
**ПРИКЛАДНАЯ И КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА,
АВТОМАТИЧЕСКАЯ ОБРАБОТКА ТЕКСТА
APPLIED AND COMPUTATIONAL LINGUISTICS,
AUTOMATIC TEXT PROCESSING**

*Вестник Челябинского государственного университета. 2026. № 5 (511). С. 110–118.
Bulletin of Chelyabinsk State University. 2026;(5(511)):110-118.*

Научная статья

УДК 81.33 + 811.161.1

DOI: 10.47475/1994-2796-2026-511-5-110-118

КОНТЕКСТНЫЕ ФАКТОРЫ В АНАЛИЗЕ ТОНАЛЬНОСТИ

Анастасия Юрьевна Зиновьева¹, Светлана Олеговна Шереметьева²

¹Южно-Уральский государственный университет (национальный исследовательский университет), Челябинск, Россия, zinovevaai@susu.ru, 0000-0002-7658-7376, Scopus ID: 57211475130

²Южно-Уральский государственный университет (национальный исследовательский университет), Челябинск, Россия, sheremetevaso@susu.ru, 0000-0003-1245-4213, Scopus ID: 6603648623

Аннотация. В статье рассмотрены лингвистические факторы (лексемы-модификаторы), в контексте которых возможны нейтрализация, ослабление, усиление, а также полное изменение знака контекстно-независимой тональности лексем, что значительно осложняет автоматизированное вычисление тональности высказываний в целом. Цель исследования — выявление контекстных факторов при многоаспектном анализе тональности на основе словаря тональной лексики (е-лексикона), определение лексических единиц, посредством которых данные факторы реализуются в тексте, а также увеличение за счет этих единиц покрываемости аннотационного словаря. В исследовании использованы методы корпусного анализа, лексико-онтологического моделирования и контекстного анализа. В результате исследования выделены классы специфических и неспецифических контекстных факторов. К числу первых относятся ирреалис, выражения иронии и сарказма, тонально окрашенная лексика, используемая не по отношению к объекту оценки, и оценочные слова в тонально-нейтральных многокомпонентных лексических группах; к числу вторых — остальные модификаторы. Исследованы лексические воплощения контекстных факторов каждого из классов и условия реализации их конкретного модифицирующего влияния на контекстно-независимую тональность лексем. Выделенные из корпуса глоссы контекстных факторов внесены в е-лексикон платформы аннотирования с соответствующей маркировкой, кодирующей онтологические концепты, которые отражают их модифицирующие значения. Исследование выполнено в рамках проекта по автоматизированному аспектно-тональному анализу, обеспечивает достаточно широкую покрываемость е-лексикона, что в свою очередь повышает корректность автоматического аннотирования и позволяет на следующих этапах проекта разработать формальные правила для автоматизированной модификации контекстно-независимой тональности лексем. Исследование проведено на материале русскоязычного корпуса отзывов об услугах медицинских центров Челябинска.

Ключевые слова: анализ тональности, контекстные факторы, автоматическое аннотирование, корпус отзывов, медицинские услуги, русский язык

Для цитирования: Зиновьева А. Ю., Шереметьева С. О. Контекстные факторы в анализе тональности // Вестник Челябинского государственного университета. 2026. № 5 (511). С. 110–118. DOI: 10.47475/1994-2796-2026-511-5-110-118

Original article

CONTEXTUAL FACTORS IN SENTIMENT ANALYSIS

Anastasia Yu. Zinoveva¹, Svetlana O. Sheremetyeva²

¹South Ural State University (National Research University), Chelyabinsk, Russia, zinovevaai@susu.ru, 0000-0002-7658-7376, Scopus ID: 57211475130

²South Ural State University (National Research University), Chelyabinsk, Russia, sheremetvaso@susu.ru, 0000-0003-1245-4213, Scopus ID: 6603648623

