

УДК 372.881.161.1

<https://doi.org/10.25587/2587-5604-2026-1-45-58>

Оригинальная научная статья

Оценка сложности текста с помощью лингвистического анализатора RuLingva (на примере текстов русских народных песен)

М. Б. Казачкова¹, Х. Н. Галимова²

¹Одинцовский филиал Московского государственного института международных отношений (университет), г. Одинцово, Российская Федерация

✉ m.kazachkova@odin.mgimo.ru

²Казанский инновационный университет им. В. Г. Тимирязова (ИЭУП),

г. Казань, Российская Федерация

✉ galikha@mail.ru

Аннотация

Цель исследования: продемонстрировать возможности функционала платформы RuLingva при отборе и модификации текстов для преподавания русского как иностранного (РКИ) на примере материала русских народных песен. Материал исследования: 107 русских народных песен общим объемом 10 517 словоформ. Исследование включало следующие алгоритмы: сбор материала исследования и ранжирование песен по тематическим циклам; создание корпуса русских народных песен; определение индекса удобочитаемости FRGL, индекса лексического разнообразия TTR, уровня языка по общеевропейской компетенции владения иностранным языком и количества вхождений единиц культурного кода русских народных песен. В ходе исследования выявлено, что русская народная песня предстает в процессе обучения РКИ, как оригинальный учебный текст, факт культуры иностранного языка, целесообразный для учебной коммуникации и имеющий знаковую функцию. Лингвистический анализатор RuLingva позволяет определить сложность текста и соответствие его сложности уровню знаний обучающихся, а также позволяет оптимизировать учебные материалы для конкретных групп обучающихся и дает оценку прогресса в виде разработанных тестов на оценку словарного запаса. Данная программа может быть полезна учителям, разработчикам контрольно-измерительных материалов (КИМ) и применима для учебного русскоязычного текста.

Ключевые слова: текст, русский как иностранный, песня, сложность текста, лингвистический анализатор, индекс читабельности, лексическое разнообразие, культурный код, Индекс Флеша-Кинкейда, профайлера RuLingva.

Финансирование. Исследование не имело финансовой поддержки.

Для цитирования: Казачкова М.Б., Галимова Х.Н. Оценка сложности текста с помощью лингвистического анализатора RuLingva (на примере текстов русских народных песен). *Вестник Северо-Восточного федерального университета им. М.К. Аммосова, Vestnik of North-Eastern Federal University. Серия «Педагогика. Психология. Философия»*. *Pedagogics. Psychology. Philosophy*. 2026: 41 (1): С. 45-58. DOI: 10.25587/2587-5604-2026-1-

Original article

Assessing text complexity using the RuLingva linguistic analyzer: The case of Russian folk song lyrics

Mariia B. Kazachkova¹, Khalida N. Galimov²

¹Odintsovo Branch, Moscow State Institute of International Relations (University),

Odintsovo, Russian Federation

✉ m.kazachkova@odin.mgimo.ru

²V. G. Timiryasov Kazan Innovative University, Kazan, Russian Federation

✉ galikha@mail.ru

Abstract

The objective of the study was to demonstrate the capabilities of the RuLingva platform for selecting and modifying texts for teaching Russian as a foreign language (RFL) using Russian folk songs as an example. The study included 107 Russian folk songs, comprising 10,517 word forms. The study included the following algorithms: collecting research material and ranking songs by thematic cycles; creating a corpus of Russian folk songs; determining the FRGL readability index, the TTR lexical diversity index, the language level according to the Common European Framework of Reference for Languages (CEFR), and the number of occurrences of cultural code units in Russian folk songs. The study revealed that Russian folk songs are presented in the process of teaching RFL as an original educational text, a fact of foreign language culture, suitable for educational communication and possessing a symbolic function. The RuLingva linguistic analyzer allows one to determine the complexity of a text and its correspondence to the level of knowledge of students, as well as to optimize teaching materials for specific groups of students and provides an assessment of progress in the form of developed tests for assessing vocabulary. This program can be useful for teachers and developers of assessment and measurement materials (AMMs) and is applicable to Russian-language educational texts.

Keywords: text, Russian as a foreign language, song, text complexity, linguistic analyzer, readability index, lexical diversity, cultural code, Flesch–Kincaid Index, RuLingva profiler.

Funding: The study did not have financial support.

For citation: Kazachkova M.B., Galimova Kh.N. Assessing text complexity using the RuLingva linguistic analyzer: The case of Russian folk song lyrics. *Vestnik of North-Eastern Federal University. Pedagogics. Psychology. Philosophy.* 2026: 41 (1). Pp. 45-58. DOI: 10.25587/2587-5604-2026-1-45-58

Введение

Правильный отбор учебного текста всегда предполагает его соответствие лингвистическим и когнитивным способностям обучающихся. «Государственная экспертиза современных российских учебников предъявляет к учебному тексту требование ориентированности на возраст учащихся, однако не предусматривает оценку сложности текста и его соответствия читательскому адресу» [1, с. 90].

Сложность текста – это комплексная характеристика, которая показывает, насколько труден текст для восприятия и понимания конкретным читателем или целевой аудиторией. Это относительная мера, зависящая от взаимодействия особенностей самого текста и подготовленности читателя.

Лингвистическая сложность текста рассчитывается на основе трёх факторов: количественных и качественных параметров текста, а также уровня подготовки читателей [2].

Учебные тексты, должны быть адаптированы к уровню понимания целевой аудитории. Здесь на помощь приходят анализаторы сложности текста – инструменты, позволяющие не только оценить уровень сложности текста и, при необходимости, предложить пути его улучшения [3].

