



DOI: 10.22363/2313-2299-2025-16-3-760-782

EDN: DCBVEJ

УДК 811.111'32'37:159.942


Научная статья / Research article

Сентимент анализ как инструмент лингвоэмотиологии: оценка потенциала систем анализа тональности текста

О.И. Максименко¹  , М.В. Беляков² 

¹Государственный университет просвещения, Москва, Российская Федерация

²Московский государственный институт международных отношений (университет) МИД
России, Москва, Российская Федерация

 maxbel7@yandex.ru

Аннотация. Оценка тональности текста в больших информационных потоках решается как качественными так и количественными методами. Качественные методы включают в себя, в первую очередь, методы лингвоэмотиологии, включая составление тональных словарей, используемых в том числе в компьютерных системах оценки тональности текста. В статье рассматриваются принципы функционирования систем автоматического анализа текста как метода компьютерного анализа текста, приводится анализ ряда современных систем анализа тональности текста. Определяются характеристики систем, на языковом материале размеченных корпусов кинокритик и отзывов на товары известного маркетплейса выявляются достоинства и недостатки анализируемых систем. Особое внимание уделяется лингвистическим причинам недостатков оценки тональности, таким как многоязычие, разные способы представления текста пользователями, включая сокращенные формы или аббревиатуры, расшифровка которых затрудняет процесс анализа, жанровое разнообразие, имплицитные оценки, полисемия и омонимия лексики, модификаторы полярности, ирреальное наклонение, сарказм, ирония и пр. По результатам исследования делается вывод, что наиболее эффективным функционалом определения тональности, необходимым оценочном средстве для лингвоэмотиологии и лингвоконфликтологии, обладают программы, использующие гибридный метод. В работе предлагаются возможные подходы к оптимизации функционирования программ. Исследование позволяет приблизиться к более четкому пониманию феномена выявления тональности текста и выбора для этих целей систем сентимент анализа, основанных на разных принципах функционирования. Такие системы, с одной стороны, решают прикладные задачи сентимент анализа, а с другой — являются источником материала для исследований в рамках лингвистической теории эмоций.

Ключевые слова: сентимент, эмотивность, лингвистические теории эмоций

Вклад авторов: вклад авторов равнозначен на всех этапах исследования и подготовки текста статьи.

Заявление о конфликте интересов: авторы заявляют об отсутствии конфликта интересов.

© Максименко О.И., Беляков М.В., 2025



This work is licensed under a Creative Commons Attribution 4.0 International License
<https://creativecommons.org/licenses/by-nc/4.0/legalcode>

История статьи: дата поступления: 28.12.2024; дата приема в печать: 15.03.2025.


Для цитирования: Максименко О.И., Беляков М.В. Сентимент анализ как инструмент лингво-эмотиологии: оценка потенциала систем анализа тональности текста // Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. 2025. Т. 16. № 3. С. 760–782. <https://doi.org/10.22363/2313-2299-2025-16-3-760-782>

Sentiment Analysis as a Tool of Linguistic Emotionology: Assessment of the Text Tonality Analysis Systems Potential

Olga I. Maksimenko¹  , Mikhail V. Belyakov² 

¹ Federal State University of Education, *Moscow, Russian Federation*

² Moscow State Institute of International Relations (University) Ministry of Foreign Affairs of the Russian Federation, *Moscow, Russian Federation*

 maxbel7@yandex.ru

Abstract. The assessment of text tone in large information flows is solved using both qualitative and quantitative methods. Qualitative methods include, primarily, the methods of linguistic emotionology, such as the compilation of tone dictionaries used in computer systems to evaluate the tone of a given text. The article discusses the principles of functioning of automatic text analysis systems as a method of computer-aided text analysis. It also provides an analysis of several modern tone analysis systems. Characteristics of these systems are identified, and advantages and disadvantages are revealed based on linguistic material from marked-up film and product reviews from a well-known online marketplace. Special attention is given to linguistic reasons behind the challenges in assessing tone, such as multilingualism, different ways of presenting text by users, including abbreviated forms, can make the process difficult to understand. Genre diversity, implicit assessments, polysemy, homonymy, polarity modifiers, subjunctive mood, sarcasm, irony are all factors that can complicate the process of determining the tonality of a piece of text. Based on the results of our study, we conclude that programs using a hybrid method have the most effective functionality for detecting tonality. These programs are an important tool for linguistic emotionology and linguoconflictology, as they provide a necessary evaluative component. The research suggests possible approaches for optimizing the functioning of these programs. The study allows us to gain a better understanding of the challenges associated with detecting tonality in text and selecting sentiment analysis systems that operate on different principles. These systems not only solve practical problems related to sentiment analysis but also serve as a valuable source of material for research within the linguistic theory of emotions.

Keywords: opinion sentiment, emotivity, linguistic theory of emotions

Authors' contribution: the authors contributed equally to this article and preparation of the article text.

Conflicts of interest: the authors declare no conflict of interest.

Article history: received: 28.12.2024; accepted: 15.03.2025.

For citation: Maksimenko, O.I., & Belyakov, M.V. (2025). Sentiment Analysis as a Tool of Linguistic Emotionology: Assessment of the Text Tonality Analysis Systems Potential. *RUDN Journal of Language Studies, Semiotics and Semantics*, 16(3), 760–782. (In Russ.). <https://doi.org/10.22363/2313-2299-2025-16-3-760-782>

Введение

Лингвистическая теория эмоций, или лингвоэмотиология, за последние десятилетия прошла этап становления и превратилась в самостоятельное и чрезвычайно востребованное направление современной лингвистики.

Виктор Иванович Шаховский, один из основоположников лингвистической теории эмоций, писал: «Эмоции как психическое явление отражают (то есть воспроизводят) в сознании человека его эмоциональные отношения к миру. Эти эмоциональные отношения являются хотя и субъективными, но социально осознанными и потому в определенной степени типизированными» [1. С. 88]. В семантику любого слова входят «актуальные или потенциальные семы *эмоция* и *оценка*» [2. С. 24], таким образом, оценка, эксплицитно или имплицитно реализованная в коммуникации, выражает мнение об объекте, действии и пр. и определяет статус эмотивной семантики лексической единицы. Типизация вербализованных эмоций позволяет проводить исследование тональности текста в автоматическом режиме. «В рамках дискурс-анализа понятие *тональности* не равно понятию *оценки*, а отражает стиль и «тон» коммуникации. За последнее время семантика термина *тональность* настолько расширилась, что можно говорить о фактической омонимии термина *тональность* как стилистического маркера и термина, связанного с оценкой высказывания в рамках анализа тональности (мнения) или компьютерного сентимент-анализа» [3. С. 186].

Сентимент-анализ как метод компьютерной лингвистики

С ростом интернет-сообщества увеличивается объем неструктурированной информации в сети: обзоры, комментарии разных событий, происходящих в мире. Извлечение сведений и фактов из материалов значительного объема с учетом высказываемого в них мнения вручную представляется задачей архисложной, а в ряде случаев и невозможной. Сентимент анализ является одним из современных инструментов аналитики, позволяющим выявить тональность текста при помощи методов NLP (Natural Language Processing), статистики и машинного обучения [4–6]. Под тональностью следует понимать эмоциональную составляющую, которая отражает отношение субъекта (автора высказывания) к объекту (к тому, о чем высказывается мнение).