Abstract. This article examines the linguistic factors (modifier lexemes), in the context of which the neutralization, weakening, strengthening, or complete sign change of the lexeme context-independent tonality are possible, which significantly complicates automated tonality calculation. The aim of the study is to identify contextual factors in multi-aspect sentiment analysis based on a dictionary of tonal vocabulary (e-lexicon), to determine lexical units by means of which these factors are realized in the text, and to increase the coverage of the annotation lexicon by means of these units. We use such methods as corpus analysis, lexico-ontological modeling and contextual analysis. We identify two classes of contextual factors: specific and non-specific. The first class encompasses irrealis moods, expressions of irony and sarcasm, evaluative vocabulary related to objects different from that of evaluation, and evaluative words found within tonally neutral multi-component phrases. The second class includes other types of modifiers. We analyze the lexical embodiments of each class of contextual factors and examine the conditions under which they exert their specific modifying influence on the lexeme context-independent tonality. The contextual factor glosses extracted from the corpus are put into an e-lexicon of an annotation platform with markings that encode ontological concepts reflecting their modifying meanings. This study is part of a project focused on automated aspect-based sentiment analysis, and it offers a sufficiently wide coverage of the e-lexicon, which, in turn, increases the correctness of automatic annotation and enables, at the next stages of the project, to develop formal rules for the automated modification of the lexeme context-independent tonality. The study was conducted using a corpus of Russian reviews concerning the services of medical centers in Chelyabinsk.

Keywords: sentiment analysis, contextual factors, automatic annotation, review corpus, medical services, Russian language

For citation: Zinoveva AYu, Sheremetyeva SO. Contextual Factors in Sentiment Analysis. *Bulletin of Chelyabinsk State University*. 2026;(5(511)):110-118. (In Russ.). DOI: 10.47475/1994-2796-2026-511-5-110-118

Введение

В последние двадцать лет одной из наиболее востребованных и динамично развивающихся областей автоматической обработки текста является анализ тональности (сентимент-анализ). Его цель состоит в определении, извлечении и систематизации оценок и оценочных суждений, представленных в постах в социальных сетях, блогах, интернет-отзывах о товарах и т. п. В зависимости от задачи исследования выделяют два основных вида анализа тональности: общий, в рамках которого определяют общую эмоциональную окраску целого текста или его части, и таргетированный, к которому относят, например, аспектный анализ, извлечение позиции пользователя по определенной теме, а также извлечение оценочных отношений [4].

В настоящее время выделяют три класса методов анализа тональности:

1) лексические (алгоритмические) методы, предполагающие использование словаря тональной лексики, единицам которого присваиваются веса, позволяющие модифицировать оценку [2–4; 11];

2) методы на основе машинного обучения, в рамках которых анализ тональности рассматри-

вается как задача классификации, решаемая обучением классификатора на аннотированном тональными метками корпусе текстов [6; 11; 14];

3) гибридные методы, комбинирующие машинное обучение с использованием словарей [1; 9; 10; 12].

Вне зависимости от используемого метода задача анализа тональности осложняется рядом факторов, которые могут быть разделены на две основные группы: лексические и контекстные. Первая группа включает факторы, связанные с многозначностью оценочной лексики или изменением исходной тональности слова или выражения в конкретной предметной области [4]. Проблема воздействия этих факторов может быть решена ориентацией модели анализа тональности на данную предметную область. Вторая группа представлена факторами, связанными с изменением исходной тональности слова или выражения в контексте использования [2; 4; 8; 13].

В группу контекстных факторов, по мнению разных авторов, могут быть отнесены слова, модифицирующие тональность соседних слов; оценочные слова в составе нейтральных

многокомпонентных лексических групп; ирреалис, «при котором сообщаемое в предложении не относится к тому, что реально происходит» [4]; идиомы [8]; противительные союзы, ирония [13] и др.

Проблема воздействия контекстных факторов на анализ тональности в основном решается с помощью словарей и лингвистических правил. Например, в работе [2] предлагается решение данной проблемы с использованием словарей операторов и стоп-выражений. Словарь операторов содержит такие слова, как *очень*, *значительно* и др., усиливающие положительную или отрицательную оценку соседнего выражения, а также слова *не*, *отменить* и др., изменяющие оценку выражения на противоположную, т. е. изменяющие ее полярность. Словарь стоп-выражений содержит многокомпонентные лексические группы, которые, сами являясь нейтральными, включают тональные компоненты. Применение лингвистических правил в совокупности со словарями также позволяет обнаружить ирреалис и снизить его влияние на вычисление тональности: в частности, снижается оценка клауз, начинающихся с союза *если* или содержащих частицы *бы*, *ли*. В работе [13] предлагается изменять веса тональных лексических единиц с учетом слов-модификаторов. Например, слова с отрицательной окраской инвертируют тональность соседних слов: *clever* +2 в совокупности с модификатором *not* дает *not clever* -2. При этом отрицательные слова выступают в качестве инверторов только в тех случаях, когда соседние с ними слова уже имеют какую-либо тональную окраску, что следует учитывать при создании правил изменения оценки. В этой же работе описаны и другие модификаторы, позволяющие интерпретировать пресуппозицию, иронию и другие факторы.

