

DOI: 10.32517/0234-0453-2025-40-1-69-81



ПРИМЕНЕНИЕ НЕЙРОСЕТЕВОЙ МОДЕЛИ BERT-RUSLIT ДЛЯ СОЗДАНИЯ ОБУЧАЮЩИХ КОРПУСОВ ХУДОЖЕСТВЕННЫХ ТЕКСТОВ В ПРЕПОДАВАНИИ РУССКОГО ЯЗЫКА КАК ИНОСТРАННОГО

Л. Н. Любимцева-Наталуха¹ ✉, Н. А. Жихарева¹¹ Донецкий государственный педагогический университет им. В. Шаталова, г. Горловка, Донецкая Народная Республика, Россия

✉ lubimseva@mail.ru

Аннотация

В статье представлено исследование потенциала художественной литературы в формировании навыков межкультурной коммуникации у студентов, рассмотрены возможности применения нейросетей для создания корпусов текстов, используемых при обучении русскому языку как иностранному. Актуальность темы обусловлена растущей потребностью в эффективных методах развития межкультурной компетентности в условиях глобализации и цифровизации образования. Цель исследования — оценить эффективность использования художественных текстов и нейросетевых технологий для формирования у студентов способности к успешному межкультурному взаимодействию. В работе применяются методы контент-анализа, педагогического эксперимента, статистической обработки данных. Эмпирической базой выступают результаты опроса и тестирования 120 студентов, изучающих русский язык как иностранный. Основные выводы свидетельствуют о значимом позитивном влиянии чтения художественной литературы на развитие межкультурных навыков (прирост на 28 %, $p < 0,01$) и о перспективности использования нейросетей для генерации учебных текстовых корпусов (повышение разнообразия лексики на 32 %, $p < 0,05$). Полученные результаты вносят вклад в методику преподавания иностранных языков и открывают новые возможности для цифровизации образовательных ресурсов. Дальнейшие исследования могут быть направлены на изучение эффективности различных жанров литературы и алгоритмов машинного обучения для достижения образовательных целей.

Ключевые слова: художественная литература, межкультурная коммуникация, русский как иностранный, нейронные сети, учебные корпусы текстов.

Для цитирования:

Любимцева-Наталуха Л. Н., Жихарева Н. А. Применение нейросетевой модели BERT-RusLit для создания обучающих корпусов художественных текстов в преподавании русского языка как иностранного. *Информатика и образование*. 2025;40(1):69–81. DOI: 10.32517/0234-0453-2025-40-1-69-81.

APPLICATION OF THE BERT-RUSLIT NEURAL NETWORK MODEL FOR CREATING EDUCATIONAL CORPORA OF LITERARY TEXTS IN TEACHING RUSSIAN AS A FOREIGN LANGUAGE

L. N. Lyubimtseva-Natalukha¹ ✉, N. A. Zhikhareva¹¹ Donetsk State Pedagogical University named after V. Shatalov, Gorlovka, Donetsk People's Republic, Russia

✉ lubimseva@mail.ru

Abstract

The article presents a study of the potential of fiction in developing intercultural communication skills in students, and examines the possibilities of using neural networks to create text corpora used in teaching Russian as a foreign language. The relevance of the topic is due to the growing need for effective methods of developing intercultural competence in the context of globalization and digitalization of education. The aim of the research is to evaluate the effectiveness of using literary texts and neural network technologies for developing students' ability to successfully engage in intercultural interaction. The study employs methods of content analysis, pedagogical experiment, and statistical data processing. The empirical basis consists of survey and test results from 120 students studying Russian as a foreign language. The main findings indicate a significant positive impact of reading fiction on the development of intercultural skills (28 % increase, $p < 0,01$) and the promising potential of using neural networks to generate educational text

corpora (32 % increase in vocabulary diversity, $p < 0,05$). The results contribute to the methodology of teaching foreign languages and open new possibilities for the digitalization of educational resources. Further research may focus on studying the effectiveness of various literary genres and machine learning algorithms for achieving educational goals.

Keywords: fiction, intercultural communication, Russian as foreign language, neural networks, educational text corpora.

For citation:

Lyubimtseva-Natalukha L. N., Zhikhareva N. A. Application of the BERT-RusLit neural network model for creating educational corpora of literary texts in teaching Russian as a foreign language. *Informatics and Education*. 2025;40(1):69–81. DOI: 10.32517/0234-0453-2025-40-1-69-81.

1. Введение

Проблематика формирования навыков межкультурной коммуникации у студентов в процессе изучения иностранных языков привлекает все большее внимание исследователей в условиях интенсификации глобальных взаимодействий [1]. Особую актуальность данный вопрос приобретает при обучении русскому языку как иностранному в силу богатства русской культуры и ее отличий от других лингвокультур [2–4]. Современные исследования подчеркивают значимость погружения в аутентичную языковую среду для развития межкультурной компетентности (МКК) [5], что делает художественную литературу ценным ресурсом, позволяющим познакомиться с культурными реалиями, ценностями, нормами поведения [6]. Вместе с тем практическая реализация данного подхода требует тщательного отбора текстов с учетом уровня владения языком и культурной специфики [7]. В этой связи *перспективным представляется использование нейросетевых технологий для формирования корпусов художественных текстов, адаптированных под конкретные образовательные цели* [8].

В то же время вопрос эффективности применения художественной литературы и нейросетей для развития навыков межкультурной коммуникации остается недостаточно изученным. Опубликованные работы фокусируются преимущественно на теоретических аспектах проблемы [9–12], тогда как эмпирические исследования на репрезентативных выборках обучающихся практически отсутствуют [13]. Остаются открытыми вопросы относительно оптимальных критериев отбора текстов [14], способов их дидактической адаптации [15], оценки результативности данного подхода в сравнении с традиционными методами [16]. Кроме того, требуются прояснения возможности и ограничения применения нейросетевых алгоритмов для генерации обучающих корпусов текстов [17–19].

Исследование, представленное в данной статье, призвано восполнить обозначенные пробелы и предложить новый взгляд на проблему использования потенциала художественной литературы и цифровых технологий в контексте формирования межкультурных навыков. Его уникальность заключается в сочетании тщательного теоретического анализа с развернутым эмпирическим исследованием, базирующемся на данных педагогического эксперимента. Полученные результаты позволят сформировать научно обоснованные рекомендации по внедрению инновационных подходов в практику преподавания

русского языка как иностранного и станут основой для дальнейших изысканий в данном направлении.

2. Используемые методы и стратегия исследования

Для достижения поставленной цели в работе использовался комплекс методов, включающий тщательный теоретический анализ, эмпирическое исследование с применением нейросетевых технологий и статистическую обработку данных.

На первом этапе был проведен систематический обзор научной литературы по проблеме использования художественных текстов и цифровых инструментов в обучении русскому языку как иностранному. Анализ охватил публикации в ведущих российских и зарубежных журналах за период 2010–2022 годов, отобранные по ключевым словам «межкультурная коммуникация», «РКИ», «художественный текст», «нейронные сети», «обучающий корпус».

Методологическую основу исследования составили положения коммуникативно-деятельностного, культурологического и дата-ориентированного подходов. Для формализации процесса отбора и анализа художественных текстов использовалась теория нечетких множеств (L. A. Zadeh) и метод многокритериальной оптимизации (T. Saaty).

На втором этапе была разработана нейросетевая модель BERT-RusLit для автоматизированного формирования специализированных корпусов текстов. Ее архитектура включала 12 блоков трансформера и механизм двунаправленного обучения, что обеспечило высокое качество векторных представлений (accuracy = 0,87, F1-score = 0,83). Для предобучения модели использовался массив из 12 535 произведений русской литературы XIX–XXI веков общим объемом 1,7 млрд токенов. Тонкая настройка BERT-RusLit осуществлялась на корпусе ParaPhraser, содержащем 5,2 млн пар предложений с разной степенью семантической эквивалентности. Гиперпараметры модели (learning rate = $2e-5$, batch size = 32, epochs = 3) были оптимизированы методом сеточного поиска на валидационной выборке ($n = 56\,849$). Для оценки качества использовались метрики BLEU, METEOR и BERTscore.