Представленное исследование имеет **целью** продемонстрировать возможности функционала платформы RuLingva при отборе и модификации текстов для преподавания русского как иностранного (РКИ) на примере материала русских народных песен.

Отбор текстового материала был обусловлен тем, что именно народная песня предстает в процессе обучения как аутентичный и естественный текст, «целесообразный для учебной коммуникации и имеющий знаковую функцию» [4, с. 77].

Песня, созданная носителями языка в естественных условиях общения, а не специально для учебных целей, как ничто другое реализует принцип аутентичности. Она представляет собой уникальный сплав языка, музыки и истории, предлагая учащимся живой и эмоционально окрашенный материал для освоения. Это выражается в отборе лексического материала [5] – контента используемых песен, который не был адаптирован и, позволяющий погрузиться в культурный контекст и особенности языка.

Гипотеза исследования: текстовый анализатор RuLingva – доступный инструмент, с помощью которого можно оценить сложность текста, подобрать учебный материал для определенной категории обучающихся, тем самым вовлечь их в процесс обучения.

Материал исследования составили тексты 107 русских народных песен, общим объемом 10 517 словоформ.

Источники материала включают электронные ресурсы, содержащие сборник песен «Песни на уроках РКИ» под ред. Н. Г. Нестеровой, альбом «Народные песни (Душевные хиты)», картотеку народных песен Образовательной социальной сети, Русские народные песни на сайтах Классика на РуСтих, Правмир.

В качестве **инструмента** для оценки сложности текстов, извлечения элементов культурного кода был использован анализатор сложности RuLingva (свидетельство о государственной регистрации программы для ЭВМ. Номер свидетельства: RU 2022684717), разработанный исследователями Казанского (Приволжского) федерального университета.

Исследование. Исследование осуществлено по следующему алгоритму:

1. Сбор материала исследования и ранжирование песен по тематическим циклам: 1) песни, связанные с календарными и семейными ритуалами, 2) песни-произведения детского фольклора, 3) плясовые/хороводные песни, 4) лирические песни, 5) солдатские/героические песни [Акимова, 1966: 98] (см прил. № 1).

Результатом данной работы стал лингвистический Корпус русских народных песен (см табл. 1).

Таблица 1

Объем лингвистического Корпуса русских народных песен

Table 1

The volume of the linguistic corpus of Russian folk songs

Корпус русских народных песен	Количество песенных текстов	Слово употреблений
Песни, связанные с календарными датами и семейными ритуалами	9	788
Детский фольклор	3	182
Плясовые/хороводные	29	2 673
Лирические песни	61	6 522
Солдатские/героические	5	352
ИТОГО:	107	10 517

2. Определение уровня и сложности текстов песен на основе качественных и количественных параметров с помощью анализатора RuLingva.

Обзор литературы. Оценка доступности текста представляется чрезвычайно актуальной и трудоёмкой задачей в процессе подготовки текстов для преподавания русского языка как иностранного [6].

С развитием технологий «автоматической обработки естественного языка в мировой научной практике появляются работы, посвященные возможностям автоматизации процесса оценки доступности текста» [7]. Все анализаторы решают одну главную задачу: автоматически проанализировать текст и, принимая во внимание его количественные и качественные параметры, оценить его уровень сложности для подборки и адаптации материала под целевую аудиторию.

Среди существующих сервисов по анализу русскоязычных текстов можно отметить ресурс PlainRussian.ru (<https://www.plainrussian.ru>). Это бесплатный онлайн-инструмент для оценки понятности текстов и простоты восприятия материала с использованием пяти формул (Формула Flesch-Kincaid, Индекс Колман-Лиау, Формула Дэйла-Чейла, Automatic Readability, SMOG). В финале проверки выводится статистика: процент лексического «мусора», число знаков, букв, слов, количество сложных слов и длинных предложений [7].

Открытые платформенные решения предлагают текстовый профайлер ГлавРед – популярный инструмент для приведения текста к информационному виду. Мощный инструмент, созданный российским редактором, переводчиком, лингвистом и писателем Максимом Ильяховым и начавший свое существование в 2014 г., ориентирован в первую очередь на стилистику и чистоту языка, но также отлично справляющийся с анализом сложности. Ключевые функции (в контексте сложности): поиск канцеляритов и «слов-паразитов», например, «наличие», «осуществлять», «кстати»; поиск сложных синтаксических конструкций (выделяет слишком длинные предложения, вложенные конструкции); оценка процента «водности» и заспамленности; подсветка проблемных мест в тексте с комментариями и проверка на соответствие информационному стилю. Данный текстовый анализатор подходит для научных статей, рекламы, инструкций и новостей, но не приемлем для стихов и художественной прозы.

На базе предшествующих методических разработок был разработан онлайн-инструмент «Текстометр» для оценки сложности тестов на русском языке [8]. Позволяет получить значение уровня сложности текста по шкале CEFR, вычислить индексы читабельности, коэффициент лексического разнообразия текста, вывести ключевые слова текста, а также рассчитать время его чтения.

Командой российских ученых (НИЛ «Интеллектуальные технологии управления текстами») Казанского федерального университета (КФУ) для автоматизации рутинных арифметических, лингвистических и исследовательских операций с текстом на русском языке был создан автоматический анализатор текстов RuLingva (rulingva.kpfu.ru). Одной из целей деятельности данной научной лаборатории – разработка автоматизированной модели для успешного чтения, предполагающей возможность выбора идеального текста для любого читателя [9].

