

УДК 81 '33

Прокофьева Татьяна Владимировна
РГПУ им. А. И. Герцена
г. Санкт-Петербург, Россия

Tatyana Prokofyeva
Herzen University
Saint-Petersburg, Russia
tatyana.prokofyeva.herzen@mail.ru

Белоглазова Елена Владимировна
доктор филологических наук, доцент
РГПУ им. А. И. Герцена
г. Санкт-Петербург, Россия

Elena Beloglazova
PhD in Philology, professor
Herzen University
Saint-Petersburg, Russia
evbeloglazova@herzen.spb.ru

АНАЛИЗ ТОНАЛЬНОСТИ РУССОЦЕНТРИЧЕСКОГО ДИСКУРСА

Статья посвящена исследованию тональности руссоцентрического дискурса в современном медиапространстве. Актуальность работы определяется усилением интереса к национальной идентичности и идеологическим конструктам России, транслируемым через язык. Целью исследования является анализ тональности иноязычных текстов, посвященных России, для выявления ключевых тем, стереотипов и эмоциональных оценок. В качестве методологической основы применяются методы компьютерной лингвистики, в частности автоматический анализ тональности текста с использованием модель VADER/NLTK и моделью RoBERTa. Эмпирической базой для исследования послужил ad-hoc корпус заголовков книг о России (новостных и публицистических статей, блогов и социальных сетей, названий книг). Результаты анализа тональности разделены на «позитивные», «нейтральные» и «отрицательные». Также рассматриваются существующие ограничения используемых инструментов для автоматического анализа тональности. По данным исследования были выявлены закономерности работы используемых моделей и их особенности, которые необходимо учитывать при использовании данных инструментов.

Ключевые слова: оценка; оценочное суждение; тональность; ценность; анализ тональности; руссо-центрический дискурс.

SENTIMENT ANALYSIS OF RUSSIA-CENTRED DISCOURSE

The article is devoted to the study of sentiment in Russia-centric discourse in the contemporary media landscape. The relevance of the research is determined by the growing interest in national identity and Russia's ideological constructs transmitted through language. The aim of the study is to analyze the sentiment of foreign-language texts about Russia in order to identify key themes, stereotypes, and emotional evaluations. As a methodological basis, methods of computational linguistics are applied, in particular automatic sentiment analysis of text using the VADER/NLTK model and the RoBERTa model. The empirical basis for the study was an ad-hoc corpus of book titles about Russia (including news and journalistic articles, blogs, and social media, as well as book titles). The sentiment analysis results are categorized into “positive”, “neutral”, and “negative”. Existing limitations of the tools used for automatic

sentiment analysis are also considered. The study identified patterns in the performance of the models used and their specific characteristics, which must be taken into account when employing these tools.

Key words: evaluation; evaluative judgment; sentiment; value; sentiment analysis; Russia-centred discourse.

В современном мире в эпоху постглобализации продолжается глобализация информационного пространства, что, с учетом пристального внимания международного сообщества к России обуславливает актуальность исследования тональности руссоцентрического дискурса. Иноязычные описания России напрямую формируют общественное мнение как внутри страны, так и за её пределами, при издании книг и публикации новостных репортажей в других государствах. В условиях информационных войн и культурных различий формальный анализ оценочности иноязычных описаний России становится важным инструментом для понимания внешнего восприятия образа страны и её места в мире.

“Тайна оценочного значения на протяжении веков волновала философов, логиков и лингвистов”, констатирует Н. Д. Арутюнова, оценивая проблему как “твердый орешек”. С момента этой констатации наука, в том числе лингвистика, испытала существенные перемены, связанные с появлением технологий автоматизированного анализа текста с привлечением больших массивов данных.

Цель данной работы заключается в оценке адекватности автоматизированного анализа тональности на самобытном материале текстов иноязычного руссоцентрического дискурса.

Теоретический фундамент

Исследование имеет две теоретико-методологические предпосылки.

Во-первых, анализ тональности, являясь прикладной задачей, направленной на извлечение из текста оценочного компонента (“opinionmining”), требует уяснения лингвистической природы оценки.

Н. Н. Болдырев выделяет оценку как одну из трех типов категорий в языке. Под оценочной категоризацией исследователь понимает «группировку объектов и явлений по характеру их оценки в соответствующие оценочные классы и категории, то есть систему оценочных категорий (статический аспект), или мыслительное соотнесение объекта или явления с определенной оценочной категорией (динамический аспект)» [Болдырев 2002: 104]. Именно включение в поле оценочности динамического аспекта и делает оценку столь интересной проблемой для изучения: оценка есть не только неотъемлемое свойство определенного класса языковых единиц, но может возникать в коммуникации как косвенный прагматический эффект высказывания.