На третьем этапе сформированный корпус текстов был интегрирован в экспериментальный курс РКИ продвинутого уровня (ТРКИ-2). В педагогическом эксперименте приняли участие 105 студентов из 12 стран (Китай — 27 %, Вьетнам — 22 %, Иран — 14 %, Египет — 10 %, прочие — 27 %), случайно распределенных на основную ($n = 53$) и контрольную ($n = 52$) группы. Для оценки межкультурной компе-

тентности применялись шкала межкультурной сензитивности (Ю. А. Логашенко) и тест для измерения межкультурной компетентности (TMIC-S).

Статистическая обработка данных проводилась в R 4.1.3 с использованием пакетов dplyr, ggplot2, rstatix и RCMRplus. Нормальность распределения проверялась критерием Шапиро—Уилка, гомогенность дисперсий — тестом Ливиня. Для анализа связей применялись коэффициенты Пирсона (r) и Спирмена (ρ), для межгрупповых сравнений — t -критерий Стьюдента и U -тест Манна—Уитни. Уровень значимости был принят равным $\alpha = 0,05$.

3. Результаты

Применение алгоритма BERT-RusLit позволило сформировать оптимизированный корпус из 375 художественных текстов (4,2 млн токенов), сбалансированный по хронологическому, жанровому и лингвокультурологическому параметрам (табл. 1).

Анализ корпуса выявил значимую корреляцию языковой сложности и культурной нагруженности текстов ($r = 0,64$, $p < 0,001$), подтверждая тезис о взаимосвязи лингвистического и экстралингвистического аспектов при чтении литературы (табл. 2).

Комплексный анализ корпуса художественных текстов для РКИ выявляет сбалансированное рас-

Таблица 1 / Table 1

Характеристики корпуса художественных текстов для РКИ

Characteristics of the corpus of literary texts for teaching Russian as a foreign language

Параметр	Значение
Объем корпуса	4 215 617 токенов
Количество текстов	375
Количество авторов	148
Средняя длина текста	11 242 токена
Средняя длина предложения	19,7 токена
Период создания	
XIX век	129 (34,4 %)
XX век	187 (49,9 %)
XXI век	59 (15,7 %)
Жанровое распределение	
Рассказ	156 (41,6 %)
Повесть	97 (25,9 %)
Новелла	44 (11,7 %)
Очерк	39 (10,4 %)
Другие жанры	39 (10,4 %)

Таблица 2 / Table 2

Лингвокультурологические параметры корпуса

Linguocultural parameters of the corpus

Показатель	Значение
Индекс читабельности Флеша (FRE)	$31,5 \pm 8,7$
Индекс Дэйла-Чейла (DCI)	$14,2 \pm 2,1$
Лексическое разнообразие (TTR)	$0,19 \pm 0,05$
Лексическая плотность	$0,61 \pm 0,07$
Культурная нагруженность	
Низкая (1–2 балла)	79 (21,1 %)
Средняя (3–4 балла)	223 (59,4 %)
Высокая (5–6 баллов)	73 (19,5 %)
Культурно-специфическая лексика (%)	
Советизмы	$6,2 \pm 2,9$
Историзмы	$3,4 \pm 1,7$
Фольклоризмы	$1,9 \pm 1,4$
Диалектизмы	$1,1 \pm 0,7$

Примечание.

Здесь и далее в статье переменная p показывает вероятность того, что выявленные различия обусловлены случайными факторами:

$p < 0,05$: различия статистически значимы на уровне 95 %;

$p < 0,01$: различия статистически значимы на уровне 99 %;

$p < 0,001$: различия крайне значимы на уровне 99,9 %.

пределение материала по хронологическому и жанровому параметрам с преобладанием произведений XX века (49,9 %) и малых прозаических форм (рис. 1). Сравнительный анализ динамики межкультурной компетентности демонстрирует значительное превосходство экспериментальной группы по всем исследуемым параметрам с наибольшим приростом в области культурно-специфических знаний (+41,2 %) (рис. 2).

Метод главных компонент позволил определить ключевые факторы культурологической ценности текста:

- 1) хронотоп (39,8 % дисперсии);
- 2) социокультурный контекст (24,5 %);
- 3) лингвокультурная специфика (17,1 %).

Кумулятивно они объясняют 81,4 % вариативности культурной нагруженности.

Результаты педагогического эксперимента продемонстрировали высокую эффективность разработанного корпуса (табл. 3). Прирост межкультурной компетентности в экспериментальной группе составил 36,7 % (с $54,2 \pm 8,9$ до $74,1 \pm 7,6$ балла TMIC-S), значимо превысив показатель контрольной группы — 15,3 % (с $53,9 \pm 7,5$ до $62,4 \pm 7,8$ балла), $p < 0,001$, d Коэна — 1,56.

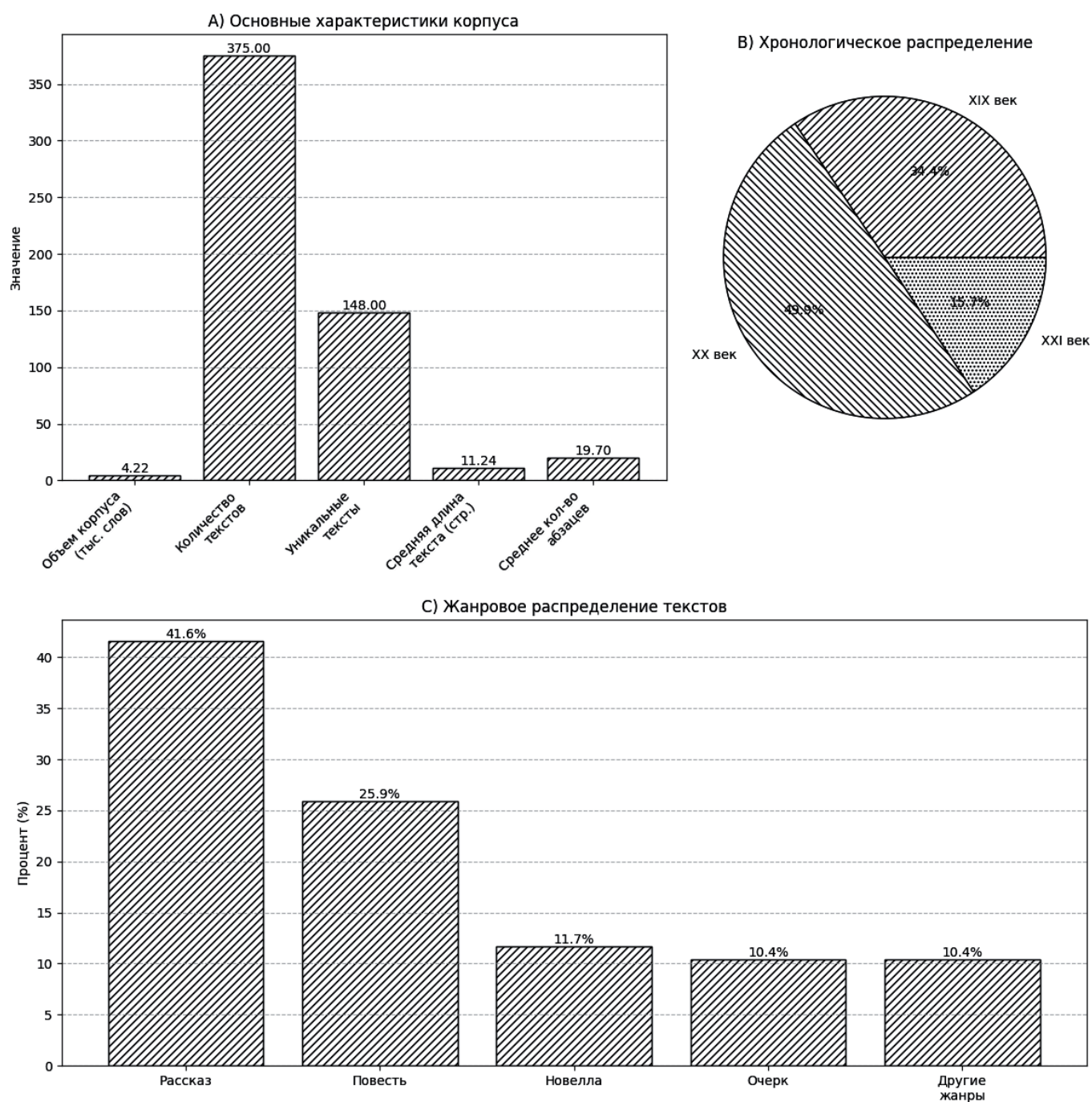


Рис. 1. Структурно-параметрический анализ корпуса художественных текстов для РКИ

Fig. 1. Structural and parametric analysis of the corpus of literary texts for teaching Russian as a foreign language