Оценка и идентификация уровня сложности текстов на RuLingva основаны на выявленных корреляциях параметров текстов и критериев читателей (возраст, образование, словарный запас и т.д.).

Платформа RuLingva предлагает комплексный анализ текстовой сложности, оценивая параметры текста для разных категорий пользователей:

- носителей русского языка;
- изучающих русский как иностранный.

Для РКИ система анализирует 8 параметров: количество токенов, количество лемм, лексическое разнообразие (англ. Type-Token Ratio TTR), количество предложений, средняя длина предложений, количество слогов, средняя длина слов (в слогах), индекс Флеш-Кинкейда (FRGL).

Помимо данных параметров, RuLingva предоставляет следующую информацию по анализируемому тексту:

1. Распределение по уровням (количество токенов, количество лемм, кумулятивная доля токенов, кумулятивная доля лемм, уровни CEFR, доля слов разной частотности в тексте).
2. Список частотности.
3. Культурный код.

Как работать с RuLingva?

Несмотря на многочисленность функций, анализатор RuLingva имеет предельно простой интерфейс. На рис. 1 указано, что заданный текст необходимо ввести вручную или автоматически (сгенерировать свой текст) (см. рис. 1).

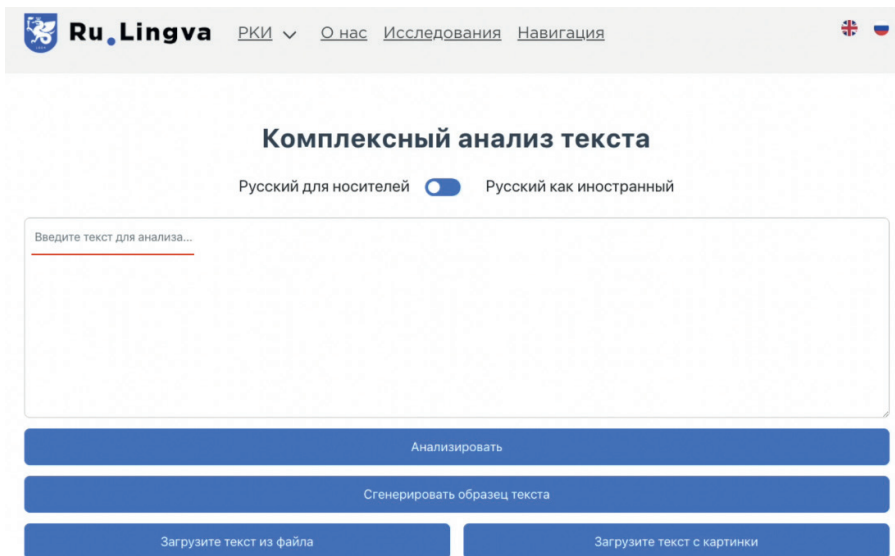


Рис. 1. Интерфейс RuLingva
Fig. 1. RuLingva interface

Далее RuLingva осуществляет автоматический лингвистический анализ текста объемом до 50 000 слов и оценивает метрики 47 параметров текстов на русском языке, «включая количество словоформ и слов, среднюю длину слов загруженного текста в слогах, среднюю длину предложений в словах, индексы лексического разнообразия и читабельности, связность, индекс абстрактности, количество терминов, ряд морфологических параметров и др. RuLingva позволяет выгружать и сохранять данные в формате excel-таблиц» [10].

Рассмотрим основные параметры, определяющие сложность тестового материала: индекс Флеша-Кинкейда (FRGL) и индекс лексического разнообразия (TTR) [10].

TTR – это классическая количественная мера, которая показывает богатство словарного запаса (лексикона) автора текста. Рассчитывается при помощи формулы: количество

уникальных лексических единиц / количество слов в тексте. Например, если в тексте из 60 слов имеется 30 уникальных лексических единиц (повторяющиеся слова считаются только один раз), то TTR будет равно 0,5 (30/60) [11]. Высокой информационной плотностью и сложностью обладает текст с высоким коэффициентом лексического разнообразия.

Автоматизация оценки TTR, несмотря на кажущуюся простоту, требует особого подхода. Главный недостаток TTR: зависимость от текстового объема. Сравнивать можно тексты только одной длины, т.к. с ростом числа токенов новые слова (типы) встречаются реже, вследствие чего лексическое разнообразие монотонно убывает. Короткий текст (например, состоящий из 100 слов) почти всегда будет иметь высокий TTR, чем длинный (например, состоящий из 5000 слов), даже если у последнего словарный запас объективно богаче [12]. Соответственно, интерпретировать TTR нужно с учетом жанра, цели и, что критически важно, длины анализируемого текста. Для глубокого анализа его используют с другими метриками [8].

Для оценки читабельности учебных текстов и художественной прозы на сайте RuLingva также установлена формула читабельности Флеша – Кинкейда, модифицированная И.В. Оборновой для русского языка. Данный параметр помогает определить, насколько текст легко читать и понимать. Рассчитывается по следующей формуле: $FRGL = 0,36 \times \text{средняя длина предложения} + 5,76 \times \text{среднее число слогов в слове} - 11,97$ [13]. Данная формула позволяет определить, для какого уровня образованности / класса подходит изучаемый текст. Значения индекса варьируются от 0 до 18, где 18 соответствует наиболее сложному тексту. Например, текст, имеющий читабельность $FRGL = 6,83$, ориентирован на школьников 7-8 классов.

Для текстов РКИ на данном сайте также рассчитать доли лексики (от A1 до C2) по международной градации CEFR (рис. 2).