Из сказанного явствует, что оценка проявляется на всех уровнях речевой коммуникации, она может выражаться через различные лексические единицы, грамматические и стилистические средства, которые позволяют

выражать субъективное отношение субъекта к объекту оценки. Согласно исследованиям Е. М. Вольф, оценочные суждения в языке включают в себя субъективный (отражает личное отношение говорящего к предмету) и объективный (отражает признаки самого объекта оценки) [Вольф, 2002].

Словами, специализирующимися на выражении оценки, принято считать прилагательные, оценочные существительные и глаголы с явной оценкой, которые выражают положительное или отрицательное оценочное значение, данные слова служат основными показателями оценки в языке. Помимо данных слов, выделяют такие языковые явления, которые содержат в себе оттенки оценки:

1. Уменьшительно-ласкательные формы слов («глаза» – «глазоньки»);
2. Суффиксальные формы слов («дом» – «домик»);
3. Коннотация («политик» – «политикан»);
4. Различные интенсификаторы («очень», «довольно», «ультра-» и т. д.);
5. Междометия («Браво!», «Фу!» и т. д.);
6. Модальность;
7. Экспрессивные средства языка (интонация, грамматические конструкции, фразеологизм, метафора, сравнение и т. д.) [Вольф, 2002; Темигразина, 2015].

Оценка и тональность тесно связаны друг с другом, но оценка имеет более широкий спектр смысловых и прагматических аспектов (когнитивные, эмоциональные и поведенческие), в то время как тональность – более узкое понятие, является выражением эмоциональной окраски высказывания, его субъективной модальности и направленности (положительной, нейтральной, отрицательной) [Чернявская, 2020; Нефедов, 2019]. Таким образом, можно считать, что тональность есть частный случай оценки. Тональность формируется на основе оценочных суждений, но не всякая оценка выражается в тональности. Она зависит от оценки и отражает ее эмоциональный аспект. Анализ взаимосвязи оценки и тональности позволяет утверждать, что оценочные структуры лежат в основе формирования тональности текста, а их интерпретация требует комплексного подхода, учитывающего как лингвистический, так и культурно-прагматический контекст.

Во-вторых, исследование тональности представляет интерес с точки зрения изучения своеобразия иноязычного руссоцентрического дискурса, который можно определить как «особое использование языка для вербализации представлений о России и русской культуры, складывающееся в определенном дискурсивном сообществе под влиянием факторов культурной и идеологической дистанции». [Белоглазова, 2025: с 134] Определенные особенности данного дискурса описаны в работах Е. В. Белоглазовой и Н. А. Осьмак, где авторы отмечают своеобразные для дискурса оценочно-заряженные концептуальные доминанты.

Материалы и методы

Основными инструментами для автоматического анализа тональности стали модель VADER/NLTK и модель RoBERTa.

Для работы с моделью VADER/NLTK были заданы определенные параметры, относительно которых она определяла какой тип тональности необходимо выставить для каждого текста. Так, текстопределялся:

1. «положительным», если общее значение текста заголовка равнялось или превышало показатель 0,5;
2. «нейтральным», если общее значение текста заголовка находилось в диапазоне от -0,5 до 0,5;
3. «отрицательным», если общее значение текста заголовка равнялось или было ниже -0,5.

При использовании стандартных настроек модель RoBERTa не выделяет тексты с нейтральной тональностью, разделяя их только на положительные и отрицательные. Однако, можно самостоятельно задать параметры, в рамках которых тексты будут определены как тексты с нейтральной тональностью (в данном случае значение $< 0,6$).

Для проведения анализа тональности был выбран ad-hoc корпус заголовков англоязычных статей и книг, посвященных России, Rossica-T, включающий 662 заголовка. Выбор материала обусловлен минимальной значимостью в нем контекста, что позволяет в наиболее чистом виде оценить действие словарного подхода к анализу тональности.

Результаты исследования

Результаты проведения анализа тональности представлены в виде круговых диаграмм с процентным соотношением автоматически определенных положительной, нейтральной и отрицательной тональностей (рис. 1 и рис. 2).

Результаты анализа тональности (VADER/NLTK)