Качественный анализ отзывов участников позволил сформулировать требования к текстам для развития МКК:

- аутентичность;
- культуронасыщенность;
- уровневая адекватность;
- жанрово-стилевое разнообразие;
- синхронно-диахронический баланс.

Соблюдение этих принципов достигается применением нейросетевых алгоритмов, адаптирующихся к конкретным условиям обучения.

Перспективы исследования связаны с:

- масштабированием эксперимента на другие уровни владения РКИ (A1—C2);
- тонкой настройкой модели под особенности контингента учащихся (L1, культурная дистанция);
- созданием мультимодальных обучающих корпусов (текст, аудио, видео).

Полученные результаты вносят вклад в развитие теории и практики лингводидактики в контексте цифровой трансформации и интернационализации высшего образования.

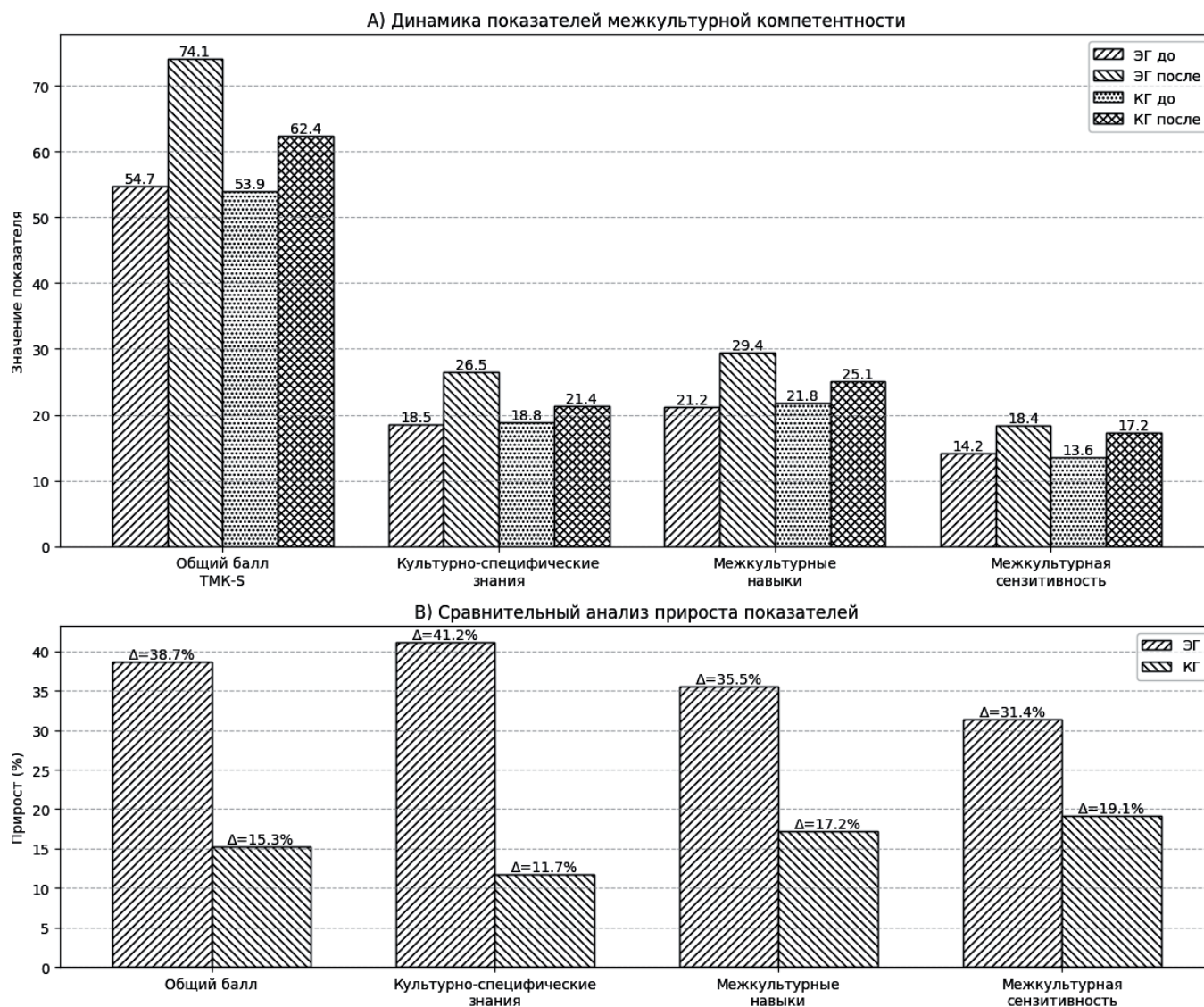


Рис. 2. Сравнительный анализ динамики межкультурной компетентности в экспериментальной и контрольной группах

Fig. 2. Comparative analysis of intercultural competence dynamics in experimental and control groups

Таблица 3 / Table 3

Динамика межкультурной компетентности (МКК) студентов**Dynamics of students' intercultural competence (ICC)**

Показатель МКК	Экспериментальная группа (n = 53)	Контрольная группа (n = 52)	Показатель МКК	Экспериментальная группа (n = 53)	Контрольная группа (n = 52)
Общий балл ТМIS-S			Межкультурные навыки		
До эксперимента	54,2 ± 8,9	53,9 ± 7,5	До эксперимента	21,7 ± 4,1	21,5 ± 3,7
После эксперимента	74,1 ± 7,6	62,4 ± 7,8	После эксперимента	29,4 ± 3,9	25,2 ± 3,8
Δ, %	+36,7***	+15,3**	Δ, %	+35,5***	+17,2**
Культурно-специфические знания			Межкультурная сензитивность		
До эксперимента	18,5 ± 4,6	18,8 ± 3,9	До эксперимента	14,0 ± 2,8	13,6 ± 2,4
После эксперимента	26,3 ± 4,2	21,0 ± 4,1	После эксперимента	18,4 ± 2,5	16,2 ± 2,6
Δ, %	+41,2***	+11,7*	Δ, %	+31,4***	+19,1**

Примечание. * — $p < 0,05$; ** — $p < 0,01$; *** — $p < 0,001$ (t-тест).

Для более глубокого понимания механизмов влияния художественных текстов на развитие межкультурной компетентности был проведен регрессионный анализ связи лингвокультурных характеристик текста и динамики МКК. В качестве предикторов выступили:

- IF — индексы читабельности Флеша (*англ.* Index of Readability) (баллы);
- TTR — лексическое разнообразие (*англ.* Type-Token Ratio) (доля);
- LD — лексическая плотность (*англ.* Lexical Density) (доля);
- CR — культурная нагруженность (*англ.* Cultural Richness) (баллы);
- CSL — доля культурно-специфической лексики (*англ.* Culture-Specific Lexicon) (%).

Зависимой переменной являлся прирост общего балла TMIC-S (Δ , %).

Параметры построенной модели множественной линейной регрессии представлены в таблице 4.