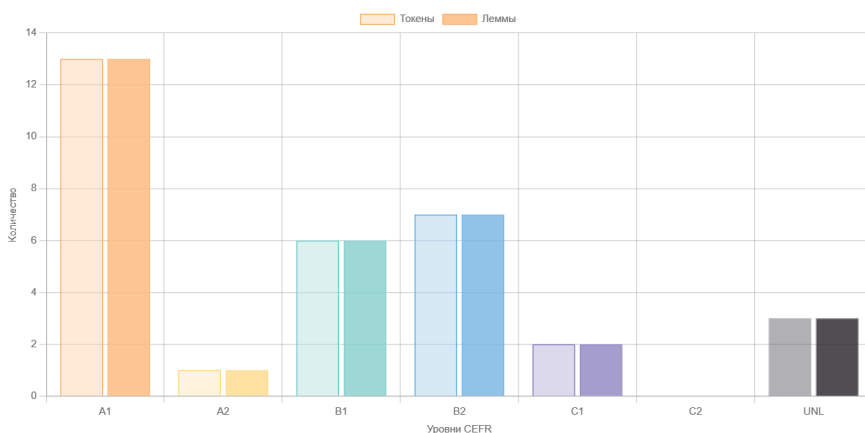


Рис. 2. Лексический анализ текста РКИ на Ru.Lingva
Fig. 2. Lexical analysis of the Russian language text on Ru.Lingva

Кроме того, лингвистический анализатор RuLingva позволяет определить культурный код текстового материала, что очень важно для иностранцев, изучающих русский язык. Текст РКИ, как правило, должен нести еще и лингвокультурологическую составляющую, поскольку основополагающим принципом, определяющим систему требований к содержанию процесса обучения, является культурологическая маркированность учебного материала направленность [14].

Программа выводит количество вхождений лексических единиц культурного кода, распределяет их по следующим тематическим группам: *География, Русский язык, Социальные слои, Традиции/быт, История* (рис. 3).

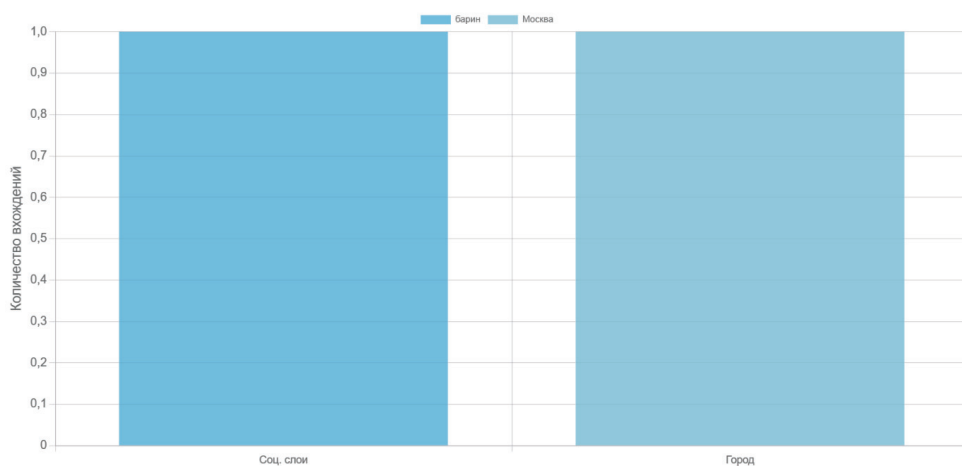


Рис. 3. Определение культурного кода на RuLingva
Fig. 3. Definition of cultural code on RuLingva

Данный лингвистический анализатор позволяет и оптимизировать учебные материалы для конкретных групп обучающихся, а также дает оценку прогресса в виде разработанных лексических тестов на оценку словарного запаса.

Разработанный тест на вербальный интеллект на базе платформы помогает не только оценить словарный запас пользователя, а также выявить области, требующие развития.

Сначала необходимо выбрать соответствующий уровень языка (см. рис. 4).

Приветствуем тебя!

Ниже будут предложены слова и изображения. Укажи правильный вариант

Выбери уровень:

A1

A2

B1

Рис. 4. Анализ словарного запаса
Fig. 4. Vocabulary analysis

Далее обучающиеся отвечают на 60 вопросов (см. рис. 5).

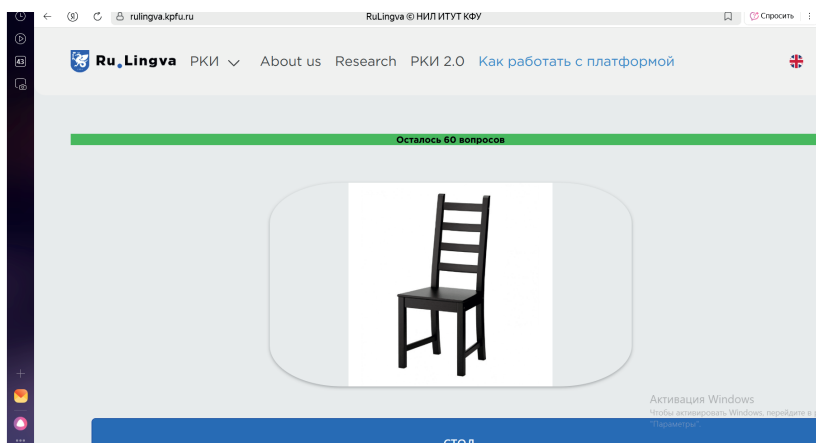


Рис. 5. Образец теста
Fig. 5. Sample test

В заключении прохождения теста система выводит результаты и статистику прохождения теста (см. рис 6).