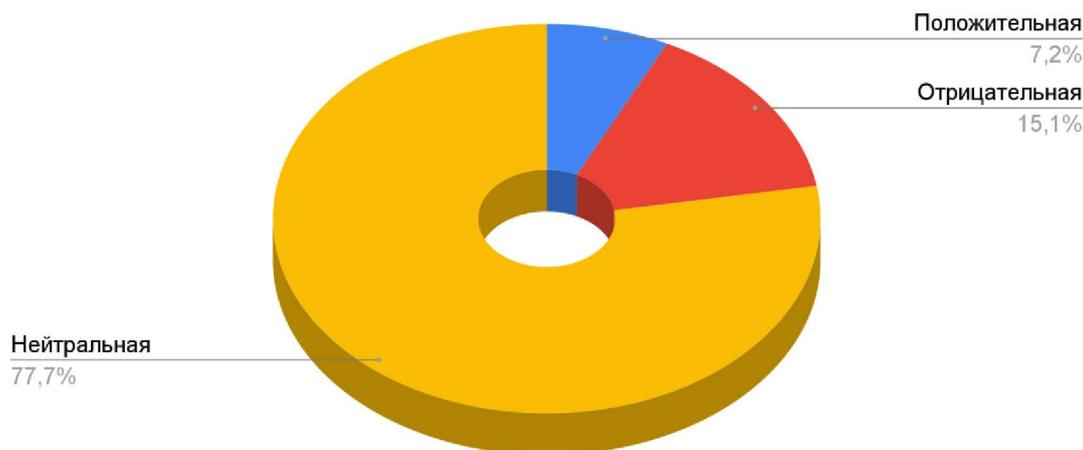


Рис. 1. Диаграмма с результатами анализа тональности (VADER/NLTK)

Результаты анализа тональности (модель RoBERTa)

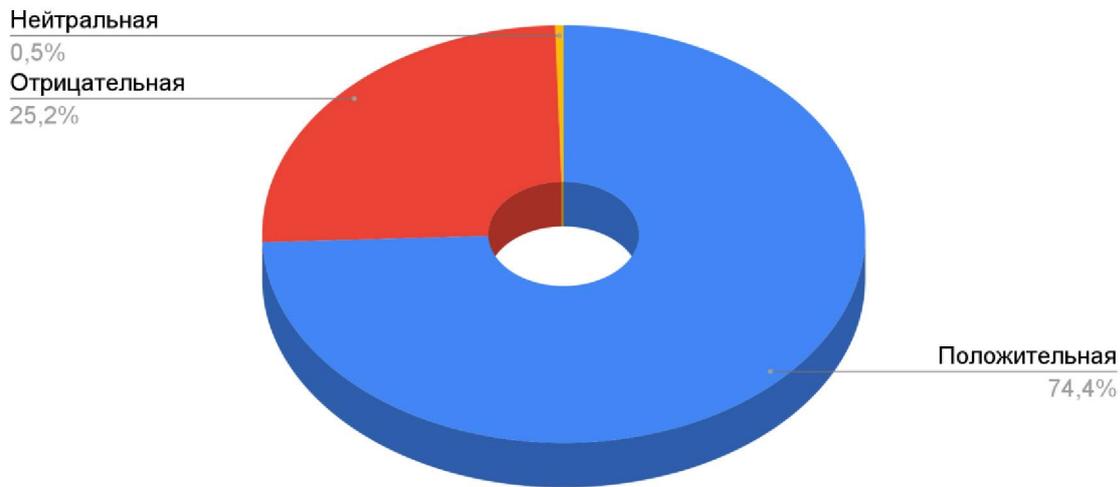


Рис. 2. Диаграмма с результатами анализа тональности (модель RoBERTa)

При сравнении результатов анализа тональности выбранными инструментами можно заметить большую разницу в соотношении текстов. Так, модель RoBERTa выдает более точные результаты, где тексты положительной тональности занимают 74,4 %, с нейтральной тональностью составляют 0,5 %, а с отрицательной – 25,2 %. В результате анализа с помощью данной модели можно увидеть, что положительная тональность превалирует над отрицательной. В то же время модель VADER/NLTK выделяет большую часть заголовков, как тексты с нейтральной тональностью – 77,7 %, соотношение текстов с положительной и отрицательной тональностями противоположно тому, что показывает диаграмма модели RoBERTa, заголовки с отрицательной тональностью составляют 15,1 %, а с положительной – 7,2 %.

Данные, полученные в результате исследования, подтверждают актуальность применения методов автоматического анализа тональности для исследования образа России в англоязычном дискурсе. Использование двух инструментов – библиотеки NLTK и модели RoBERTa – позволило выявить как общие закономерности, так и значительные расхождения в оценке одного и того же корпуса текстов.

С одной стороны, модель RoBERTa показала большую чувствительность к эмоционально окрашенной лексике, что было отражено в процентном соотношении текстов с положительной тональностью (74,4 %) и отрицательной тональностью (25,5 %), в то время как заголовки с нейтральной тональностью составили всего 0,5 %. Это может быть связано как с особенностями обучающего корпуса модели, так и с контекстной глубиной анализа, обеспечиваемой архитектурой данной модели.

С другой стороны, модель VADER/NLTK показала большой процент текстов с нейтральной тональностью – 77,7 %, это может быть обусловлено тем, что данная модель ориентирована на словарные значения и ограничена в способности считывать дискурсивный контекст.