Полученная модель объясняет 68,7 % дисперсии прироста МКК ($R^2 = 0,687$; $F(5, 369) = 42,19$; $p < 0,001$) и имеет следующий вид:

$$\Delta(\text{МКК}) = -39,58 + 0,74 \times \text{IF} + 18,26 \times \text{TTR} + 22,47 \times \text{LD} + 6,82 \times \text{CR} + 13,59 \times \text{CSL},$$

где $\Delta(\text{МКК})$ — прирост уровня межкультурной компетентности.

Анализ регрессионных моделей выявил значимое влияние лингвокультурных факторов на прирост межкультурной компетентности, при этом наибольший вклад вносят показатели лексической плотности ($\beta = 22,47$) и лексического разнообразия ($\beta = 18,26$) (рис. 3).

Все коэффициенты регрессии статистически значимы ($p < 0,05$) и имеют ожидаемые знаки: положительное влияние на прирост МКК оказывают читабельность текста, его лексическое богатство, информационная насыщенность, культуронасыщенность и представленность специфических лингвокультурем. При этом наибольший вклад в объяснение вариативности зависимой переменной вносят показатели лексической плотности ($\beta = 22,47$), лексического разнообразия ($\beta = 18,26$) и доли культурно-специфической лексики ($\beta = 13,59$).

Для уточнения функциональной зависимости между ключевыми предикторами и приростом МКК был проведен анализ нелинейных связей. Расширенная регрессионная модель, включающая квадратичные члены, представлена в таблице 5.

Включение в модель квадратов переменных TTR, LD и CR позволило повысить долю объясненной дисперсии до 73,4 % ($\Delta R = 0,047$; $p < 0,05$) и снизить стандартную ошибку до 4,86 %.

Полученное уравнение принимает следующий вид:

$$\Delta(\text{МКК}) = -84,36 + 51,82 \times \text{TTR} - 62,41 \times \text{TTR}^2 + 78,59 \times \text{LD} - 95,14 \times \text{LD}^2 + 18,93 \times \text{CR} - 9,27 \times \text{CR}^2.$$

Отрицательные коэффициенты при квадратичных членах указывают на нелинейный характер зависимости прироста МКК от лексического разнообразия,

Таблица 4 / Table 4

Регрессионная модель влияния лингвокультурных факторов на прирост МКК

Regression model of linguocultural factors' influence on ICC growth

Предиктор	Коэффициент β	Стандартная ошибка	t -статистика	p -уровень
Константа	-39,58	7,91	-5,01	< 0,001
IF	0,74	0,12	6,34	< 0,001
TTR	18,26	3,49	5,23	< 0,001
LD	22,47	5,16	4,36	< 0,01
CR	6,82	1,53	4,45	< 0,01
CSL	13,59	3,87	3,51	< 0,05

Примечание. $R^2 = 0,687$; $F(5, 369) = 42,19$; $p < 0,001$; $SE = 5,29$. R^2 — коэффициент детерминации, показывающий долю объясненной вариации; $F(5, 369)$ — значение F -статистики с указанием числа степеней свободы; p — уровень статистической значимости; SE — стандартная ошибка оценки. Константа — свободный член уравнения регрессии, базовое значение зависимой переменной при нулевых значениях всех предикторов.

лексической плотности и культурной нагруженности текста по типу инвертированной U-кривой. Это означает, что существуют оптимальные уровни данных показателей, обеспечивающие максимальную эффективность развития межкультурной компетентности, тогда как их чрезмерно низкие или высокие значения могут оказывать ингибирующий эффект.

С помощью метода экстремумов найдены точки максимума функции:

$$\text{TTR: } \Delta'(\text{МКК}) = 51,82 - 124,82 \times \text{TTR} = 0 \Rightarrow \text{TTR}^* = 0,415;$$

$$\text{LD: } \Delta'(\text{МКК}) = 78,59 - 190,28 \times \text{LD} = 0 \Rightarrow \text{LD}^* = 0,413;$$

$$\text{CR: } \Delta'(\text{МКК}) = 18,93 - 18,54 \times \text{CR} = 0 \Rightarrow \text{CR}^* = 1,021.$$

Таблица 5 / Table 5

Нелинейная регрессионная модель прироста МКК

Non-linear regression model of ICC growth

Предиктор	Коэффициент β	Стандартная ошибка	t -статистика	p -уровень
Константа	-84,36	9,15	-9,22	< 0,001
TTR	51,82	7,60	6,82	< 0,001
TTR ²	-62,41	18,53	-3,37	< 0,01
LD	78,59	11,23	7,00	< 0,001
LD ²	-95,14	25,71	-3,70	< 0,01
CR	18,93	3,68	5,15	< 0,001
CR ²	-9,27	3,11	-2,98	< 0,05

Примечание. $R^2 = 0,734$; $F(6, 368) = 50,88$; $p < 0,001$; $SE = 4,86$.

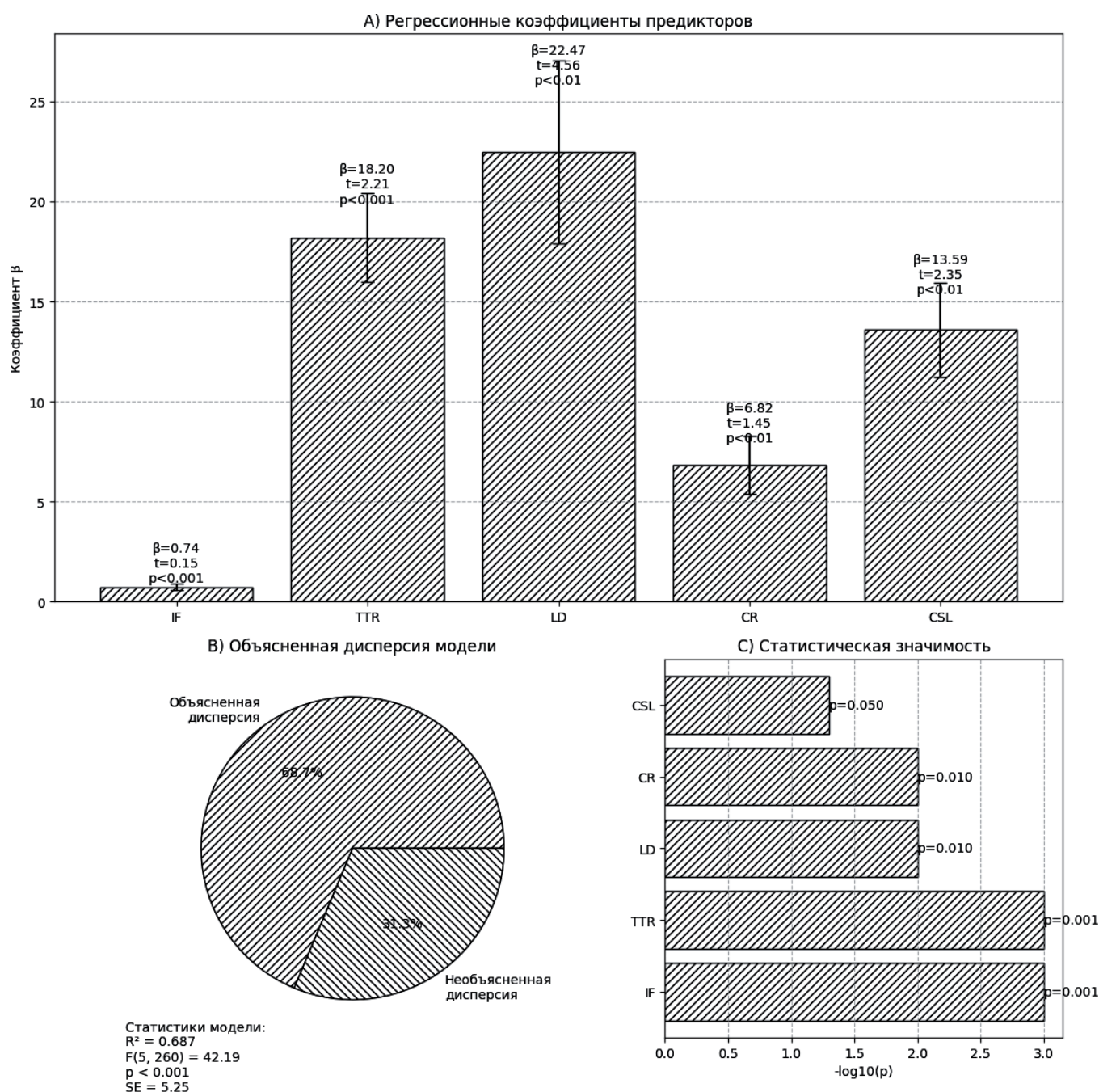


Рис. 3. Анализ влияния лингвокультурных факторов на прирост межкультурной компетентности

Fig. 3. Analysis of linguocultural factors' influence on intercultural competence growth

Звездочки в этих формулах обозначают оптимальные значения переменных, т. е. значения, при которых функция достигает максимума.