Настоящая работа направлена на выявление контекстных факторов при многоаспектном анализе тональности на основе словаря тональной лексики, определение лексических единиц, посредством которых данные факторы реализуются в тексте, а также увеличение за счет этих единиц покрываемости аннотационного словаря. Исследование выполнено на материале русскоязычных отзывов об услугах, оказываемых частными медицинскими центрами Челябинска.

Лингво-информационные ресурсы и методы исследования

При выполнении задач исследования использованы следующие предварительно созданные лингво-информационные ресурсы, в число которых входят:

1. Исходный корпус объемом 113 606 словоупотреблений, содержащий 1601 отзыв о восьми медицинских центрах Челябинска. Корпус был создан методом сплошной выборки с сайта электронного справочника 2GIS¹ по состоянию на март 2023 г.

2. Ориентированная на задачу аспектно-тонального анализа трехуровневая онтология предметной области отзывов об услугах медицинских учреждений, которая имеет следующую структуру:

ALL [ASPECT [Q CH EQ ... L] SENTIMENT [POS NEG POSX NEGX] MODIFIER [MAG ANT REV] OTHER],

где ALL — концепт первого уровня (корень онтологического дерева); четыре концепта второго уровня: ASPECT — аспект анализа, SENTIMENT — тональность, MODIFIER — модификатор тональности и OTHER, а также 25 концептов третьего уровня, представляющих собой подклассы концепта ASPECT (качество лечения (Q), компетентность врача (CH), техническое оснащение (EQ) и т. д.), подклассы концепта SENTIMENT (POS — положительная, POSX — крайне положительная, NEG — отрицательная, NEGX — крайне отрицательная тональность) и подклассы концепта MODIFIER (MAG — усиление, ANT — ослабление, REV — изменение контекстно-независимой тональности на противоположную).

3. Исходный онтолексикон — список релевантных для аспектно-тонального анализа лексем исходного корпуса, концептуальные значения которых соответствуют значениям концептов онтологии. Особенностью онтолексикона является то, что он содержит не только отдельные слова, как в большинстве проектов по аннотированию корпусов и анализу тональности, но и многокомпонентные лексические группы (именные, глагольные, предложные, наречные, группы прилагательных), а также, когда это практически целесообразно, клаузы, что позволяет снизить многозначность лексем и облегчить последующий анализ тональности.

4. Платформа концептуального аннотирования [7], состоящая из электронной лексико-онтологической базы знаний (е-лексикона) и концептуально-аспектного таггера.

5. Корпус с автоматической аспектно-тональной разметкой, выполненной на платформе концептуального аннотирования на основе знаний исходного е-лексикона аспектной и тональной

¹ URL: <https://2gis.ru/chelyabinsk> (дата обращения: 29.01.2025).

лексики предметной области отзывов об услугах медицинских учреждений.

6. Корпус с «золотой» аспектно-тональной разметкой [15], полученный посредством ручного постредактирования автоматически аннотированного корпуса на основе исходного е-лексикона.

В исследовании использованы методы корпусного анализа, лексико-онтологического моделирования и контекстного анализа, с помощью которых на основе сравнения исходного, автоматически аннотированного и «золотого» корпусов проведен глубокий анализ контекстных факторов (лексических модификаторов), с целью уточнения и увеличения онтолексиконов исходного е-лексикона, а также условий реализации их конкретного модифицирующего влияния на контекстно-независимую тональность лексем. Проведено исследование как узкого контекста в пределах одного-двух шагов влево и вправо от тональной лексической единицы, так и широкого контекста всего отзыва для более глубокого понимания воздействия контекстных факторов на определение тональности.