При определении тональности текста, как правило, на первый план выходят несколько базовых компонентов:

- автор мнения (opinion holder), иногда именуемый источником мнения (opinion source), то есть некий субъект (отдельное лицо, группа людей, организация и т.п.), которому принадлежит мнение;
- объект (opinion target/entity) — сущность, об аспекте которой высказано мнение;
- сам аспект объекта, находящийся в центре внимания;

- содержание мнения (opinion content) — смысл, который вложил автор в свое высказывание.

Тональность (opinion sentiment), которая определяется путем анализа содержания (opinion content), показывает, какой сентимент несет в себе это изречение, то есть что чувствует, думает автор относительно того, о чем он пишет. Может проводиться анализ контекста (opinion context), что позволяет выявить, при каких условиях и обстоятельствах было высказано то или иное мнение и время выражения мнения (time) [7; 8].

Существуют разные виды тональной классификации. Наиболее простым способом является бинарное распределение, когда используются только два класса: положительный и отрицательный. Несмотря на кажущуюся простоту такого вида классификации, в некоторых случаях задача может быть усложнена. Например, иногда нужно определить, является ли политическая речь «за» или «против» обсуждаемого вопроса на дебатах или распределить рецензии на фильмы на классы: рекомендованные к просмотру или не рекомендованные [9]. Серьезным минусом данной классификации является то, что не всегда выражения можно однозначно отнести к определенному классу, так как они могут быть нейтральными. Более того, некоторые тексты могут содержать как позитивное, так и негативное отношение автора. В.И. Шаховский писал, что эмоции (и оценка) могут быть амбивалентными, т.е. «любовь и ненависть могут овладеть человеком одновременно. Данный факт указывает на полярность некоторых эмоций. Они могут образовывать оппозиции по типу оценочного знака (положительные/отрицательные), что отражается в лексической системе языка в сферах номинации эмоций (любовь/ненависть) и их выражения (замечательный/отвратительный)» [10. С. 30]. Таким образом, бинарная классификация эффективна только в случае выражения автором эксплицитной оценки (строго позитив или негатив). Усложненной версией бинарного распределения является классификация с добавлением нейтрального класса и дополнительных подклассов.

Автоматическое классифицирование текстовых фрагментов по критерию доминирующей в них эмоции еще один вид классификации. Так, возможно распределение текстов по 9 эмоциональным категориям, предложенным Г. Левхеймом: 1) интерес — возбуждение; 2) удовольствие — радость; 3) удивление; 4) страдание — тоска; 5) страх — ужас; 6) стыд — унижение; 7) брезгливость — отвращение; 8) злость — гнев; 9) нейтральные тексты [11]. Первая эмоция в классе отражает слабую степень выраженности, а вторая — наивысшее проявление [12; 13]. Использование n-балльного шкалирования позволяет получить числовое значение тональности. Текстовому фрагменту (t) присуждается числовое значение, например от -5 до 5 , то есть если $t > 0$, то тональность положительная, если $t < 0$, то отрицательная, в случае когда $t = 0$, сентимент нейтральный.

Важнейшим понятием, связанным с тональностью, является субъективность. Тексты, содержащие мнения, оценки, эмоции, предположения попадают в категорию субъективных. К объективному классу относятся текстовые фрагменты, выражающие фактическую информацию без вмешательства личных интерпретаций. Укоренившаяся классификация по делению языковых единиц на субъективные и объективные единицы была признана неэффективной. Интерес исследователей перешел к классификации по релевантности. «Тональная релевантность — это понятие, позволяющее отличать информативное содержание для определения тональности документа от неинформативного. Оно контрастирует с обычным различием между субъективным и объективным содержанием» [14. С. 40]. Например, предложение “*Bruce Banner, a genetics researcher with a tragic past, suffers a horrible accident*” является субъективным, так как словосочетания *tragic past* и *horrible accident* — субъективные понятия, вызывающие разные ассоциации у читателя. Несмотря на то, что предложение имеет субъективный негативный окрас, оно нерелевантное, поскольку связано с сюжетом фильма и может появляться в положительных отзывах. Объективное предложение “*The movie won a Golden Globe for best foreign film*” будет релевантным, так как выражает позитивное мнение автора [15. С. 955].

Анализировать тональность можно на уровне документа, в данном случае следует выявить общее мнение, выраженное во всем тексте; на уровне предложения происходит разметка каждого предложения в тексте; на уровне фраз в центре внимания не целые предложения, а отдельные фрагменты; определение тональности на аспектуальном уровне является самым точным и в то же время самым сложным видом анализа.

Принято выделять несколько подходов к распознаванию тональности:

- на основе словарей оценочной лексики, когда каждому слову присваивается индекс в виде числовой характеристики из словаря, затем подсчитывается среднее арифметическое всех слов в тексте и определяется тональность;
- на основе правил (в этом случае экспертом составляется перечень правил в зависимости от языка и предметной области анализируемого фрагмента);
- с применением машинного обучения с учителем, когда классификатор обучается на размеченной коллекции текстов, составленной вручную, после чего строится классификатор, определяющий тональность новых документов;
- с применением машинного обучения без учителя, когда отличие от вышеупомянутого метода заключается в отсутствии заранее размеченной выборки. Идея заключается в выделении терминов с наибольшим весом. В эту категорию попадают термины, которые встречаются чаще всего в конкретном тексте и присутствуют в небольшом количестве в остальных текстах из коллекции, на их основе определяется тональность;

- гибридный метод, который предполагает использование нескольких рассмотренных выше способов, к примеру, тональные словари и правила.

Несмотря на большое количество трудов и проведенных экспериментов по тональному анализу эксперты сталкиваются с рядом факторов, препятствующих точному анализу, среди них многоязычие; разные способы представления текста пользователями, например, сокращенные формы или аббревиатуры, расшифровка которых затрудняет процесс; жанровое разнообразие; имплицитные оценки; многозначность лексики; модификаторы полярности; ирреальное наклонение; определение реальных чувств сообщаемой информации; успех или неудача одной стороны относительно другой; сарказм и ирония и пр.

Материал исследования

Существующие программные решения в большинстве случаев ориентированы на анализ тональности английских текстов. Количество систем для русского языка весьма ограничено, при этом большая часть из них находится в закрытом доступе. В начале 2025 г. был представлен следующий список лучших программ сентимент анализа: Brand24, Qualtrics, SproutSocial, OpenText, Semantria, Meltwater, Dialpad, MonkeyLearn, Sentigem, SentiStrength, Reputation.com, Mentionlytics, Hootsuite, SentiSum, Mediallia¹. Некоторые из них входили в списки надежных систем и ранее.

С целью тестирования программ был проведен обзор находящихся в свободном доступе систем автоматического определения тональности для английских текстов. Среди рассматриваемых были системы, работа которых основана на разных принципах. Принцип машинного обучения: Eureka Engine, SummarizeBot, MonkeyLearn, Microsoft Text Analytics API, Sentiment analyzer, Texterra, Megaputer. Принцип рекурсивных нейросетей — Stanford NLP; рекуррентных нейросетей — ParallelDots, словарный метод — Sentistrength, гибридный метод — Repustate, работа на правилах — Watson Tone Analyzer, Pattern; Sentigem — принцип неизвестен.

Тестирование проводилось на открытых корпусах кинокритик (IMDb²) и отзывов на товары электронной торговой площадки (Amazon³). Определяющими факторами при выборе языкового материала стали следующие:

- 1) тексты подобного рода априори предполагают выражение мнения об объекте и, соответственно, его оценку, выраженную вербально;
- 2) возможность получения доступа к уже размеченному корпусу текстов.

¹ Режим доступа: <https://brand24.com/blog/best-sentiment-analysis-tools/> (дата обращения: 15.12.2024).