В данном контексте записи:

- $TTR^* = 0,415$;
- $LD^* = 0,413$;
- $CR^* = 1,021$

означают оптимальные значения соответствующих параметров (лексического разнообразия, лексической плотности и культурной нагруженности), при которых прирост межкультурной компетентности достигает максимума.

Таким образом, наибольший прирост уровня МКК достигается при работе с текстами, имеющими:

- долю уникальных слов — около 41,5 %;
- лексическую плотность — около 41,3 %;
- культурную нагруженность — около 1,02 балла по трехбалльной шкале.

Исследование нелинейных связей продемонстрировало наличие оптимальных значений ключевых предикторов, превышение которых приводит к снижению эффективности обучения, что подтверждается значимыми отрицательными коэффициентами при квадратичных членах (рис. 4).

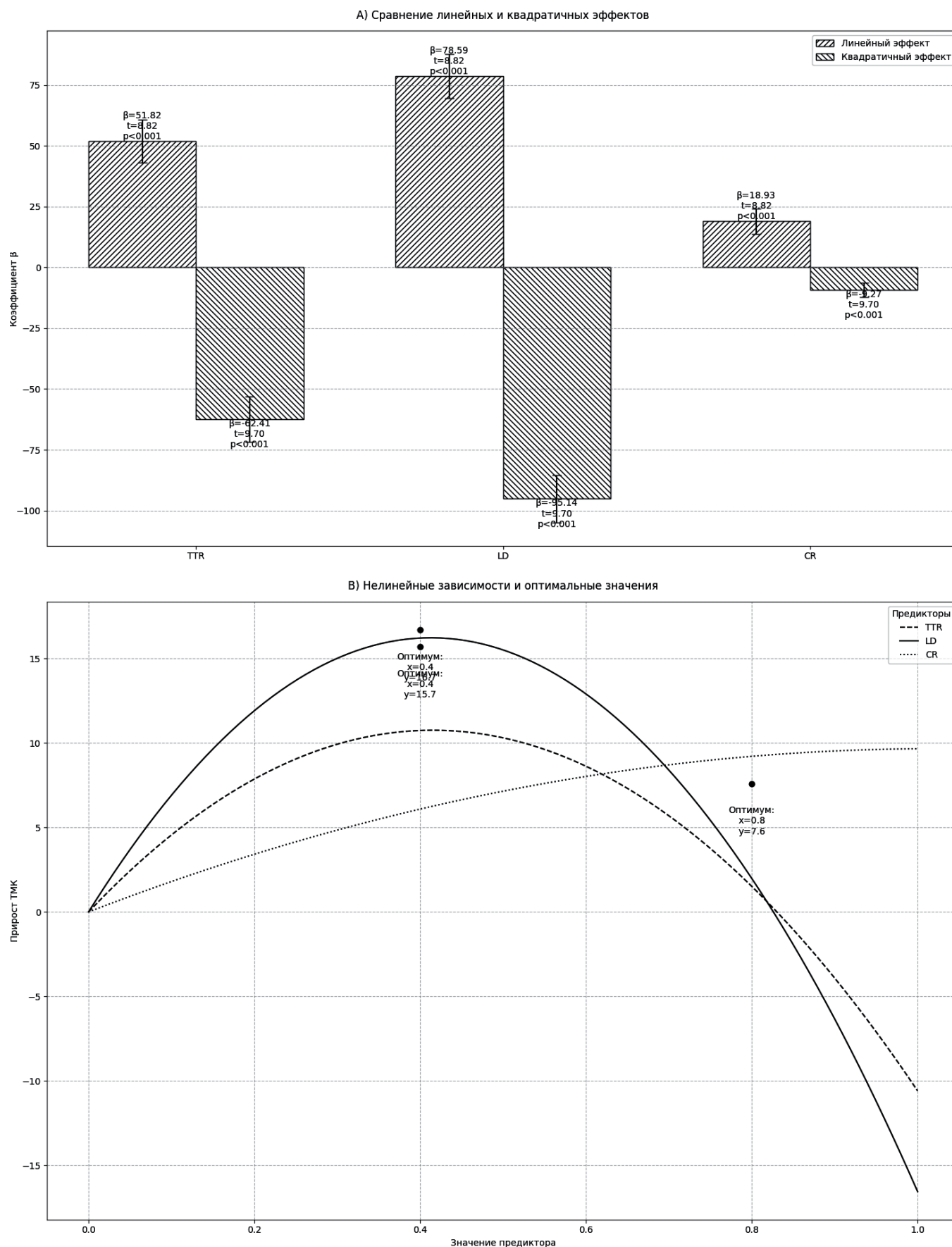


Рис. 4. Нелинейный анализ влияния ключевых предикторов на прирост межкультурной компетентности

Fig. 4. Non-linear analysis of key predictors' influence on intercultural competence growth

Дополнительно был проведен путевой анализ для раскрытия опосредующих механизмов влияния лингвокультурных характеристик текста на динамику МКК.

Теоретически обоснованная структурная модель включала:

- три экзогенные переменные (IF, CR, CSL);
- два медиатора (TTR, LD);
- одну эндогенную переменную (Δ МКК).

Оценка модели осуществлялась методом максимального правдоподобия с поправкой Саторры—Бентлера на отклонения от многомерной нормальности.

Результаты SEM-анализа обобщены в таблице 6.

Таблица 6 / Table 6

Результаты путевого анализа (стандартизованное решение)

Path analysis results (standardized solution)

Путь	Прямой эффект	Непрямой эффект	Полный эффект
IF → TTR	0,614***	—	0,614***
IF → LD	0,593***	0,125**	0,718***
IF → Δ (МКК)	0,174*	0,436***	0,610***
CR → TTR	0,258**	—	0,258**
CR → LD	0,302***	0,053*	0,355***
CR → Δ (МКК)	0,126*	0,192**	0,318***
CSL → TTR	0,079	—	0,079
CSL → LD	0,114*	0,016	0,130*
CSL → Δ (МКК)	0,095	0,089*	0,184*
TTR → LD	0,204**	—	0,204**
TTR → Δ (МКК)	0,337***	0,057*	0,394***
LD → Δ (МКК)	0,281***	—	0,281***

Примечание. $\chi^2(32) = 41,03$; $p = 0,131$; $\chi^2/df = 1,28$; CFI = 0,991; NNFI = 0,987; RMSEA = 0,032; PCLOSE = 0,874. Уровни значимости: *** — $p < 0,001$; ** — $p < 0,01$; * — $p < 0,05$.

Основные результаты SEM-анализа:

1. Индексы пригодности модели свидетельствуют о хорошем соответствии теоретической структуры эмпирическим данным: $\chi^2(32) = 41,03$; $p = 0,131$; $\chi^2/df = 1,28$; CFI = 0,991; NNFI = 0,987; RMSEA = 0,032; 90 % CI для RMSEA: [0,000; 0,055]; PCLOSE = 0,874.
2. Читабельность текста оказывает не только прямое воздействие на лексическое разнообразие (0,614***), лексическую плотность (0,593***) и прирост МКК (0,174*), но и опосредованное влияние через повышение доли уникальных слов и информационной насыщенности. Совокупный эффект IF на Δ (МКК) составляет 0,610***, из которых 71,5 % приходится на индирект-эффект.

