой	связкой	бечевкой	моткой	отмычкой
5	разговорами	очерками	отрывками	рассказами	анекдотами	мифами
6	слободе	избе	заимке	ярмарке	глуши	деревне

2.2. Корпус текстов для генерации заданий

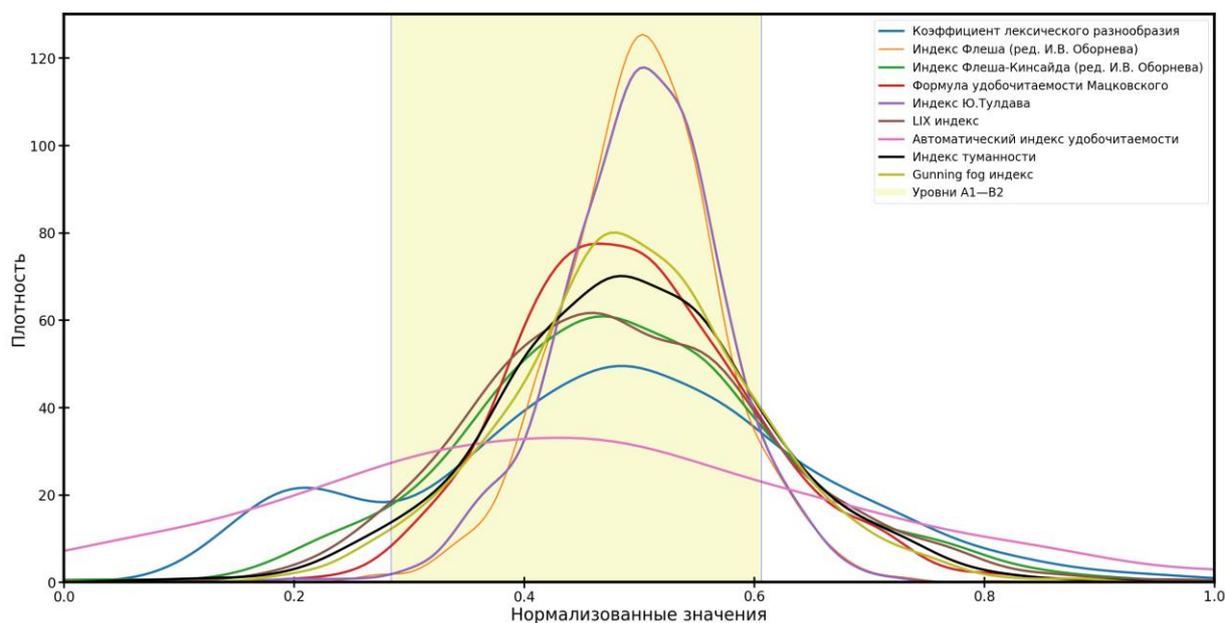
С учетом опыта коллег [Цзинцзин 2005] в качестве источников для корпуса была выбрана детская и подростковая художественная литература, адаптированные тексты

¹ Common European Framework of Reference

² 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

для изучающих РКИ. Тексты отбирались из Библиотеки М. Мошкова³, электронной библиотеки издательства Златоуст⁴ и 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 различных авторов и изданий; 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 метрики удобочитаемости [Laposhina et al. 2018; Solovyev et al. 2018 и др.]: индексы Флэша, Флэша-Кинкейда, их адаптации для русского языка [Оборнева 2006], SMOG, Automated Readability Index, (Gunning) Fog, формулы Мацковского, Тулдавы и т.д., см. графики на рис. 1. Были применены инструменты оценки удобочитаемости, в частности, «Текстометр»⁶ [Лапошина, Лебедева 2021]. Данные метрики задействуют разные шкалы интерпретации их значений (возраст носителей, год обучения в школе/ВУЗе). При отборе текстов мы полагали, что носитель достигает уровня, эквивалентного B2–C1, приблизительно к 8–10 году обучения в школе, или 15–16 годам [Дубинина, Птюшкин 2021; Goodier 2018], при этом учли различия в обучении РКИ и тестировании школьников и взрослых, связанные с методическими источниками, с недостаточной разработанностью лексических минимумов для школьников и стандартов для оценки знаний по шкале CEFR. После вычисления метрик для текстов произведена нормализация их значений с использованием библиотеки scikit-learn⁷, установлено соответствие границам уровней (нижняя A1 и верхняя B2) и отобраны соответствующие этим параметрам тексты.



³ URL: <http://lib.ru/>

⁴ URL: <https://zlatoust.store/>

⁵ URL: <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 URL: <https://textometr.ru/>

⁷ URL: <https://scikit-learn.org/stable/>

Рисунок 1 – Соотношение количества текстов и метрик удобочитаемости текстов⁸
Figure 1 – The ratio of the number of texts and the readability metrics

Далее была произведена разметка и фильтрация корпуса: тексты приведены к нижнему регистру, удалены стоп-слова и нетекстовые токены, средствами библиотеки `spacy`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.