² Режим доступа: <https://www.kaggle.com/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews?select=IMDB+Dataset.csv> (дата обращения: 15.12.2024).

³ Режим доступа: <https://www.kaggle.com/datasets/kritanjali/jain/amazon-reviews> (дата обращения: 15.12.2024).

Процедура и результаты

Для тестирования систем автоматического определения тональности с целью выявления имеющихся недостатков и достоинств было отобрано пять программ, различающихся по принципам работы.

1. *Sentistrength*

Программа строится на основе словаря эмоционально окрашенных слов с соответствующими весами, отражающими силу тональности. Международная социальная сеть MySpace является источником создания встроенных словарей. При скачивании программы в обязательном порядке загружается архив, в котором помимо английского словаря также присутствуют: словарь слов-усилителей и словарь слов, понижающих тональность; словарь эмотиконов; словарь идиом; словарь сленга, словарь вопросительных и отрицательных лексем. Следует отметить, что слова, включенные в словарь, прошли процесс стемминга и представлены без аффиксов. Это значит, что при анализе будут учтены все словоформы, представленные в рецензиях. Для получения результатов отдельно суммируются веса положительных и отрицательных лексем. Для оценки тональности рецензии суммировались два полученных результата (см. Табл. 1).

Таблица 1 / Table 1

Результаты тестирования программы *Sentistrength* / Test results of *Sentistrength*

Метрики классификации / Rating Metrics	Кинорецензии / Film Reviews	Отзывы Amazon / Amazon Reviews
Точность определения положительных отзывов / Accuracy of positive feedback	0,68	0,65
Точность определения отрицательных отзывов / Accuracy of negative feedback	0,77	0,7
Правильность классификации / Correct classification	70/100	60/100

Источник: составлено О.И. Максименко, М.В. Беляковым. / Source: compiled by Olga I. Maksimenko, & Mikhail V. Belyakov.

Возможные причины ошибок:

В отзывах с описанием сюжетов фильмов о военных событиях и последствиях войны превалирует отрицательно окрашенная лексика, например: *quick-and-dirty battle, disgusted with the gruesome war, it's hellish effect on the land is timelessly relevant, terrible repercussions*. Говоря о сюжете, пользователь выражает позитивное мнение о том, как представлены военные картины (*“It is a unique vision of an era that is sure not only to entertain but also to truly absorb the audience into the lives of a people torn apart by a wa”, “an inspiring war drama”*), однако программа определяет общую тональность как отрицательную, поскольку при подсчете весов отрицательной лексики оказывается больше. В данном случае следует говорить о тональной релевантности,

поскольку описание сюжета фильма не имеет отношения к тональности текста. То же самое касается и рецензий, в которых автор высказывает мнение о триллерах или фильмах ужасов. Как правило, главными механизмами воздействия таких жанров являются тревога и напряжение, в связи с чем отзывы насыщены лексикой с отрицательной коннотацией, что, в свою очередь, служит причиной ошибочного отнесения отзыва к отрицательному классу. Примеры: “*beats their host (Sam Levene) to death*”, “*a gay-bashing murder*”, “*Ryan’s murderous rage*”, “*Restlessness*”, “*the mad scientist is shot and drops dead shortly afterwards*”, “*Ryan, naturally, does his prototypical Angry White Male*”, “*they report these murders to the police*”, “*a mission of revenge*” и пр.

В данном случае отнесение выделенных лексем к отрицательному классу неправомерно, поскольку они отражают специфику жанра и являются элементом описания сюжета, а не оценки.

Еще одной причиной некорректности работы системы является присвоение множеству окрашенных слов нулевого веса, что говорит о возможном отсутствии в тональном словаре значимого количества лексем, вызванного несистематичностью проведения обновлений. Отсутствие актуализирующих словарь систематических обновлений является причиной того, что многие выражения, идиомы, сокращения, используемые в настоящее время в сети, в словаре не представлены. Например, прилагательным *taut* и *gripping* присвоен нулевой вес, однако они являются положительно окрашенными и выражают мнение автора.

К недостаткам программы следует отнести и то, что её работа строится без учета синтаксиса, каждое слово анализируется отдельно, а это ведет к ошибке в результатах. Например:

- вне контекста существительное *restlessness* можно отнести к отрицательным, но в словосочетании *cope with their restlessness* тональность меняется на положительную;
- существительным *fight* и *evil* программа справедливо присвоила отрицательный вес (-2), однако если объединить их в тональную цепочку, то словосочетание *fight against evil* следует отнести к разряду положительных;
- прилагательное *crude* имеет несколько значений. Программа помечает данное слово как отрицательное, поэтому можно предположить, что оно было рассмотрено в значении «грубый», «резкий» или «оскорбительный», однако в контексте, в котором оно встречается в отзыве (*crude crayon drawing*), прилагательное имеет нейтральную окраску. Это подтверждает, что слова при обработке рассматриваются по отдельности, а не в словосочетаниях.

Такая же ошибка программы допущена при классификации одного из отзывов, когда существительное *charger* (зарядное устройство) получило отрицательный индекс ввиду проекции значения глагола *charge* (взимать плату; обвинять) на данное существительное. Это свидетельствует об ограниченности

встроенных в программу лексиконов и о необходимости расширения диапазона значений.

Несмотря на заявленный инструмент по коррекции правописания, система не распознает характерные для языка интернета сокращения, выставляя им нулевую тональность. Например: *risible'n'ridiculous (risible and ridiculous)*.

К особенностям работы системы относится определение тональности всех значимых лексических единиц без исключения, что ведет к ошибке. “Night of Terror” (название фильма) в рецензии было выделено кавычками, но при анализе данный знак препинания был удален, существительное *terror* получило отрицательный вес, что повлияло на общую тональность. Избежать появления подобных ошибок можно путем введения правила, запрещающего программе оценивать именованные сущности, однако в таком случае система уже будет работать по гибриднему принципу (лексиконы + правила).

К серьезным недостаткам следует отнести неверное определение частей речи и снятие омонимии, принципиально важных при анализе, поскольку одно и то же слово может иметь разную тональность в зависимости от того, какой частью речи в предложении оно является. Например, *like* может выступать в роли наречия, которое является нейтральным, однако нередко оно имеет положительный вес, указывающий на то, что данное слово было рассмотрено как существительное или глагол; подобная ситуация нередка и для слова *kind*, которое может быть существительным с нейтральной окраской, а не прилагательным.

Стоит отметить, что отрицательные отзывы классифицированы с меньшей точностью. Это можно объяснить игнорированием отрицательных слов, которые в ряде предложений меняют тональность на противоположную или снижают вес.

Среди недостатков, ведущих к ошибочной работе систем, была выделена проблема имплицитного выражения мнения. Так, в одном из отзывов пользователь перечисляет утомительную последовательность действий, которую ему пришлось выполнить для установки игры. Отзыв не насыщен отрицательно окрашенной лексикой, а содержит объективные факты, поэтому программа определила данный отзыв как нейтральный. Несмотря на наличие отрицательных модификаторов полярности *will not*, *no*, а также отрицательного наречия *begrudgingly*, финальное решение было выведено неверно.

В случае, где речь шла о необходимости адаптера для работы устройства, автор также не использует отрицательную лексику, давая объективную оценку. Имплицитная оценка — одно из наиболее актуальных препятствий для систем тональной классификации.