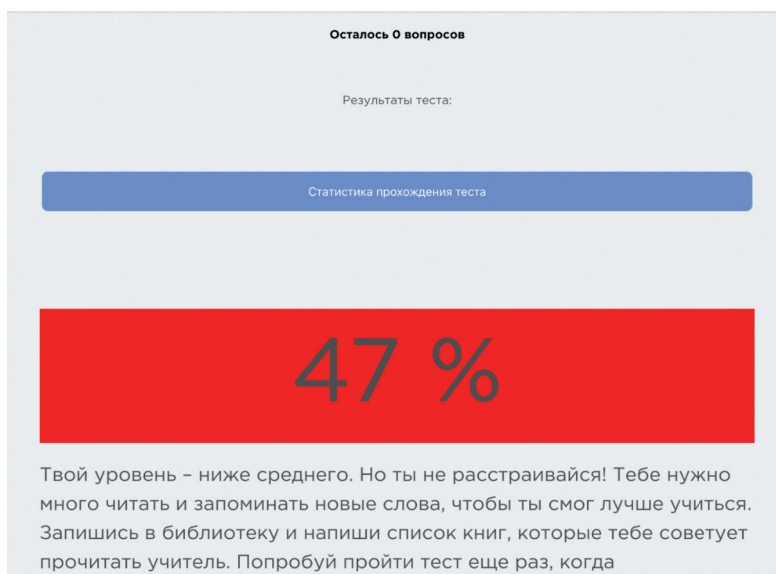


Рис. 6. Образец результата тестирования
Fig. 6. Result of Sample test

Таким образом, на базе программы RuLingva стало возможным подобрать текст соответствующего уровня сложности для различной целевой аудитории [13].

Обсуждение

С помощью профайлера RuLingva были проанализированы все отобранные нами русские народные песни. Русская песня является действенным педагогическим инструментом, позволяющим решать широкий спектр задач в обучении РКИ. Народные песни содержат богатую, естественную лексику, устойчивые выражения, пословицы и поговорки. Например, русская народная песня «Блины» вызывает ассоциации с традиционным рус-

ским блюдом, которое имеет большое значение в русской культуре. Блины символизируют праздник, достаток и семейные традиции.

Песня состоит из нескольких куплетов, каждый из которых описывает процесс приготовления блинов и радость, связанную с этим. Повторения фразы «Ой, блины, блины, блины» создают ритм и настроение праздника. Так же существует элемент визуализации, когда описываются действия старшей сестры, что позволяет представить процесс приготовления блинов «Моя старшая сестрица печь блины-то мастерица...». Использование уменьшительно-ласкательных форм («блиночки») создаёт тёплую и уютную атмосферу.

Песня наполнена радостью, ностальгией и теплотой. Чувства счастья и единства передаются через упоминания о блинах, которые связывают людей, вызывая воспоминания о домашнем уюте и заботе.

Текст песни очень легко читается: индекс удобочитаемости по Флешу-Кинкейду – 3.88, индекс лексического разнообразия достаточно низкий – 0.48, а значит в тексте, обнаруживается тенденция к повторению одних и тех же лексем, за счёт чего его TTR снижается: «Мы давно блинов не ели, / Мы блинчиков захотели. / Ой, блины, блины, блины, / Вы, блиночки мои! / В квашне новой растворили, / Два часа блины ходили. / Ой, блины, блины, блины, / Вы, блиночки мои!». Большую часть лексики составляют слова, относящиеся к уровню А1 (рис. 7). Количество лемм А1 – 24 (55.81%), А2 – 2 (4.65%), В1 – 3 (6.98%), В2 – 3 (6.98%), С1 – 2 (4.65%), С2.

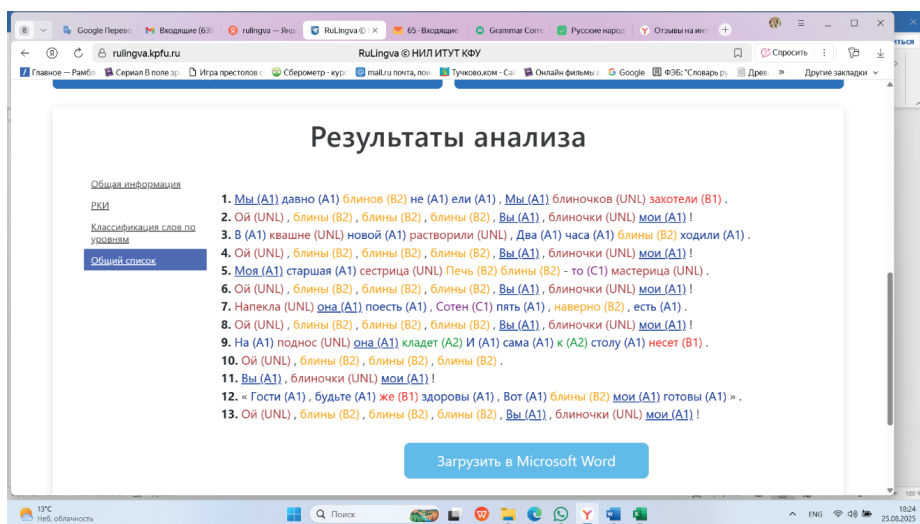


Рис. 7. Результаты анализа текста

Fig. 7. Results of text analysis

Выявлен культурный код: 1) объект традиционного быта – «печь» и 2) кухня – «блин».