2.3. 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:

$$\text{similarity} = \cos\theta = \frac{\vec{a} \times \vec{t}}{\|\vec{a}\| \times \|\vec{t}\|}.$$

Для предсказания замен слова в контексте более результативна архитектура `cbow` (continuous bag of words). Модель обучалась с параметрами размерности вектора $d=300$ и ширины контекстного окна $w=10$ в пять эпох до сходимости с минимальным значением функции потерь.

2.4. Алгоритм выбора предложений-кандидатов

При генерации заданий применимы алгоритмы экстрактивной суммаризации [Miller 2017; Pilán et al. 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.

2.5. Алгоритм выбора целевого слова в предложении

При выборе ЦС для замены в предложениях мы руководствуемся работами [Agarwal, Mannem 2011; Pilán 2016]. ЦС должны быть восстановимы по контексту и отвечать ряду требований. В базовой версии алгоритм считает целевыми словами все токены кроме стоп-слов, именованных сущностей, 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.

⁸ Для визуализации использовалась библиотека `matplotlib` URL: <https://matplotlib.org/>

⁹ URL: <https://spacy.io/>

¹⁰ URL: <https://pypi.org/project/bert-extractive-summarizer/>

2.6. 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, 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 `pymorphy2`¹² образуются формы с теми же грамматическими характеристиками, что и ЦС. Наконец, случайным образом отбирается нужное число дистракторов.

2.7. Программная реализация

Нами было создано веб-приложение¹³, демонстрирующее применение метода генерации заданий. Веб-интерфейс реализован с помощью библиотеки `streamlit`¹⁴ для Python. Материалом для генерации заданий является пользовательский текст, который загружается напрямую в поле веб-приложения или в виде текстового файла. Для одного текста можно получить задания различной сложности в зависимости от выбранного пользователем лексического минимума. Количество дистракторов может варьировать от 2 до 9. 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.

3. Эксперименты по генерации лексико-грамматических заданий для РКИ

Для оценки качества генерируемых заданий проведены два эксперимента с совпадающими тестовыми данными и отличающихся целевой аудиторией и запрашиваемыми у респондентов данными.

Для генерации заданий из пособий по РКИ были отобраны тексты из разделов «Чтение». Для каждого текста с помощью инструмента «Текстометр» был определен уровень для выбора лексических минимумов. С помощью нашего метода 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 было необходимо оценить наибольшее возможное число заданий и дистракторов и предусмотреть возможность фильтрации дистракторов преподавателями.

Каждый тест был создан в двух версиях: с маркированными и немаркированными правильными ответами. В первом эксперименте участниками являлись наивные носители РЯ и студенты-лингвисты, не специализирующиеся на РКИ. Во втором

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 URL: <https://pymorphy2.readthedocs.io/en/stable/>

¹³ URL: <https://a-v-beliy-rfl-task-generator.streamlit.app/>.

¹⁴ URL: <https://streamlit.io>

эксперименте профессиональную оценку проводили пять экспертов-преподавателей РКИ, в том числе ЦЯТ (Центра языкового тестирования) СПбГУ. Такой выбор позволил получить значимые данные от обучающихся и составителей тестов.

В первом эксперименте респондентам предлагалось пройти тесты, указывая 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?» со шкалой от 1 до 3, где 1 – «Совсем не уверен(-а)», а 3 – «Полностью уверен(-а)». Формулировки определяются спецификой формы проведения эксперимента (на платформе GoogleForms) и соответствуют рекомендациям по проведению психолингвистических экспериментов [Ягунова 2005].

Во втором эксперименте эксперты оценивали **задания** в целом на пригодность по сложности и целям тестирования, однозначность решения и релевантность дистракторов.

Результаты первого эксперимента (57 ответов респондентов) представлены в Таблице 1. Релевантными дистракторами считались только выбранные абсолютным большинством респондентов, «полностью уверенных» в своем выборе. Были вычислены значения полноты (recall), точности (precision), аккуратности (accuracy) в видоизмененной форме:

$$recall = \frac{УПО}{УПО+НОО}; precision = \frac{УПО}{УПО+НПО}; accuracy = \frac{УПО+УОО}{O}, \text{ где}$$

УПО – количество уверенно-правильных ответов,

УОО – количество уверенно-ошибочных ответов,

НОО – количество неуверенно ошибочных ответов,

НПО – количество неуверенно правильных ответов,

O – общее количество ответов.

Таблица 1 – Результаты первого эксперимента

Table 1 – Results of the first experiment

Показатели	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

В первом эксперименте 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.

Во втором эксперименте 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 наш 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, обозначений видов растений, животных, единиц измерения, переносных употреблений слов.

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

4. Заключение

Авторами был разработан и реализован 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 его пригодность 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.

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. Режим доступа:

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*. Н. П. Андрюшина (ред.) 5-е изд., СПб.: (электронное издание), 2015.

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 СПбГУ, 2017. 684 с.