Несмотря на достаточно высокий результат, работа системы могла бы стать более корректной при внесении таких изменений, как:

- регулярное обновление и пополнение всех включенных словарей, которое даст возможность распознавания большего количества лексики;

- создание тематических лексиконов для каждого набора текстов (в данном случае — для кинокритик и отзывов о товарах, преимущественно технических устройств), что позволит системе «адаптироваться» под конкретную тематику и выводить правильные веса для лексем (например, в контексте фильмов-ужасов прилагательные «кровавый», «ужасный» могут получить положительный вес);
- подключение синтаксического анализатора, который поможет выделять слова в синтаксические цепочки и анализировать словосочетания, а не отдельные слова;
- введение правила по присуждению нейтрального веса именам собственным, что также может повысить правильность и точность классификации.

2. SummarizeBot

Эта программа использует принцип машинного обучения для определения тональности, потому в результате дается не просто ответ о принадлежности текста к классу, а приводится подробный анализ, что позволяет анализировать причины допущенных ошибок. Каждому предложению присваивается положительный или отрицательный вес в зависимости от его тональности. Слова и выражения с позитивным сентиментом выделяются в анализируемом тексте зеленым цветом, с негативным — красным. Кроме того, выделяются объекты, по отношению к которым выражается мнение. Общая тональность выводится путем подсчета весов всех предложений (см. Табл. 2).

Таблица 2 / Table 2

Результаты тестирования программы SummarizeBot / Test results of SummarizeBot

Метрики классификации / Rating Metrics	Кинокритик / Film Reviews	Отзывы Amazon / Amazon Reviews
Точность определения положительных отзывов / Accuracy of positive feedback	0,79	0,72
Точность определения отрицательных отзывов / Accuracy of negative feedback	0,75	0,78
Правильность классификации / Correct classification	77/100	75/100

Источник: составлено О.И. Максименко, М.В. Беляковым. / Source: compiled by Olga I. Maksimenko, & Mikhail V. Belyakov.

Возможные причины ошибок:

Главным недостатком систем, которые используют машинное обучение в качестве подхода, является предметная ориентированность обучающей выборки. Программа хорошо справляется с текстами, которые максимально приближены к тренировочным, в других случаях возникают следующие неточности: если части сложносочиненного предложения соединены союзом «но», то в большинстве случаев доминирующей тональностью будет часть, идущая после коннектора, которая противопоставляется предыдущей. В ряде

случаев видим, что данное правило не учитывается, и общая тональность определяется путем сложения весов всех тонально окрашенных слов, выделенных программой, и это ведет к ошибкам.

Данное правило актуально не только внутри одного предложения, но и между самими предложениями. Программа учитывает модификаторы полярности. При появлении отрицательных наречий, вспомогательных или модальных глаголов с частицей *not* тональность последующей лексемы меняется на противоположную (в случае отрицательной коннотации слова).

Стоит уточнить, что отрицательные слова не всегда меняют тональность, например, устойчивая конструкция *no less* не выражает отрицание, а используется, чтобы подчеркнуть важность чего-либо. Например, в случае “*And Courtenay is no less convincing as the mincing dresser*”, где конструкция *less than* не является отрицанием, программа не распознала это выражение. *If not* в ряде контекстов также не меняет тональность на отрицательную; *can't help* — устойчивое выражение без негативного сентимента; *no... like this* — конструкция, имеющая в контексте отзыва ярко положительную окраску и означающая, что описываемый объект единственный в своем роде, ему нет равных, также в отзыве “*No another grill like this*” была неверно классифицирована; существительное с явно отрицательной коннотацией *violence* получает позитивный окрас, что объясняется стоящей перед ним отрицательной частицей *not* (“*Not just violence, but injustice...*”), которая меняет тональность, однако в данном случае это является ошибкой, поскольку в данном случае это не частица, а парный союз «не только..., но и...», в английском варианте — *not only..., but also*. Причиной данной неточности может быть тот факт, что в рецензии данный союз представлен в видоизмененном виде — *not just..., but...* Возможно, в тренировочной выборке присутствовали тексты, где союз был только в исходной форме, поэтому в рецензии распознан не был.

Заявлено, что в программе есть лингвистический анализатор, который автоматически проводит сегментацию текста, разбивая его на предложения, осуществляет процесс лемматизации, определяет часть речи у каждого слова, выявляет синтаксические цепочки. Следует отметить, что части речи не всегда определяются верно, допущенные ошибки могут повлиять на общую тональность: *kind* может иметь нейтральную тональность, поскольку является существительным в значении «вид», но программа классифицировала слово как положительное, подразумевая значение прилагательного «добрый». Несмотря на то, что программа определяет и слова, имеющие эмоциональную окраску, и объект, к которому они относятся, тональные цепочки не рассматриваются как единое целое, каждое слово оценивается отдельно, что тоже ведет к ошибкам, например, конфликтосодержащие существительные *corruption* и *sabotage* обозначены красным цветом, что означает негативную оценку, однако они сочетаются с глаголом *deal*

with, поэтому оценка должна быть противоположной; глагол *raise* оценивается как положительный, но в сочетании с существительным *question* тональность, скорее, нейтральная; числительное «много» нередко определяется как имеющее положительный вес, однако не всегда увеличение количества или изобилие чего-либо будет иметь позитивный sentiment. Прилагательное *sharp* в сочетании с существительным *wit* будет иметь положительный окрас; если рассматривать прилагательное *large* в тональной цепочке *large scale war effort*, то тональность будет явно отрицательной. Условные предложения не распознаются системой, а классифицируются как обычные, что является серьезной ошибкой, поскольку, употребляя глаголы в сослагательном наклонении, пользователи выражают желания по поводу сюжетов фильмов или купленных товаров, а не говорят о реальных событиях. По статистике условные предложения в английском языке занимают 8 % всего текста. Можно предположить, что причиной ошибок является или малое количество таких предложений в обучающей выборке, или их отсутствие выражает sentiment.

3. *Pattern*

Работа программы осуществляется при помощи созданных экспертом правил, имеющих вид «если А, то Б». Текст разбивается на слова или последовательности слов. Полученные данные используются для выделения часто встречающихся шаблонов, которым присваивается положительная или отрицательная оценка. Создание правил является достаточно трудоемкой задачей, поскольку система должна содержать правило для каждой комбинации тонально окрашенных слов, находящихся в тональной библиотеке и для каждого типа предложений в тексте. Системы с таким подходом эффективны только для текстов конкретной тематики, перенос алгоритма на незнакомые документы не даст высоких результатов. Помимо правил готовятся тональные лексиконы, представляющие собой списки позитивных и негативных лексем. Программа соотносит каждую единицу входного текста с имеющимися единицами в словарях для определения веса (см. Табл. 3).

Таблица 3 / Table 3

Результаты тестирования программы *Pattern* / Test results of *Pattern*

Метрики классификации / Rating Metrics	Кинорецензии / Film Reviews	Отзывы Amazon / Amazon Reviews
Точность определения положительных отзывов / Accuracy of positive feedback	0,57	0,87
Точность определения отрицательных отзывов / Accuracy of negative feedback	0,69	0,38
Правильность классификации / Correct classification	60/100	61/100

Источник: составлено О.И. Максименко, М.В. Беяковым. / Source: compiled by Olga I. Maksimenko, & Mikhail V. Belyakov.

Поскольку в результате анализа выдается только ответ, к какому классу принадлежит текст, то, в связи с отсутствием доступа к ходу работы программы, анализу были подвергнуты рецензии и отзывы, которые были неверно классифицированы.