Одна из самых известных русских народных песен «Живёт моя отрада». Это лирическая композиция, в которой рассказывается о настоящей и страстной любви, для которой не существует никаких преград и препятствий. Текст песни показывает более высокие параметры сложности. Индекс удобочитаемости 4.97, т.е. текст легко читается и воспринимается среднестатистическим обучающимся. Лексическое разнообразие – 0.67 указывает на богатый словарный состав текста «Я знаю, у красотки / Есть сторож у крыльца, / Но он не загородит / Дороги молодца. / Войду я к милой в терем / И брошусь в ноги к ней! / Была бы только ночь / Сегодня потемней».

Выявлены объекты культурного кода, относящиеся: к традиционному быту русского народа: «*терем*» и «*тройка*» – упряжка в три лошади.

Количество лемм А1 – 20 (50.00%), А2 – 3 (7.50%), В1 – 4 (10.00%), В2 – 2 (5.00%), С1 – 6 (15.00%), С2 – 0).

Популярная как плясовая песня «*Валенки*» описывает не только такой предмет русского быта, как валенки, но и рассказывает о чувствах, отношениях между молодыми людьми. Основная тема песни – любовь с искренними чувствами и простыми словами, без лишних эпитетов, обещаний, и клятв «*Суди, люди, суди, Бог, / как же я любила, / По морозу босиком/ К милому ходила*».

Индекс удобочитаемости 7.75, текст песни довольно простой и легкий для понимания, низкое значение TTR – 0.58, указывает на повторы лексических единиц в тексте песни: «*Валенки да валенки, / Ой, да не подишты стареньки, / Нельзя валенки носить, / Не в чем к милому сходить. / Валенки, валенки, / Эх, не подишты, стареньки, / Валенки да валенки, / Эх, не подишты стареньки*».

Количество лемм А1 – 17(45.95%), А2 – 2 (5.41%), В1 – 8 (21.62%), В2 2 (5.41%), С1 – 3 (8.11%), С2 – 1 (2.70%).

Объекты культурного кода не выявлены.

Были проанализированы все русские народные песни. Пример анализа представлен в таблице (см. табл. 2).

Таблица 2

Анализ русских народных песен

Table 2

Analysis of Russian folk songs

Название песни	Кол-во словоформ	Кол-во лемм	FRGL	TTR	Культурный код (кол-во вхождений)	Уровень
Посою лебеду на берегу	73	40	9.37	0,55	2 (4)	A1
Блины	91	43	3.88	0,48	2 (20,2)	A1
Валенки	63	37	7.75	0,59	нет	A1
Живет моя отрада	60	40	4.97	0.67	2 (5)	A1
Во поле береза стояла	55	24	9.61	0.44	нет	A1

Результаты

С помощью лингвистического анализатора RuLingva были рассчитаны индекс удобочитаемости FRGL, индекс лексического разнообразия TTR, уровень языка по общеевропейской компетенции владения иностранным языком и количество вхождений единиц культурного кода русских народных песен.

С точки зрения читабельности тексты русских народных песен не вызовут сложности понимания у обучающихся. Показатели индекса удобочитаемости Флеша-Кинкейда варьируются от 13.6 («Ах, ты Волга, Волга матушка») до 1.23 («Что мне жить и тужить»). Песни с наиболее высоким уровнем сложности: «Во кузнице» (12.85), «Я на горку шла» (12.69), «Ты воспой в саду соловейко» (11.29), «Ах! Пашинька, Парасковьюшка» (11.42), «Утенушка луговая» (10.39). Песни с наименьшим уровнем сложности: «Как хотела меня мать» (2.95), «Блины» (3.88), «Светит месяц, светит ясный» (3.84), «Уж ты, сад» (3.88).

Среднее число индекса FRGL – 7,1. Среднее число индекса лексического разнообразия составляет 0,56. В песнях «Во саду ли, в огороде» (0,81), «Ой, полна, полна коробушка» (0,87), «Уж ты пташечка» (0,83), «Хоть Москва в руках французов» (0,9), «Липа вековая» (0,87) относительно высокий коэффициент лексического разнообразия. Песни «Миленький ты мой» (0,15), «Ой, вставала я ранёшенько» (0,29), «Ты воспой в саду соловейко» (0,26), «По Дону гуляет» (0,26), «Во кузнице» (0,24), «Как хотела меня мать» (0,23) демонстрируют низкий уровень лексического разнообразия.

С помощью анализатора RuLingva выявлен культурный код песенного материала, подсчитано количество вхождений лексических единиц культурного кода, классифицированы по тематическим группам. Например, *География*: тайга (1), Байкал (3), Сибирь (1) («По диким степям Забайкалья»); *Русский язык*: товарищ (2) («Степь да степь кругом»), (3) («Раскинулось море широко»); *Социальные слои*: барышня (4) («Крутится, вертится шар голубой»), казак (2), казачка (2) («Посею лебеду на берегу»), ямщик (2) («Вот мчится тройка почтовая»); *Традиции/быт*: село (1) («Во саду ли, в огороде»), лапоть (4) («Лапти»), баня (5) («Топится, топится в огороде баня»); *История*: золотое кольцо (1) («У церкви стояла карета»), Кутузов (1) («Хоть Москва в руках французов»), Красная армия (1) («Как родная меня мать провожала»).

Все проанализированные песни относятся к уровню А1, что предполагает базовые знания русского языка и ориентированы на тех обучающихся, которые только начинают изучение русского языка.