При сравнении моделей можно сделать вывод, что для более точного анализа больше подходит модель RoBERTa, так как полученные результаты обладают большей точностью, но для анализа данных требуется больше времени. Модель VADER/NLTK же подходит для первичного анализа, либо для определенных исследований, в которых необходимо задавать собственные параметры определения тональности при анализе. RoBERTa выдает результаты с выраженной бинарной полярностью, в то время как VADER демонстрирует склонность к оценочной умеренности.

При необходимости анализа большого количества данных, автоматический анализ тональности значительно сокращает время, необходимое для его реализации. Использование автоматического анализа тональности позволит эффективнее изучить иноязычного руссоцентрического дискурс. Однако, необходимо учитывать особенности используемых инструментов, настройку моделей с учетом специфики культурного, жанрового и идеологического контекста, а также ведение контроля качества полученных данных.

ЛИТЕРАТУРА

1. Арутюнова Н. Д. Типы языковых значений: Оценка. Событие. Факт. / Арутюнова Н. Д., Степанов Г. В. – М.: Наука, 1988. – 339 с.
2. Болдырев Н. Н. Структура и принципы формирования оценочных категорий // С любовью к языку: Сб. Научных трудов. – Москва-Воронеж: ИЯ РАН, Воронежский государственный университет, 2002. – С. 103–114.
3. Вольф Е. М. Функциональная семантика оценки / Е. М. Вольф – 2-е изд., доп. – М.: Едиториал УРСС, 2002. – 280 с. – Текст: электронный – URL: <https://reallib.org/reader?file=579811&pg=111> (дата обращения: 12.01.2025).
4. Темиргазина З. К. Лингвистическая аксиология: оценочные высказывания в русском языке: монография / З. К. Темиргазина. – 2-е изд., стер. – Москва: ФЛИНТА, 2015. – 247 с. – ISBN 978-5-9765-2118-6. – Текст: электронный // Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/70403> (дата обращения: 04.01.2025). – Режим доступа: для авториз. Пользователей.
5. Чернявская В. Е. Язык оценок в научном дискурсе: терминологическое поле и методологические подходы / В. Е. Чернявская // Вестник Пермского университета. Российская и зарубежная филология. – 2022. – Т. 14, № 3. – С. 44-55. – DOI 10.17072/2073-6681-2022-3-44-55. – EDN SVXEVH.
6. Нефедов С. Т. Варьирование оценки в коммуникативных практиках научного дискурса / С. Т. Нефедов // Вестник Санкт-Петербургского университета. Язык и литература. – 2021. – Т. 18, № 4. – С. 760–778. – DOI 10.21638/spbu09.2021.408. – EDN HPNSWE.

7. Белоглазова Е. В. Образ России в зеркале японоязычных кинокритических рецензий / Белоглазова Е. В., Сосункевич В. О. // Социальные и гуманитарные науки на Дальнем Востоке. – 2025. – Том 22, N 1. – С. 132–137. –DOI: 10.31079/1992-2868-2025-22-1-132-137.
8. Белоглазова Е. В. Rossica и Belarossica: общие и культурно-специфические маркеры дискурсов иноязычного описания России и Беларуси / Белоглазова Е. В., Генидзе Н. К. ; RossicaV. Belarossica: UniversalandCulture-Specific Features of Russia- and Belarus-Centered Discourses // Вопросы когнитивной лингвистики. – 2023. –N 4. – С. 108–115.
9. Белоглазова Е. В., Осьмак Н. А., Шувалова Е. К. Формальные маркеры и содержательные доминанты руссоцентрического дискурса: кросс-языковое корпусное исследование // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2024. Т. 22, No 1. С. 105–119. DOI 10.25205/1818-7935-2024-22-1-105-119.
10. Осьмак Н. А. Образ Сибири в современном финноязычном новостном дискурсе = The Image of Siberia in Contemporary Finnish-language News Discourse / Осьмак Н. А. // Журнал Сибирского федерального университета. Гуманитарные науки. – 2025. – Том 18, N 1. – С. 21–33. – URL: <https://elibrary.ru/item.asp?id=80336744>.
11. Семина Т. А. Тональность текста: синтаксические паттерны выражения отношений между сущностями : специальность 10.02.19 «Теория языка» : защищена 20.05.2020 / Семина Т. А.; Российский университет дружбы народов; научный руководитель Максименко О. И. – Мытищи, 2020. – 24 с. – Текст: электронный // Elibrary: электронно-библиотечная система. – URL: https://www.elibrary.ru/download/elibrary_45677042_93462808.pdf (дата обращения 15.04.2025). – Режим доступа: для авториз. Пользователей.
12. More than a Feeling: Accuracy and Application of Sentiment Analysis / Jochen Hartmann, Mark Heitmann, Christian Siebert, Christina Schamp // International Journal of Research in Marketing. – 2023. – Vol. 40, Issue 1. – pages 75–87. – ISSN 0167-8116.