В нескольких текстах, использовавшихся при анализе работы системы, присутствуют окказионализмы, которых нет в тональном словаре, поэтому программа их пропускает во время разметки, присуждая нулевой вес, среди них: *“A woman-behind-the-man”*, *“We see the evil mad scientist Dr. Krieger played by Udo Kier”*, *“making Genetically-Mutated-soldiers”*, *“to bring a ton of look-alike creepy-woman based horrors”*, *“indulgence in long-takes-without-cutaways”*, *“Lifetime for Women with nothing better to do” movie*, *“I mean it lacked all dramaticness”*.

Причинами могут быть несоблюдение грамматических форм, сокращения, возможные опечатки, бленды разных типов, включая написание слов с использованием чисел. Программа не распознает такие фрагменты, поскольку правила не справляются с потоком неструктурированной речи. Например:

“Three names that actually have made them selfs pretty big in the movie biz” (them selfs = themselves; biz = business);

“c’mom, the entire mission gathers to see this elder sent home” (c’mom = come on);

“I really like Salman Kahn so I was really disappointed when I seen this movie”;

“It’s look like the DVD is due a Region 2 release soon” (2 вместо «to»);

“And then I’m comin’ upstairs” (comin’ = coming);

“And how you can be such a busybody and talk about ppl, when ur own life isn’t sorted out” (ppl = people; ur = your)

Субстандартная лексика, а именно ненормативные и неформальные выражения, инвективы, которые, вероятнее всего, не вошли в тональный лексикон, также оказывают влияние на классификацию: *“that’s poop for you simpleton”*; *“«take the a**-whupping they deserve”*, *“against the notorious mutated hillbillies on their last day”*, *“Emmy has me absolutely gob-smacked”*, *“themes from the chum bucket”*, *“deep philosophy is bull*****”*.

Большинство отзывов, в которых встречаются условные предложения, также классифицированы неверно. Можно предположить, что в программе отсутствует правило, указывающее, как следует анализировать данный вид предложений. R. Narayanan [16] утверждает, что для такого вида предложений должны быть созданы специальные правила, поскольку условные предложения могут содержать эмоционально окрашенные слова, но не во всех случаях они выражают сентимент:

“If only Ed Wood could bring his subtle sense of flair and dignity to these remarkable scripts”;

“If it were longer, it might make more sense, but it still wouldn’t be much good”;

“I think if the movie did this it could have been very funny since both actresses are quite funny in their own ways and sitting here I can think of numerous scenarios that would have been a riot”;

“I suppose if you were going to sit in the same room and have line-of-sight with the device, lamp, etc. plugged into this control, it might work. Maybe”

Рецензии и отзывы с риторическими вопросами также представляют сложности для системы. Вопросительные предложения как способ выражения автором оценок не были достаточно изучены в области тонального анализа. По всей видимости, в программе отсутствует правило, которое бы «указывало» системе, как классифицировать такие предложения. Например:

“How can one recognize such brilliance and then see fit to replace it with such mediocrity?”;

“Canaanites wandered the earth...really? What were the scriptwriters thinking?”;

“If this book is so great, then why are there so many people who give it only one or two stars?”

Необходимость составления предметно-ориентированных словарей получила подтверждение и в ходе тестирования данной системы. Например, в тексте одной из проанализированных рецензий было следующее предложение: *“I found myself looking at my watch more and more as the film went on”*. В данном предложении отсутствуют слова с негативной коннотацией, однако тот факт, что во время просмотра фильма зритель все «больше и больше смотрел на часы», свидетельствует о том, что он дает негативную оценку. Можно допустить, что включение в тональный словарь выражения *look at watch* и создание правила, маркирующего, что появление данного словосочетания в контексте просмотра пользователем фильма дает отрицательную тональность, позволит программе верно классифицировать рецензию. В случае с выборкой Amazon в отзывах также часто встречаются оценки, не содержащие негативной лексики, однако вывод о тональности можно сделать по объективной информации, изложенной пользователем. Например, *“As directed, I downloaded all of the files from the SE site (70MB on dial up! It is too much!), and then downloaded all of the user guides”* — автор пишет о большом объеме данных, что определенно не является положительной характеристикой объекта. Тем не менее, система не классифицирует данный отзыв как негативный, поскольку в ее базе нет предметно-ориентированного словаря и релевантного для данной предметной области набора правил.

“This film laboured along with some of the most predictable story lines and shallow characters ever seen. The writer obviously bought the playbook “How to write a space disaster movie” and followed it play by play”. Прилагательные *predictable* и *shallow* в контексте отзывов о фильмах будут иметь негативный сентимент, что требует отдельно внести их в тональный словарь данной тематики, поскольку в других контекстах прилагательные могут быть нейтральными.

4. *ParallelDots* представляет собой пакет прикладных исследований с использованием искусственного интеллекта, в котором задача сентимент анализа является одним из включенных в пакет инструментов. Долгая краткосрочная память (long short-term memory, LSTM) считается особой разновидностью архитектуры рекуррентных нейронных сетей, на основе которых осуществляется работа системы. Превосходство рекуррентных нейросетей над другими видами заключается в использовании нейронами скрытого слоя предыдущих состояний сети для вычисления текущего (см. Табл. 4). Таким образом, полученная ранее информация о тональности какого-либо текстового фрагмента может повлиять на тональность анализируемого текста в данный момент. «LSTM сети были разработаны для решения проблемы долговременных зависимостей. Запоминание информации на продолжительный срок — это одна из основных особенностей этих сетей, не требующая продолжительного обучения»⁴

Таблица 4 / Table 4

Результаты тестирования программы *ParallelDots* / Test results of *ParallelDots*

Метрики классификации / Rating Metrics	Кинорецензии / Film Reviews	Отзывы Amazon / Amazon Reviews
Точность определения положительных отзывов / Accuracy of positive feedback	0,56	0,82
Точность определения отрицательных отзывов / Accuracy of negative feedback	0,53	0,69
Правильность классификации / Correct classification	54/100	68/100

Источник: составлено О.И. Максименко, М.В. Беляковым. / Source: compiled by Olga I. Maksimenko, & Mikhail V. Belyakov.

Возможные причины ошибок:

Определение тональности с использованием рекуррентной нейронной сети LSTM показывает лучшие результаты по сравнению с другими нейросетями, но достигаемая точность классификации существенно зависит от того, насколько корпус, на котором обучается модель, близок по составу, теме и стилю текста к анализируемому корпусу, в данном случае — кинорецензий или отзывов о товарах с Amazon [17]. По результатам анализа трех метрик можно сказать, что тренировочная выборка, которая использовалась для обучения, существенно отличается от текстов кинорецензий и более приближена к выборке отзывов с портала Amazon. Кроме того, эффективность работы зависит от количества нейронов скрытого и входных слоев. Например, в работе А. Кузнецова и Д. Кочурова⁵ экспериментально доказано, что увеличение нейронов скрытого слоя приводит к ухудшению работы тестовой выборки,

⁴Режим доступа: <https://digiratory.ru/575> (дата обращения: 15.12.2024).

⁵Режим доступа: <https://scienceforum.ru/2018/article/2018009449> (дата обращения: 15.12.2024).

а увеличение нейронов входного слоя, наоборот, улучшает. Однако система является закрытой, в свободном доступе присутствует демоверсия, информация об алгоритме работы не представлена. Известно, что системы, использующие нейросетевой метод работы, способны обучаться. Чем больше тестовая выборка, и чем больше итераций было произведено, тем качественнее работа.