Обсуждение результатов

Информация «об уровне языковой сложности текста является важнейшей, но не единственной характеристикой текста, влияющей на выбор текста преподавателем» [11]. Например, важным критерием является и информация о возможности использования данного текста для целей контроля, какая лексика может быть изучена на его материале и насколько она полезна для целевой аудитории. Поэтому, помимо уровня сложности текста, сервис «RuLingva» предлагает дополнительную информацию о тексте, представляющую ценность для его подготовки к занятию по РКИ: количество вхождений единиц культурного кода, статистику по покрытию текста лексическими минимумами ТРКИ (CEFR), частотный словарь, построенный на основе текста, а также лексический текст на вербальный интеллект, что дает возможность определить уровень владения языком.

Заключение

Анализ состояния текстовой аналитики показывает, что для решения стоящих перед ней задач современная лингвистика успешно обращается к междисциплинарным подходам [15]. Все большую актуальность приобретают «инструменты лингвистического профилирования, опирающиеся на лингвистическую статистику, компьютерную лингвистику и искусственный интеллект» [11]. RuLingva – удобный инструмент для анализа текста и определения основных параметров, по которым можно судить о сложности текста и о его соответствии уровню обучающихся.

Разрабатываемый инструмент позволяет измерить уровень владения русским языком, а также проводить «сопоставительные анализы в синхронии и диахронии, осуществлять кросс-культурные исследования» [11]. Очевидно, что любая автоматизация является долгим процессом и требует тщательного тестирования и доработки [14]. Ближайшая перспектива развития RuLingva связана с расширением спектра предикторов сложности и внедрением на платформе функции автоматического определения предметной области учебного текста.

Данная программа может быть полезна преподавателям и методистам, разработчикам контрольно-измерительных материалов (КИМ) и применима для учебного русскоязычного текста.

Литература

1. Вахрушева А. Я., Солнышкина М. И., Куприянов Р. В., Гафиятова Э. В., Климагина И. О. Лингвистическая сложность учебных текстов. *Вопросы журналистики, педагогики, языкознания*. 2021: 1: 89-99. URL: <https://cyberleninka.ru/article/n/lingvisticheskaya-slozhnost-uchebnyh-tekstov> (дата обращения: 21.01.2026).
2. Fisher D., Frey N., Lapp D. *Text complexity: raising rigor in reading*. Newark, International Reading Association, 2012: 212. (in English).
3. Pulido D. *Modeling the Role of Second Language Proficiency and Topic Familiarity in Second Language Incidental Vocabulary Acquisition through Reading*. 2004: 469- 523. (in English).
4. Арутюнов А.Р. *Теория и практика создания учебника русского языка для иностранцев*. М : Русский язык, 1990: 76.
5. Гриднева Н.А. Использование аутентичных материалов в обучении иностранному языку на уровне А1. *Самарский научный вестник*. 2017: 21(4): 210–214.
6. Лапошина А.Н., Лебедева М.Ю. Корпусный подход к решению проблемы отбора лексики в обучении РКИ. *Slavica Helsingiensia*. 2019: 52: 359–368.
7. Лапошина А. Н., Лебедева М. Ю. Текстомер: онлайн-инструмент определения уровня сложности текста по русскому языку как иностранному. *Русистика*. 2021:3:331-345. URL: <https://cyberleninka.ru/article/n/tekstometr-onlayn-instrument-opredeleniya-urovnya-slozhnosti-teksta-po-russkomu-yazyku-kak-inostrannomu> (дата обращения: 21.01.2026).
8. Лапошина А. Н. Опыт экспериментального исследования сложности текстов по РКИ. *Динамика языковых и культурных процессов в современной России: материалы VI Конгресса РОПРЯЛ* (Уфа, 11–14 октября 2018 г.): сборник статей. 2018: (6): 1544–1549.
9. Солнышкина М. И., Кисельников А. С. Сложность текста: этапы изучения в отечественном прикладном языкознании. *Вестник Том. гос. ун-та. Филология*. 2015: 38):6:86-99.
10. Чурунина А. А., Солнышкина М. И., Ярмакеев И. Э. Лексическое разнообразие как предикатор сложности учебников по русскому языку. *Russian Language Studies*: 2023: 21(2): 212-227.
11. Казачкова М. Б., Галимова Х. Н. Лексическое разнообразие как параметр сложности текста. *Вестник Марийского государственного университета*. 2021: 15(3): 384–390.
12. Солнышкина М. И., Соловьев В. Д., Эбзеева Ю. Н. Подходы и инструменты лингвистического профилирования текста на русском языке. *Русистика*. 2024: (4): 501-517.
13. Иванова Т.К. Объективация критериев сложности текста медиасферы на основе использования анализатора RuLingva. Медиалингвистика. Вып. 11. *Язык в координатах массмедиа: мат-лы VIII Междунар. научн. конференции* (Санкт-Петербург, 26–29 июня 2024 г.). Науч. ред. Л. Р. Дускаева, отв. ред. А. А. Малышев. СПб.: Медиапир, 2024: 750.
14. Милославская С. К. Учебник русского языка как иностранного -уникальное средство формирования образа России в мире: К теоретическому обоснованию лингвопедагогической имагологии. *Вестник РУДН. Вопросы образования: языки и специальность*. 2008: (4): 10-15.
15. Лапошина А. Н., Лебедева М. Ю. Текстомер: онлайн-инструмент определения уровня сложности текста по русскому языку как иностранному. *Русистика*. 2021: (3): 331-345.