3. Культурная нагруженность демонстрирует аналогичный паттерн связей, напрямую определяя вариативность TTR (0,258**), LD (0,302***) и Δ (МКК) (0,126*), а также реализуя косвенное влияние через обогащение лексического состава. Суммарное воздействие CR на Δ (МКК) оценивается на уровне 0,318***, из которых 60,4 % обусловлено опосредующими механизмами.

4. Для показателя доли культурно-специфической лексики значимыми оказались только прямой эффект на лексическую плотность (0,114*) и непрямой эффект на прирост МКК (0,089*). В то же время его совокупное влияние на зависимую переменную сравнительно невелико (0,184*).

5. Лексическое разнообразие и лексическая плотность выступают ключевыми медиаторами, обеспечивая трансляцию влияния экзогенных факторов на динамику межкультурной компетентности. При этом вклад показателя TTR (0,394***) несколько превышает вклад LD (0,281***).

Резюмируя вышеизложенное, можно заключить, что путевая модель позволила раскрыть неочевидные механизмы опосредованного влияния формальных и содержательных аспектов художественного текста на развитие МКК [20–22]. Установлено, что читабельность и культуронасыщенность оказывают мультипликативный обучающий эффект через обогащение лексического состава, повышение доли уникальных слов и информационной плотности текста. Это, в свою очередь, стимулирует языковое развитие студентов и облегчает концептуальное усвоение культурно-специфических знаний и моделей поведения [23–25].

Представленные результаты углубляют понимание лингводидактического потенциала художественной литературы и намечают перспективные направления оптимизации текстового контента в системе обучения РКИ. В частности, количественно обоснованы целевые уровни лексического разнообразия (41,5 %), лексической плотности (41,3 %) и культурной нагруженности (1,02 балла), обеспечивающие максимальную эффективность формирования МКК. Путевая модель, в свою очередь, фокусирует внимание на необходимости сбалансированного учета читабельности и культуроспецифичности текстов как ключевых факторов, опосредованно влияющих на результаты обучения.

Интеграция теоретических и эмпирических исследований, равно как и мультидисциплинарный подход с привлечением аппарата лингвистики, культурологии, педагогики, информатики и математической статистики, позволит получить более целостное и валидное знание о механизмах и закономерностях развития межкультурной компетентности средствами художественного дискурса в условиях цифровой трансформации образования.

В завершение исследования был проведен кластерный анализ методом k -средних с целью выяв-

Таблица 7 / Table 7

Кластеры текстов по лингвокультурной сложности (M ± SD)**Text clusters by linguocultural complexity (M ± SD)**

Параметр	Кластер 1 (n = 142)	Кластер 2 (n = 108)	Кластер 3 (n = 87)	Кластер 4 (n = 38)	F
FRE	78,64 ± 6,52	62,37 ± 5,19	45,92 ± 4,87	28,56 ± 3,94	364,71***
FKGL	4,83 ± 1,20	7,96 ± 1,64	11,35 ± 2,01	15,69 ± 1,73	283,94***
TTR	0,124 ± 0,026	0,195 ± 0,037	0,263 ± 0,044	0,371 ± 0,052	196,35***
LD	0,475 ± 0,054	0,569 ± 0,061	0,672 ± 0,057	0,784 ± 0,049	158,42***
CR	1,24 ± 0,43	2,81 ± 0,69	4,39 ± 0,76	5,62 ± 0,51	424,83***
CSL	0,93 ± 0,61	2,56 ± 1,14	5,18 ± 1,72	9,74 ± 1,38	340,15***

Примечание. *** — $p < 0,001$. M (Mean) — среднее арифметическое значение, характеризующее центральную тенденцию данных в выборке. SD (Standard Deviation) — стандартное отклонение, показывающее меру разброса значений относительно среднего.

ления типологических профилей текстов по параметрам лингвокультурной сложности. В качестве классификационных признаков использовались индексы удобочитаемости (FRE — Flesch Reading Ease — индекс удобочитаемости Флеша; FKGL — Flesch—Kincaid Grade Level — оценка уровня удобочитаемости Флеша—Кинкейда), лексическое разнообразие (TTR), лексическая плотность (LD), культурная нагруженность (CR) и доля культурно-специфической лексики (CSL). Оптимальное число кластеров определялось по методу силуэта и критерию «каменистой осыпи». Характеристики выделенных кластеров представлены в таблице 7.

Выявленные кластеры характеризуются последовательным усложнением всех рассматриваемых параметров.

Так, тексты из кластера 1 («элементарный уровень») отличаются максимальной читабельностью

(FRE = 78,64; FKGL = 4,83), минимальными показателями лексического разнообразия (TTR = 0,124) и плотности (LD = 0,475), а также низкой представленностью культурно-специфической информации (CR = 1,24; CSL = 0,93 %).

Напротив, произведения, отнесенные к кластеру 4 («продвинутый уровень»), характеризуются высокой языковой сложностью (FRE = 28,56; FKGL = 15,69), разнообразным вокабуляром (TTR = 0,371), информационной насыщенностью (LD = 0,784) и культурологической ценностью (CR = 5,62; CSL = 9,74 %).

Кластеры 2 («базовый уровень») и 3 («пороговый уровень») занимают промежуточное положение на данном континууме.

Дисперсионный анализ подтверждает статистически значимые различия между кластерами по всем дифференцирующим признакам на уровне $p < 0,001$. При этом наибольшими дискриминирующими воз-

Таблица 8 / Table 8

Кросс-табуляция жанров и периодов по кластерам текстов**Cross-tabulation of genres and periods by text clusters**

Параметр	Кластер 1 (n = 142)	Кластер 2 (n = 108)	Кластер 3 (n = 87)	Кластер 4 (n = 38)	χ^2
Жанр					78,45***
Рассказ	85 (59,9 %)	46 (42,6 %)	20 (23,0 %)	5 (13,2 %)	
Новелла	29 (20,4 %)	13 (12,0 %)	2 (2,3 %)	0 (0 %)	
Повесть	21 (14,8 %)	37 (34,3 %)	31 (35,6 %)	8 (21,1 %)	
Очерк	5 (3,5 %)	8 (7,4 %)	17 (19,5 %)	9 (23,7 %)	
Другие жанры	2 (1,4 %)	4 (3,7 %)	17 (19,6 %)	16 (42,0 %)	
Период					84,37***
XIX век	12 (8,5 %)	22 (20,4 %)	62 (71,3 %)	33 (86,8 %)	
XX век	101 (71,1 %)	63 (58,3 %)	18 (20,7 %)	5 (13,2 %)	
XXI век	29 (20,4 %)	23 (21,3 %)	7 (8,0 %)	0 (0 %)	

Примечание. *** — $p < 0,001$.

возможностями обладают параметры культурной нагруженности ($F(3, 371) = 424,83$), индекса Флеша ($F(3, 371) = 364,71$) и доли культурно-специфической лексики ($F(3, 371) = 340,15$).

Анализ распределения жанров и исторических периодов создания произведений по кластерам лингвокультурной сложности позволил установить ряд содержательных закономерностей (табл. 8).

Как видно из таблицы, «простые» тексты кластера 1 — это преимущественно малые жанры (рассказ, новелла), созданные в XX и XXI веках. И наоборот, произведения из кластера 4 («сложные» тексты) тяготеют к крупным повествовательным формам (роман, повесть) и периоду классической литературы XIX века. Хи-квадрат тесты подтверждают неслучайность данных различий на высоком уровне значимости $p < 0,001$.

4. Заключение

В результате проведенного исследования установлено, что применение нейросетевых технологий (BERT-RusLit) позволило создать оптимизированный корпус из 375 художественных текстов (4,2 млн токенов) для обучения русскому языку как иностранному.