5. Repustate

Программы, в основе которых лежит машинное обучение, могут демонстрировать высокие результаты, однако не всегда достигается правильная классификация. В совместной работе греческих и швейцарских исследователей [18] была выдвинута гипотеза о том, что машинное обучение, дополненное подходом, основанным на правилах, улучшит результаты анализа, поскольку помимо алгоритма, обучающегося на тренировочной выборке данных, правила будут «направлять» систему⁶. Однако стоит отметить, что результаты напрямую зависят от сходства текстов, подаваемых системе на вход, и текстов, которые использовались в качестве обучающей выборки (см. Табл. 5). Обновленная платформа предлагает следующие услуги: сентимент-анализ, извлечение данных (Text Mining), мультилингвальный подход⁷.

Таблица 5 / Table 5

Результаты тестирования программы Repustate / Test results of Repustate

Метрики классификации / Rating Metrics	Кинорецензии / Film Reviews	Отзывы Amazon / Amazon Reviews
Точность определения положительных отзывов / Accuracy of positive feedback	0,72	0,9
Точность определения отрицательных отзывов / Accuracy of negative feedback	0,68	0,83
Правильность классификации / Correct classification	70/100	84/100

Источник: составлено О.И. Максименко, М.В. Беляковым. / Source: compiled by Olga I. Maksimenko, & Mikhail V. Belyakov.

Возможные причины ошибок:

Как отмечалось ранее при анализе систем, использующих машинное обучение или правила в качестве подхода тональной классификации, причинами ошибок являются, прежде всего, отличие тестируемых текстов от обучающей выборки. В документации указано, что программа отлично справляется с эмоджиками и разного рода аббревиатурами, используемыми в интернете, следовательно, можно предположить, что тестирование проходило

⁶ Режим доступа: <https://www.zora.uzh.ch/id/eprint/39614/1/lrec2010.pdf> (дата обращения: 15.12.2024).

⁷ Режим доступа: <https://www.siachen.com/repustate/> (дата обращения: 15.12.2024).

на неформальных интернет-текстах, составленных пользователями, однако многое зависит от тематики текстов. При подаче на вход системы незнакомой лексики или синтаксических структур возникают ошибки, поскольку программа их не распознает и пропускает. Кроме того, создание правил требуется для каждого тематического блока текстов, так как пользовательский контент довольно разнообразен. То, что можно применить для комментариев в социальных сетях, не даст точных результатов при анализе кинокритик или отзывов. Более того, если рассматривать кинокритик, составленные 10 лет назад, и отзывы, собранные за 2024 г., а также правила для их анализа, логично предположить, что для данных выборок необходим разный набор правил ввиду постоянного изменения языка.

По результатам тестирования пяти программ, использующих разные методы работы, на двух размеченных по бинарному принципу выборках — кинокритик с сайта IMDb и отзывы пользователей на товары портала Amazon — получены следующие результаты:

- Sentistrength

Недостатки: устаревшие тональные лексиконы; неправильное определение частей речи; анализ без опоры на контекст; определение веса каждой единицы независимо от степени важности лексем; частичное игнорирование отрицательных слов, являющихся модификаторами.

Достоинства: возможность использования собственного словаря или существующего размеченного лексикона, который прошел предобработку.

- SummarizeBot

Недостатки: неверная классификация условных и сложносочиненных предложений с противительным союзом «но», учет модификаторов полярности только в непосредственной близости; ошибки при классификации устойчивых выражений; неверное определение частей речи; разрыв тональных цепочек.

Достоинства: определение объекта, о котором выражается мнение; подробное представление анализа (раздельное выделение цветом положительных и отрицательных лексем), способность работать с неструктурированными данными.

- Pattern

Недостатки: нет развернутого анализа; нет информации о используемом лексиконе и о правилах, встроенных в систему.

Достоинства: наряду с определением тональности система дает сведения о уровне субъективности текста.

- Repustate

Недостатки: отсутствие развернутого анализа.

Достоинства: сочетание двух методов классификации (правила и машинное обучение); возможность проведения анализа на 20 языках; обновление системы и обучение на новых данных.

- ParallelDots

Недостатки: отсутствие развернутого анализа; низкая точность определения отзывов.

Достоинства: широкий спектр текстов, на которых проходило обучение; использование усовершенствованного варианта архитектуры рекуррентных нейронных сетей.

Результаты сравнительного анализа приводятся в Таблице 6.

Таблица 6 / Table 6

Сравнительная таблица с результатами анализа программ /
Comparative table of programme analysis

Программа / Programme	Метод работы / Method of operation	Точность определения положительных отзывов / Accuracy of positive feedback		Точность определения отрицательных отзывов / Accuracy of negative feedback		Правильность классификации / Correct classification	
		IMDb	Amazon	IMDb	Amazon	IMDb	Amazon
Summarize Bot	машинное обучение machine / learning	0,79	0,72	0,75	0,78	77	75
Senti-strength	словарный / vocabulary	0,68	0,65	0,77	0,7	70	60
Repustate	гибридный метод / hybrid method	0,72	0,9	0,68	0,83	70	84
Pattern	правила / rule	0,57	0,87	0,69	0,38	60	61
Parallel-Dots	рекуррентные нейросети / recursive neural networks	0,56	0,82	0,53	0,69	54	68

Источник: составлено О.И. Максименко, М.В. Беляковым. / Source: compiled by Olga I. Maksimenko, & Mikhail V. Belyakov.

Определяющим фактором при тональной классификации является степень схожести лексических, синтаксических и стилистических особенностей анализируемого текста с данными, которые заложены в систему. Независимо от подхода (словарного, на основе правил, с использованием машинного обучения или гибридного) программа не даст высоких результатов, если данные, подающиеся на вход, неизвестны системе. Следовательно, необходимо настраивать программу на определенную предметную область: создать словари, включающие оценочные слова и словосочетания из текстов предполагаемого тестирования. Создание подобных словарей входит в задачи

лингвоэмотиологии. Также необходимым условием является составление правил на основе выделения особенностей предварительно проанализированного текста. С программами, обучаемыми на эталонном корпусе, задача представляется более сложной, поскольку для эффективной работы требуется большой размеченный корпус. В свободном доступе имеются готовые аннотированные данные, однако далеко не для всех предметных областей.

Таким образом, можно сделать вывод, что подбирать программу для автоматической тональной классификации необходимо, в первую очередь, исходя из текстов, которые предполагается анализировать. Необходимо предварительно изучить документацию систем с тем, чтобы:

- понять на каких текстах проходило обучение, если в основе машинное обучение;
- проанализировать словарь, который используется системой, если работа базируется на словарном подходе;
- изучить правила, если информация о них имеется в открытом доступе.

Комментарии, отзывы, сообщения, которые представлены в интернете, меняются, как и язык в целом, даже иконические компоненты такие, как эмодзи могут со временем выглядеть иначе. Необходимо постоянное обновление тональных лексиконов и обучающих выборок. Как показал анализ системы Sentistrength, словари которой обновляются нерегулярно, отсутствие новых данных ведет к ошибкам в классификации. Для сравнения, программа Repustate была обновлена в 2024 г., разработчики провели переобучение алгоритмов на новых данных, в результате система показала наиболее точный анализ.