Балашова и др. 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)*: учеб. пособие. М.: Русский язык. Курсы, 2004. 240 с.

Блинова, Тарасов 2021 — Блинова О. В., Тарасов Н. А. Сложность русских правовых текстов: методы оценки и языковые данные. В сб.: *Труды международной конференции «Корпусная лингвистика-2021»*. Захаров В.П. (ред.). СПб.: Издательство Скифия-принт, 2021. С. 175–182.

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. Текстометр: 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. ... кандидата педагогических наук. Ин-т содержания и методов обучения РАО. 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*. О.Г. Берестнева, А. А. Мицель, Т. А. Гладкова (ред.). Томск: Изд-во ТПУ, 2017. Ч.1: С. 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*. 2017, №7(184): 128–133.

Ягунова 2005 — Ягунова Е.В. Эксперимент в психолингвистике: Конспекты лекций и методические рекомендации: учебное пособие для вузов. СПб.: Издательство «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, 2011. P. 56–64.

Anwar, Shelmanov, Panchenko, Biemann 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, 2020. P. 95–103.

Arefyev, Sheludko, Podolskiy, Panchenko 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, 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.] (ed.). Stroudsburg, 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 Inc., 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, Veselovskaya, Lebedeva, Kupreshchenko 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.

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

Pérez, 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, 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, 2016. P. 151–161.

Pilán et al. 2017 — Pilán I., Volodina E., Borin L. Candidate sentence selection for language learning exercises: from a comprehensive framework to an empirical evaluation. In: *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, NY, 2008. 28 p.

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

Solovyev et al. 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. Access mode: https://testingcenter.spbu.ru/images/webinars/webinar_6_bi.pdf (accessed 08.02.2023) (In Russian)

Андрюшина и др. 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. N. P. Andryushina (ed.) 5th ed., St. Petersburg: (electronic edition), 2015. (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: Publishing House of St. Petersburg State University, 2017. 684 p. (In Russian)

Балашова и др. 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, №1(17): 195–202. (In Russian)

Бальхина 2004 — Balykhina T. M. *Fundamentals of the theory of tests and the practice of testing (in the aspect of Russian as a foreign language)*: Textbook. M.: Russian language. Courses, 2004. 240 p. (In Russian)

Блинова, Тарасов 2021 — Blinova O. V., Tarasov N. A. The complexity of Russian legal texts: assessment methods and language data. In: *Trudy mezhdunarodnoy konferentsii «Korpusnaya lingvistika-2021»*. Zakharov V.P. (red.). SPb.: Izdatel'stvo Skifiya-print, 2021. P. 175–182. (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. *Rusistika*. 2021, Vol.19, No.2: 222–234. (In Russian)

Кручинин, Кузовикин 2022 — Kruchinin VV, Kuzovikin VV Review of existing methods for automatic generation of problems with conditions in natural language. *Komp'yuternyye instrumenty v obrazovanii*. 2022, №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. *Rusistika*. 2021, Vol.19, No.3: 331–345. (In Russian)

Малафеев 2015 — Malafeev A. Yu. Method of automatic creation of lexicogrammatical exercises in wordbank cloze format. *Inostrannyye yazyki v vysshey shkole*. 2015, №2(33): 88–95. (In Russian)

Оборнева 2006 — Osborneva I. V. *Automated assessment of the complexity of educational texts based on statistical parameters*. Dis. ... kandidata pedagogicheskikh nauk. In-t sodержaniya i metodov obucheniya RAO. M., 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*. O.G. Berestneva, A. A. Mitsel', T. A. Gladkova (red.). Tomsk: Izd-vo TPU, 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). *Vestnik Tomskogo gosudarstvennogo pedagogicheskogo universiteta*. 2017, №7(184): 128–133. (In Russian)

Ягунова 2005 — Yagunova E.V. Experiment in psycholinguistics: Lecture notes and guidelines: textbook for universities. St. Petersburg: Ostrov Publishing House, 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, 2011. P. 56–64.

Anwar, Shelmanov, Panchenko, Biemann 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, 2020. P. 95–103.

Arefyev, Sheludko, Podolskiy, Panchenko 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, 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.] (ed.). Stroudsburg, 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, 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, Veselovskaya, Lebedeva, Kupreshchenko 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.

- Miller 2017 — Miller D. Leveraging BERT for Extractive Text Summarization on Lectures. 2019. In: arXiv. Режим доступа: <https://arxiv.org/abs/1906.04165> (дата обращения 08.02.2023)
- Pérez, 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, 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, CA, 2016. P. 151–161.
- Pilán et al. 2017 — Pilán I., Volodina E., Borin L. Candidate sentence selection for language learning exercises: from a comprehensive framework to an empirical evaluation. In: *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, NY, 2008. 28 p.
- Skehan 1998 — Skehan P. *A Cognitive Approach to Language Learning*. Hong Kong/Oxford, 1998. 332 p.
- Solovyev et al. 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.