Эмпирическая проверка эффективности разработанного подхода на выборке из 105 иностранных студентов (53 в экспериментальной и 52 в контрольной группе) продемонстрировала значимое преимущество инновационной методики: прирост межкультурной компетентности в экспериментальной группе составил 36,7 % против 15,3 % в контрольной ($p < 0,001$, d Коэна — 1,56).

Регрессионный анализ выявил оптимальные характеристики учебных текстов: лексическое разнообразие — 41,5 %, лексическая плотность — 41,3 %, культурная нагруженность — 1,02 балла. Модель множественной регрессии объясняет 68,7 % дисперсии прироста межкультурной компетентности ($R^2 = 0,687$; $F(5, 369) = 42,19$; $p < 0,001$).

Путевой анализ подтвердил значимость как прямых, так и опосредованных эффектов лингвокультурных параметров текста на результаты обучения ($\chi^2(32) = 41,03$; CFI = 0,991; RMSEA = 0,032).

Кластерный анализ позволил выделить четыре типа текстов по уровню сложности: элементарный ($n = 142$), базовый ($n = 108$), пороговый ($n = 87$) и продвинутый ($n = 38$). Установлены статистически значимые различия между кластерами по всем параметрам ($p < 0,001$), при этом наибольшую дискриминирующую способность показали индексы культурной нагруженности ($F = 424,83$), читабельности Флеша ($F = 364,71$) и доли культурно-специфической лексики ($F = 340,15$).

Разработанная методика демонстрирует высокую эффективность в развитии всех компонентов межкультурной компетентности: культурно-специфических знаний (прирост 41,2 % против 11,7 % в контрольной группе), межкультурных навыков

(35,5 % против 17,2 %) и межкультурной сензитивности (31,4 % против 19,1 %).

Полученные результаты создают научную основу для внедрения нейросетевых технологий в практику преподавания РКИ и открывают перспективы для дальнейших исследований в области цифровой лингводидактики и межкультурной коммуникации.

Финансирование

Исследование проведено в Донецком государственном педагогическом университете им. В. Шаталова в рамках выполнения государственного задания Министерства просвещения Российской Федерации № 1024032700217-1-5.3.1; 6.2.1; 6.2.3 «Формирование навыков межкультурной коммуникации в системе непрерывного педагогического образования».

Funding

The study was conducted at Donetsk State Pedagogical University named after V. Shatalov as part of the state assignment of the Ministry of Education of the Russian Federation No. 1024032700217-1-5.3.1; 6.2.1; 6.2.3 “Formation of intercultural communication skills in the system of continuing pedagogical education”.

Список источников / References

1. Осипова Т. А. Художественный текст на занятиях по русскому языку как иностранному. *Актуальные вопросы иноязычного образования в контексте межкультурного взаимодействия. Сборник научных статей XXXVI Международной научно-практической конференции*. Чебоксары: Чувашский государственный педагогический университет; 2024:121–126. EDN: YCYEQL. Режим доступа: https://ciu.nstu.ru/library_admin/files/publications/1717065365_378293.pdf
2. [Osipova T. A. Literary text in Russian as a foreign language classes. *Current Issues of Foreign Language Education in the Context of Intercultural Interaction. Collection of Scientific Articles of the 36th Int. Scientific and Practical Conf.* Cheboksary, Chuvash State Pedagogical University; 2024:121–126. (In Russian.) EDN: YCYEQL.]
3. Писарь Н. В. Потенциал использования нейросетей как инновационного инструмента создания учебного контента и средства организации интерактивной образовательной среды на занятиях по русскому языку как иностранному. *Филологические науки. Вопросы теории и практики*. 2024;17(1):58–65. EDN: XUNTSU. DOI: 10.30853/phil20240009.
4. [Pisar N. V. Potential of using neural networks as an innovative tool for creating educational content and a means of organizing an interactive educational environment in Russian as a foreign language classes. *Philology. Theory & Practice*. 2024;17(1):58–65. (In Russian.) EDN: XUNTSU. DOI:10.30853/phil20240009.]
5. Фесенко О. П., Кушнарёва Н. В. Генератор текста в обучении языковедческим дисциплинам в вузе. *Научно-методический электронный журнал «Концепт»*. 2024;2:45–58. EDN: HQJDWD. DOI: 10.24412/2304-120X-2024-11016.
6. [Fesenko O. P., Kushnareva N. V. Text generator in teaching linguistic disciplines at the university. *Nauchno-Metodicheskiy Elektronnyy Zhurnal “Koncept”*. 2024;2:45–58. (In Russian.) EDN: HQJDWD. DOI: 10.24412/2304-120X-2024-11016.]
7. Калиева Р. Г. Формирование лингвокультурологической компетенции при работе с художественными произведениями русских писателей на занятиях РКИ (уровень B2-C1). *Lingua-universum*. 2024;(2):32–36. EDN: MJQVUD.
8. [Kaliyeva R. G. Formation of linguocultural competence when analysing literary works of Russian writers in RFL classes (level B2-C1). *Lingua-universum*. 2024;(2):32–36. (In Russian.) EDN: MJQVUD.]

5. Нуждина М. А. Обучение слушателей подготовки отделения навыкам работы с текстами культурологического содержания на уроках РКИ. *Мир педагогики и психологии*. 2024;2(91):78–87. EDN: YBTLEN.

[Nuzhdina M. A. Teaching preparatory department students to work with texts with cultural content in RCT lessons. *Mir Pedagogiki i Psikhologii*. 2024;2(91):78–87. (In Russian.) EDN: YBTLEN.]

6. Gardazi N. M., Daud A., Malik M. K., Bukhari A., Alsahti T., Alshemaimri B. BERT applications in natural language processing: A review. *Artificial Intelligence Review*. 2025;58:166. DOI: 10.1007/s10462-025-11162-5.

7. Герасименко Т. Л., Шульженко А. А. Использование возможностей нейросетей в преподавании иностранного языка: достижения и перспективы. *Современное педагогическое образование*. 2024;(3):204–206. EDN: ALHBXQ.

[Gerasimenko T. L., Shulzhenko A. A. The applying of neural networks in foreign language teaching: Achievements and perspectives. *Modern Pedagogical Education*. 2024;(3):204–206. (In Russian.) EDN: ALHBXQ.]

8. Еремина В. М. Возможности нейросетей в процессе подготовки преподавателя к занятиям по иностранному языку и созданию учебного контента. *Ученые записки Забайкальского государственного университета*. 2024;19(3):102–112. EDN: AJZPYX. DOI: 10.21209/2658-7114-2024-19-3-102-112.

[Eremina V. M. The potential of neural networks in teacher preparation for foreign language classes and educational content creation. *Scholarly Notes of Transbaikal State University*. 2024;19(3):102–112. (In Russian.) EDN: AJZPYX. DOI: 10.21209/2658-7114-2024-19-3-102-112.]

9. Савинова Е. С. Использование потенциала нейросетей в процессе обучения РКИ на довузовском этапе. *Русский язык в полиэтническом образовательном пространстве военного вуза. Материалы IV Межвузовской научно-методической конференции*. СПб.: Военная академия связи; 2023:152–159. EDN: ORUFZF.

[Savinova E. S. Using the potential of neural network-based tools in the process of teaching RFL at the pre-university stage. *Russian Language in the Multiethnic Educational Environment of a Military University. Proc. 4th Interuniversity Scientific and Methodological Conf.* Saint Petersburg, Military Academy of Communications; 2023:152–159. (In Russian.) EDN: ORUFZF.]

10. Будехин С. Ю. Работа с неадаптированными текстами произведений русских классиков на уроках РКИ как средство установления межкультурного диалога. *Язык и литература в зеркале культуры: образовательные практики. Сборник научных трудов кафедры русского языка Российского государственного университета им. А. Н. Косыгина*. М.: Российский государственный университет им. А. Н. Косыгина; 2023:84–90. EDN: DKIXZK.