В ходе анализа среди проблем, которые препятствовали верной классификации, были выделены следующие:

1. Неоднозначное мнение и сравнение. Некоторые пользователи перечисляют ряд недостатков, которые их разочаровали в фильме, но в конце отзыва пишут, что в целом они довольны. Иногда, напротив, выражается восхищение режиссером, другими фильмами, которые были им сняты, но делается заключение, что просмотренный фильм не имеет ничего общего с предыдущими работами. Такая же тенденция просматривается и в случае с отзывами. Системе, оценивающей каждое предложение в тексте, на основе чего далее делается заключительное решение о тональности всего документа, работать с такими данными сложно. V. Aharonson и I. Becker [19] по результатам своего исследования предложили классифицировать тональность всего документа, принимая во внимание только последнее предложение, поскольку чаще всего именно в конце текста пользователи выражают окончательное мнение. Введение такого правила поможет решить проблему с подобными отзывами, но стоит отметить, что не каждая рецензия и отзыв строятся по такому принципу.

2. Вариантом решения данной проблемы может быть разделение текстов на две группы и их отдельный анализ. Разные типы предложений требуют разных подходов при классификации. В проанализированных текстах наибольшую сложность вызвали условные и сложносочиненные предложения с противительным союзом.
3. Необходимо рассматривать слова в их окружении, а не каждую единицу отдельно. Специфика работы словарного подхода заключается в оценке всех слов (исключения составляют стоп-слова, к которым относятся, в первую очередь, служебные части речи), что некорректно. Простая разметка текстов путем сравнения каждой единицы с включенными в тональный лексикон словами — неэффективный способ классификации, подключение правил, вероятно, позволит значительно улучшить результат.
4. При тестировании анализировались все единицы, в том числе и имена собственные (именованные сущности), что неверно. Решение проблемы возможно путем создания правил тестирования.

В.И. Шаховский, формулируя основные положения лингвистической теории эмоций в работе «Категоризация эмоций в лексико-семантической системе языка», писал: «есть мир (объект), есть человек (субъект), способный отражать этот мир. Отражение мира человеком включает только то, что имеет для него значение в настоящий момент или составляет ценность» [20. С. 31–32]. Учитывая объем ежесекундно поступающей информации, имеющей в том числе и оценочную функцию, пакеты сентимент-анализа позволяют существенно ускорить процесс определения тональности входных текстов.

Заключение

Таким образом, отвечая на главный вопрос о наиболее эффективном методе определения тональности, необходимом оценочном средстве для лингвоэмотиологии и лингвоконфликтологии, отметим, что программа, использующая гибридный метод, лучше остальных справилась с задачей. Однако такие корректировки, как введение нескольких правил, подбор тематической обучающей выборки могут дать возможность значительного улучшения результатов.

Список литературы

1. Шаховский В.И. Категоризация эмоций в лексико-семантической системе языка. Воронеж : Воронеж. гос. ун-т, 1987.
2. Шаховский В.И. Лингвистическая теория эмоций: монография. М. : Гнозис, 2008.
3. Беляков М.В. Лингвоаксиология и лингвoseмиотика дипломатического дискурса (на материале открытой профессиональной дипломатии): дисс. ... д. фил. наук. М., 2022.
4. Беляков М.В. Анализ новостных сообщений сайта МИД РФ методом сентимент-анализа (ст. 2) // Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. 2016. Т. 5. № 4. С. 115–124. EDN: WZZKRZ

5. Максименко О.И. Анализ тональности текстов (сентимент-анализ) на материале текстов СМИ // IV Новиковские чтения: Функциональная семантика и семиотика знаковых систем. Сб. научных статей. 2012. Ч. 1. М. : РУДН. С. 96–105. EDN: WZZKRZ
6. Introduction to Sentiment Analysis: What is sentiment analysis, 2018. Режим доступа: <https://algorithmia.com/blog/introduction-sentiment-analysis> (дата обращения: 10.01.2025).
7. Agrawal R., Gupta N. Extracting knowledge from opinion mining // *Advances in Data Mining and Database Management*. Hershey: IGI Global, 2018. <https://doi.org/10.4018/978-1-5225-6117-0>
8. Юрганов А.А. Сентимент анализа как инструмент исследования текстов // *Проблемы современной науки и образования*. 2017. № 29(111). С. 39–41. EDN: ZDMSVF
9. Pang B., Lee L. Opining mining and sentiment analysis // *Foundations and Trends in Information Retrieval*. 2008. Vol. 2. № 1–2. P. 1–135.
10. Шаховский В.И. Эмоции как объект исследования в лингвистике // *Вопросы психолингвистики*. 2009. № 9. С. 29–43. EDN: LAULHF
11. Колмогорова А.В., Калинин А.А., Маликова А.В. Кто и о чем говорит в «радостных» и «грустных» текстах: в поисках дискриминантных черт текстов разных эмоциональных тональностей // *Известия Уральского федерального университета. Серия 2. Гуманитарные науки*. 2019. Т. 21. № 4(193). С. 219–234. <https://doi.org/10.15826/izv2.2019.21.4.078> EDN: JMPVSS
12. Колмогорова А.В., Вдовина Л.А. Лексико-грамматические маркеры эмоций как параметры для сентимент анализа русскоязычных интернет-текстов // *Вестник Пермского университета. Российская и зарубежная филология*. 2019. Т. 11. № 3. С. 38–46. <https://doi.org/10.17072/2073-6681-2019-3-38-46> EDN: VOGAAV
13. Колмогорова А.В. Вербальные маркеры эмоций в контексте решения задач сентимент-анализа // *Вопросы когнитивной лингвистики*. 2018. № 1(54). С. 83–93. <https://doi.org/10.20916/1812-3228-2018-1-83-93> EDN: QIWKPN
14. Семина Т.А. Дихотомия субъективность vs. объективность и тональная релевантность в задачах анализа тональности // *Вестник Московского государственного областного университета. Серия: Лингвистика*. 2018. № 1. С. 38–45. <https://doi.org/10.18384/2310-712X-2018-1-38-45> EDN: YRNQNH
15. Scheible C., Schutze H. Sentiment relevance // *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. Sofia, 2013. P. 954–963.
16. Narayanan R., Liu B., Choudhary A. Sentiment analysis of conditional sentences // *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*. Singapore: ACL and AFNLP, 2009. P. 180–189. Режим доступа: https://www.researchgate.net/publication/221012837_Sentiment_Analysis_of_Conditional_Sentences (дата обращения: 10.01.2025). <https://doi.org/10.3115/1699510.1699534>
17. Сбоев А.Г., Воронина И.Е. и др. Продвинутое нейросетевые модели для решения задачи определения тональности // *Вестник Воронежского государственного университета. Сер. Системный анализ и информационные технологии*. 2016. № 4. С. 178–183. EDN: XVSYZX
18. Rentoumi V., Petrakis S., Klenner M. et al. United we stand: improving sentiment analysis by joining machine learning and rule based methods // *7th International Conference on Language Resources and Evaluation (LREC 2010)*, 19 Mai 2010–21 Mai 2010. Malta : Zurich Open Repository and Archive, University of Zurich. P. 954–963.
19. Becker I., Aharonson V. Last but definitely not least: on the role of the last sentence in automatic polarity-classification // *Proceedings of the ACL 2010 Conference Short Papers*. Uppsala, 2010. P. 331–335. Режим доступа: https://www.researchgate.net/publication/220873476_Last_but_Definitely_Not_Least_On_the_Role_of_the_Last_Sentence_in_Automatic_Polarity-Classification, (дата обращения: 10.01.2025).
20. Шаховский В.И. Обоснование лингвистической теории эмоций // *Вопросы психолингвистики*. 2019. № 1(39). С. 22–37. <https://doi.org/10.30982/2077-5911-2019-39-1-22-37> EDN: SKQDYO