References

1. Vakhrusheva A. Ya., Solnyshkina M. I., Kupriyanov R. V., et al. Linguistic Complexity of Educational Texts. *Issues of Journalism, Pedagogy, and Linguistics*. 2021:(1):89–99. Available at: <https://cyberleninka.ru/article/n/lingvisticheskaya-slozhnost-uchebnyh-tekstov> (accessed: 21 January 2026) (in Russian).
2. Fisher D., Frey N., Lapp D. *Text Complexity: Raising Rigor in Reading*. Newark, International Reading Association; 2012:212.
3. Pulido D. *Modeling the Role of Second Language Proficiency and Topic Familiarity in Second Language Incidental Vocabulary Acquisition through Reading*. 2004:469–523.

4. Arutyunov A.R. *Theory and Practice of Creating a Russian Language Textbook for Foreigners*. Moscow: Rus. lang. 1990: 76 (in Russian).
5. Gridneva N.A. Using Authentic Materials in Teaching a Foreign Language at Level A1. *Samara Scientific Bulletin*. 2017:21(4):210–214 (in Russian).
6. Laposhina A.N., Lebedeva M.Yu. Corpus-Based Approach to Solving the Problem of Vocabulary Selection in Teaching Russian as a Foreign Language. *Slavica Helsingiensia*. 2019: 2: 59–368 (in Russian).
7. Laposhina A. N., Lebedeva M. Yu. Textometer: An Online Tool for Determining the Difficulty Level of Texts in Russian as a Foreign Language. *Russists*. 2021:(3):331–345. Available at: <https://cyberleninka.ru/article/n/tekstometr-onlayn-instrument-opredeleniya-urovnya-slozhnosti-teksta-po-russkomu-yazyku-kak-inostrannomu> (accessed: 21 January 2026) (in Russian).
8. Laposhina A. N. An Experimental Study of the Complexity of Texts in Russian as a Foreign Language. *Dynamics of Linguistic and Cultural Processes in Modern Russia: Proceedings of the VI Congress of the Russian Association of Foreign Languages*, Ufa, October 11–14, 2018. Collection of Articles. 2018:(6):1544–1549 (in Russian).
9. Solnyshkina M.I., Kiselnikov A.S. Text Complexity: Stages of Study in Domestic Applied Linguistics. *Bulletin of Tomsk State University. Philology*. 2015:(38):6:86–99 (in Russian).
10. Churunina A.A., Solnyshkina M.I., Yarmakeev I.E. Lexical Diversity as a Predictor of Complexity of Russian Language Textbooks. *Russian Language Studies*. 2023:21(2):212–227 (in Russian).
11. Kazachkova M.B., Galimova Kh.N. Lexical Diversity as a Parameter of Text Complexity. *Bulletin of Mari State University*. 2021:15(3):384–390 (in Russian).
12. Solnyshkina M. I., Soloviev V. D., Ebzeeva Yu. N. Approaches and Tools for Linguistic Profiling of Russian-Language Texts. *Russists*. 2024:(4):501–517 (in Russian).
13. Ivanova T.K. Objectification of Media Sphere Text Complexity Criteria Using the RuLingva Analyzer. *Medialinguistics*. Issue 11. *Language in Mass Media Coordinates: Proceedings of the VIII International Scientific Conference* (St. Petersburg, June 26–29, 2024). Ed. L. R. Duskaeva, Ed. A. A. Malyshev. St. Petersburg: Mediapapir; 2024:750 (in Russian).
14. Miloslavskaya S. K. A textbook of Russian as a foreign language is a unique means of shaping the image of Russia in the world: Towards a theoretical justification of linguopedagogical imagology. *Bulletin of RUDN University. Education Issues: Languages and Specialties*. 2008:(4):10–15 (in Russian).
15. Laposhina A. N., Lebedeva M.Yu. Textometer: an online tool for determining the difficulty level of a text in Russian as a foreign language. *Russistics*. 2021:(3):331–345 (in Russian).

Сведения об авторах

КАЗАЧКОВА Мария Борисовна – к. филол. наук, доцент кафедры английского языка, Одинцовский филиал, Московский государственный институт международных отношений (университет). ORCID: 0000-0002-0357-3010. Scopus ID: 57216807031, Researcher ID: G-7529-2018.

E-mail: m.kazachkova@odin.mgimo.ru

ГАЛИМОВА Халида Нурисламовна – к. филол. наук, доцент кафедры иностранных языков и перевода, Казанский инновационный университет им. В. Г. Тимирязова (ИЭУП). ORCID: 0000-0003-1817-5004. SPIN-code: 7931-3389.

E-mail: galikha@mail.ru

Information about the authors

KAZACHKOVA Maria Borisovna – Cand. Sci. (Philology), Associate Professor, Department of the English Language, Odintsovo Branch, Moscow State Institute of International Relations (University), Odintsovo, Russian Federation, ORCID: 0000-0002-0357-3010, Scopus ID: 57216807031, Researcher ID: G-7529-2018, e-mail: m.kazachkova@odin.mgimo.ru.

GALIMOVA Khalida Nurislamovna – Cand. Sci. (Philology), Associate Professor, Department of Foreign Languages and Translation, V. G. Timiryasov Kazan Innovative University, Kazan, Russian Federation, ORCID: 0000-0003-1817-5004, SPIN-code: 7931-3389, e-mail: galikha@mail.ru

Вклад авторов

Все авторы сделали эквивалентный вклад в подготовку публикации.

Authors' contribution

Both the authors have made equal contribution to the preparation of the manuscript.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Conflict of interests

The authors declare no conflict of interest.

Поступила в редакцию / Submitted 12.02.26

Поступила после рецензирования / Revised 16.03.26

Принята к публикации / Accepted 19.03.26