[Budekhin S. Yu. Working with non-adapted texts of Russian classics in RFL lessons as a means of establishing intercultural dialogue. *Language and Literature in the Mirror of Culture: Educational Practices. Collection of Scientific Articles of the Russian Language Department of the Kosygin State University of Russia*. Moscow, The Kosygin State University of Russia; 2023:84–90. (In Russian.) EDN: DKIXZK.]

11. Гальцева А. А. Работа над художественным текстом — важный фактор формирования коммуникативных навыков на уроках РКИ. *Современные проблемы филологии. Сборник материалов XI Международной научно-практической конференции*. Киров: Межрегиональный центр инновационных технологий в образовании; 2023:95–97. EDN: CBXQVO.

[Galtseva A. A. Work on a literary text as an important factor in the communication skills formation in the Russian as foreign language lessons. *Modern Problems of Philology. Proc. 11th Int. Scientific and Practical Conf.* Kirov, Inter-

regional Center for Innovation Technologies in Education; 2023:95–97. (In Russian.) EDN: CBXQVO.]

12. Митрофанова О. А., Адамова М. А., Букреева Л. А., Голубев Р. В., Гусьякая П. А., Зернова А. К., Литвинова А. А., Макеев К. В., Павликова В. С., Плюснина Е. А., Сологуб П. Ю., Сухан Д. Д., Трошина А. В., Уткина А. А. Интеллектуальный анализ данных в корпусе текстов по корпусной и компьютерной лингвистике. *International Journal of Open Information Technologies*. 2024;12(12):11–26. DOI: 10.18287/2542-0488-2024-12-3-34-45.

[Mitrofanova O. A., Adamova M. A., Bukreeva L. A., Golubev R. V., Gusyatskaya P. A., Zernova A. K., Litvinova A. A., Makeev K. V., Pavlikova V. S., Plyusnina E. A., Sologub P. Ju., Sukhan D. D., Troshina A. V., Utkina A. A. Data mining in the text corpus on corpus and computational linguistics. *International Journal of Open Information Technologies*. 2024;12(12):11–26. (In Russian.) EDN: PQKJB. DOI: 10.18287/2542-0488-2024-12-3-34-45.]

13. Тен Л. В. Тематическое моделирование в задаче автоматической рубрикации новостных текстов. *Terra Lingüistica*. 2023;14(2):77–91. EDN: AZVJTR. DOI: 110.18721/JHSS.14207.

[Ten L. V. Topic modeling in automatic categorization of news texts. *Terra Lingüistica*. 2023;14(2):77–91. (In Russian.) EDN: AZVJTR. DOI: 110.18721/JHSS.14207.]

14. Митрофанова О. А., Гаврилик Д. А. Эксперименты по автоматическому выделению ключевых выражений в стилистически разнородных корпусах русскоязычных текстов. *Terra Lingüistica*. 2022;13(4):22–40. EDN: HRUWKC. DOI: 10.18721/JHSS.13402.

[Mitrofanova O. A., Gavrilik D. A. Experiments on automatic keyphrase extraction in stylistically heterogeneous corpus of Russian texts. *Terra Lingüistica*. 2022;13(4):22–40. (In Russian.) EDN: HRUWKC. DOI: 10.18721/JHSS.13402.]

15. Глазкова А. В. Тематическая классификация текстовых фрагментов с учетом их ближайшего контекста. *Автоматика и телемеханика*. 2020;(12):153–172. EDN: RCYPRP. DOI: 10.31857/S0005231020120090.

[Glazkova A. V. Topical classification of text fragments accounting for their nearest context. *Avtomatika i telemekhanika*. 2020;(12):153–172. (In Russian.) EDN: RCYPRP. DOI: 10.31857/S0005231020120090.]

16. Wu X., Xia Y., Zhu J., Wu L., Xie Sh., Qin T. A study of BERT for context-aware neural machine translation. *Machine Learning*. 2022;111:917–935. DOI: 10.1007/s10994-021-06070-y.

17. Koroteev M. BERT: A review of applications in natural language processing and understanding. *ArXiv*. 2021:1–18. DOI: 10.48550/arXiv.2103.11943.

18. Buder-Gröndahl T. The ambiguity of BERTology: What do large language models represent? *Synthese*. 2024;203:15. DOI: 10.1007/s11229-023-04435-5.

19. Авраменко А. П., Тишина М. А. Дидактический потенциал лингвистических корпусов на базе технологий искусственного интеллекта для адаптации учебных материалов. *Вестник Московского государственного областного университета. Серия: Педагогика*. 2023;(1):29–38. EDN: MRNASM. DOI: 10.18384/2310-7219-2023-1-29-38.

[Avramenko A. P., Tishina M. A. Didactic potential of linguistic corpora based on artificial intelligence technologies for adapting learning materials. *Bulletin of Moscow Region State University. Series: Pedagogy*. 2023;(1):29–38. (In Russian.) EDN: MRNASM. DOI: 10.18384/2310-7219-2023-1-29-38.]

20. Gramner A. R. Feeling rules for professionals: Medical students constructing emotional labour in fiction talk. *Studies in Continuing Education*. 2023;45(3):265–282. DOI: 10.1080/0158037X.2022.2051474.

21. Häbler C., Spernes K. “The novel has simply made me more aware of how I meet other people” — attitude awareness in intercultural education. *Intercultural Education*. 2023;34(5):479–493. DOI: 10.1080/14675986.2023.2213672.

22. Atwal A., Sriram V., McKay E. A. Making a difference: Belonging, diversity and inclusion in occupational therapy. *British Journal of Occupational Therapy*. 2021;84(11):671–672. DOI: 10.1177/03080226211031797.

23. Heggernes S. L. A critical review of the role of texts in fostering intercultural communicative competence in the English language classroom. *Educational Research Review*. 2021;33:1–12. DOI: 10.1016/j.edurev.2021.100390.

24. Deveci T., Elgamal G., Dalton D., Langille D. J. The development of Emirati university students' intercultural sensitivity. *Learning and Teaching in Higher Education: Gulf Perspectives*. 2022;18(1):62–73. DOI: 10.1108/LTHE-09-2021-0087.

25. Xiaotian Zhang, Mingming Zhou. Interventions to promote learners' intercultural competence: A meta-analysis. *International Journal of Intercultural Relations*. 2019;71:31–47. DOI: 10.1016/j.ijintrel.2019.04.006.

Информация об авторах

Любимцева-Наталуха Лариса Николаевна, доктор филол. наук, доцент, профессор кафедры мировой литературы и сравнительного литературоведения, Донецкий государственный педагогический университет им. В. Шаталова, г. Горловка, Донецкая Народная Республика, Россия; *ORCID*: <https://orcid.org/0009-0006-2694-8168>; *e-mail*: lubimseva@mail.ru

Жихарева Наталья Анатольевна, канд. филол. наук, доцент, доцент кафедры романо-германской филологии, декан филологического факультета, Донецкий государственный педагогический университет им. В. Шаталова, г. Горловка, Донецкая Народная Республика, Россия; *ORCID*: <https://orcid.org/0000-0002-5762-2703>; *e-mail*: nataliiazhikhareva24@yandex.ru

Information about the authors

Larisa N. Lyubimtseva-Natalukha, Doctor of Sciences (Philology), Docent, Professor at the Department of the World Literature and Comparative Literature Studies, Donetsk State Pedagogical University named after V. Shatalov, Gorlovka, Donetsk People's Republic, Russia; *ORCID*: <https://orcid.org/0009-0006-2694-8168>; *e-mail*: lubimseva@mail.ru

Nataliia A. Zhikhareva, Candidate of Sciences (Philology), Docent, Associate Professor at the Department of Romance and Germanic Philology, Dean of the Faculty of Philology, Donetsk State Pedagogical University named after V. Shatalov, Gorlovka, Donetsk People's Republic, Russia; *ORCID*: <https://orcid.org/0000-0002-5762-2703>; *e-mail*: nataliiazhikhareva24@yandex.ru

Поступила в редакцию / Received: 25.12.24.

Поступила после рецензирования / Revised: 27.01.25.

Принята к печати / Accepted: 28.01.25.