References

1. Shakhovsky, V.I. (1987). *Categorization of Emotions in the Lexico-Semantic System of Language*. Voronezh: Voronezh State University publ. (In Russ.).
2. Shakhovsky, V.I. (2008). *Linguistic Theory of Emotions*. Moscow: Gnosis. (In Russ.).
3. Belyakov, M.V. (2022). *Linguoaxiology and Linguosemiotics of Diplomatic discourse (based on the Material of Open Professional Diplomacy)* [Doctoral Thesis]. Moscow. (In Russ.).
4. Belyakov, M.V. (2016). Analysis of News Reports on the Website of the Ministry of Foreign Affairs of the Russian Federation by the Method of Sentiment Analysis (Art. 2). *RUDN Journal of Language Studies, Semiotics and Semantics*, 6(4), 115–124. (In Russ.). EDN: WZZKRZ
5. Maksimenko, O.I. (2012). Analysis of the Tonality of Texts (Sentiment Analysis) Based on Media Texts. In: *Proceedings of IV Novikov readings: Functional Semantics and Semiotics of Sign Systems. Part 1* (pp. 96–105). Moscow: RUDN University publ. (In Russ.).
6. Introduction to Sentiment Analysis: What is Sentiment Analysis (2018). URL: <https://algorithmia.com/blog/introduction-sentiment-analysis> (accessed: 10.01.2025).
7. Agrawal, R., & Gupta, N. (2018). Extracting Knowledge from Opinion Mining. In: *Advances in Data Mining and Database Management*. Hershey: IGI Global. <https://doi.org/10.4018/978-1-5225-6117-0>
8. Yurganov, A.A. (2017). Sentiment Analysis as a Tool for Text Research. *Problems of Modern Science and Education*, 29(111), 39–41. (In Russ.). EDN: ZDMSVF
9. Pang, B., & Lee, L. (2008). Opining Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135.
10. Shakhovsky, V.I. (2009). Emotions as an Object of Research in Linguistics. *Journal of Psycholinguistics*, 9, 29–43. (In Russ.). EDN: LAULHF
11. Kolmogorova, A.V., Kalinin, A.A., & Malikova, A.V. (2019). Who Speaks about What in “Joyful” and “Sad” Texts: in Search of Discriminating Features of Texts of Different Emotional Tones. *Proceedings of the Ural Federal University. Series 2, Humanities*, 21(4(193)), 219–234. (In Russ.). <https://doi.org/10.15826/izv2.2019.21.4.078> EDN: JMPVSS
12. Kolmogorova, A.V., & Vdovina, L.A. (2019). Lexico-Grammatical Markers of Emotions as Parameters for The Sentimental Analysis of Russian-Language Internet Texts. *Bulletin of the Perm University. Russian and Foreign Philology*, 11(3), 38–46. (In Russ.). <https://doi.org/10.17072/2073-6681-2019-3-38-46> EDN: VOGAAV
13. Kolmogorova, A.V. (2018). Verbal Markers of Emotions in the Context of Solving Problems of Sentiment Analysis. *Questions of Cognitive Linguistics, Issues of Cognitive Linguistics*, 1(54), 83–93. (In Russ.). <https://doi.org/10.20916/1812-3228-2018-1-83-93> EDN: QIWKPN
14. Semina, T.A. (2018). Dichotomy Subjectivity vs. Objectivity and Tonal Relevance in the Tasks of Tonality Analysis. *Bulletin of the Moscow State Regional University. Series: Linguistics*, 1, 38–45. (In Russ.). <https://doi.org/10.18384/2310-712X-2018-1-38-45> EDN: YRNQNH
15. Scheible, C., & Schutze, H. (2013). Sentiment Relevance. In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, August 4–9*. (pp. 954–963). Sofia, Bulgaria: Association for Computational Linguistics.
16. Narayanan, R., Liu, B., & Choudhary, A. (2009). Sentiment Analysis of Conditional Sentences. In: *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (pp. 180–189). Singapore: ACL and AFNLP. URL: https://www.researchgate.net/publication/221012837_Sentiment_Analysis_of_Conditional_Sentences (accessed: 10.01.2025). <https://doi.org/10.3115/1699510.1699534>
17. Failures, A.G., Voronina, I.E., et al. (2016). Advanced Neural Network Models for Solving the Problem of Determining Tonality. *Proceedings of Voronezh State University. Series: Systems Analysis and Information Technologies*, 4, 178–183. EDN: XVSYZX
18. Rentoumi, V., Petrakis, S., Klenner, M. et al. (2010). United We Stand: Improving Sentiment Analysis by Joining Machine Learning and Rule Based Methods. In: *7th International Conference on Language Resources and Evaluation (LREC 2010)*, 19–21 May 2010 (pp. 954–963). Malta: Zurich Open Repository and Archive, University of Zurich.

19. Becker, I., & Aharonson, V. (2010). Last but definitely not least: on the role of the last sentence in automatic polarity-classification. In: *Proceedings of the ACL 2010 Conference Short Papers 11–16 July 2010* (pp. 331–335). Sweden: Uppsala, Association for Computational Linguistics URL: https://www.researchgate.net/publication/220873476_Last_but_Definitely_Not_Least_On_the_Role_of_the_Last_Sentence_in_Automatic_Polarity-Classification (accessed: 10.01.2025).
20. Shakhovsky, V.I. (2019). Substantiation of the Linguistic Theory of Emotions. *Journal of psycholinguistics*, 1(39), 22–37. (In Russ.). <https://doi.org/10.30982/2077-5911-2019-39-1-22-37> EDN: SKQDYO

Сведения об авторах:

Максименко Ольга Ивановна, доктор филологических наук, профессор, профессор кафедры теории языка, англистики и прикладной лингвистики лингвистического факультета, Государственный университет просвещения (105005, Российская Федерация, г. Москва, ул. Радио, д. 10а, стр. 2); *научные интересы*: автоматизированный анализ текстовых массивов, лингвосемиотика, лингвистическая семантика, теория интерпретации; *e-mail*: maxbel7@yandex.ru

ORCID: 0000-0002-6611-8744. SPIN-код: 7708-5901, AuthorID: 374818.

Беляков Михаил Васильевич, доктор филологических наук, доцент, профессор кафедры русского языка МГИМО (университет) МИД России (119454, Российская Федерация, г. Москва, проспект Вернадского, д. 76); *сфера научных интересов*: лингвосемиотика, дискурс-анализ, компьютерная лингвистика, семантические и когнитивные исследования языка; *e-mail*: m.belyakov@my.mgimo.ru

ORCID: 0000-0002-6230-9893. SPIN-код: 1761-5400, AuthorID: 474699.

Information about the authors:

Olga I. Maksimenko, Dr.Sc. (Philology), Full Professor, Professor of the Language Theory, Anglistics and Applied Linguistics Department, Linguistic Faculty, State University of Education (10a built 2, Radio street, Moscow, Russian Federation, 105005); *Research interests*: automatic text analysis, linguosemiotics, linguistic semantics, interpretation theory; *e-mail*: maxbel7@yandex.ru

ORCID: 0000-0002-6611-8744. SPIN-code: 7708-5901, AuthorID: 374818.

Mikhail V. Belyakov, Dr.Sc. (Philology), Associated Professor, Professor at Russian Department, Moscow State Institute of International Relations (University) (76, Vernadskogo av., Moscow, Russian Federation, 119454); *Research interests*: linguosemiotics, discourse analysis, computer science, semantic and cognitive language studies; *e-mail*: m.belyakov@my.mgimo.ru

ORCID: 0000-0002-6230-9893. SPIN-code: 1761-5400, AuthorID: 474699.