Результаты исследования и их обсуждение

В центре настоящего исследования находятся вербальные факторы, влияющие на контекстно-независимую тональность лексических единиц, встречающихся в их контексте. Глубокий сравнительный анализ функционирования лексем-модификаторов в исходном, автоматически аннотированном и «золотом» корпусах отзывов о медицинских услугах позволил выделить специфические и неспецифические контекстные факторы. К числу первых мы отнесли ирреалис, выражения иронии и сарказма, тонально окрашенную лексику, использующуюся не по отношению к объекту оценки, и оценочные слова в тонально-нейтральных многокомпонентных лексических группах; к числу вторых — остальные лексем-модификаторы контекстно-независимой тональности лексем исследуемой предметной области. При этом для фиксации значений контекстных факторов в случае появления их в узком контексте, как правило, использовано соотнесение их с концептами онтологической ветки MODIFIER, в то время как при реализации контекстных факторов в широком контексте их значение чаще всего описывалось концептами ветки SENTIMENT.

В первую очередь приведем результаты анализа неспецифических контекстных факторов, по-

¹ Термины «золотая разметка», «золотой корпус», «золотой стандарт» обозначают абсолютно корректную разметку корпуса.

скольку их частотность в корпусе намного выше частотности специфических. К неспецифическим контекстным факторам, как правило, относятся лексем-модификаторы, которые изменяют контекстно-независимую тональность лексем в узком контексте в пределах одного-двух шагов влево и вправо.

Например (здесь и далее орфография и пунктуация авторов сохранены), в высказывании «*Порядовала лишь девушка-администратор*» (1) слово *лишь* ослабляет положительную тональность, представленную словом *порядовала*. В то время как слово *всегда* в высказывании «*Медицинский центр Лотос всегда отличает профессионализм медицинского и административного персонала*» (2) усиливает положительную оценку многокомпонентной именной группы *профессионализм медицинского и административного персонала*. Изменение контекстно-независимой тональности лексем на противоположную можно проиллюстрировать на примере предложения «*До сегодняшнего дня претензий к качеству предоставляемых услуг не было*» (3), где фрагмент предложения *претензий к качеству предоставляемых услуг не было* вне контекста может быть интерпретирован как имеющий положительную тональность, однако предложная группа *до сегодняшнего дня* позволяет сделать вывод о его отрицательной тональности. Модификаторы тональности могут быть также обнаружены в более широком контексте, но в рамках сравнительно короткого предложения. Например, в высказывании «*Бывают недочеты со стороны работы администраторов, но это не критично*» (4). Здесь модификатор *не критично* расположен дистантно относительно модифицируемого им тонального слова *недочеты*. Это затрудняет обнаружение и учет влияния модификатора при автоматическом анализе тональности и, как следствие, потребует достаточно сложных правил коррекции тональности высказывания и ручного постредактирования аннотированного отзыва экспертом. Чтобы обеспечить учет такого рода модификаторов при автоматическом аннотировании корпуса, одно- и многокомпонентные лексем-модификаторы, функционирующие, как в примерах (1)–(4), включены в онтолексикон и соотнесены с концептами MAG, ANT и REV соответственно.

В отдельных случаях влияние лексем-модификаторов проявляется только в контексте аспектных лексических единиц. Например, в предложении «*Это не врач!*» (5) слово *врач* не является тонально окрашенным, это лексема

с концептуальным значением ‘компетентность врача’ (CH) — подклассом онтологического концепта ASPECT. В таких случаях в онтолексикон включается вся многокомпонентная единица, которая одновременно соотносится с двумя концептами онтологии: аспектом КОМПЕТЕНТНОСТЬ ВРАЧА (CH) и модификатором (NEGX).

Что касается специфических контекстных факторов, то в исследуемом корпусе нередко случаи использования ирреалиса, при котором автор не оценивает произошедшее событие, а выражает мнение (зачастую не свое) относительно потенциально возможного события. Например, в корпусе встречаются предложения вида (6)–(8) (жирным шрифтом выделены фрагменты текста, указывающие на ирреалис): «*Как в поликлинику хрен дозвонишься, а я думала, частная клиника высокого профиля*» (6), «*Где тут профессиональные врачи?*» (7), «*И это при том, что компания **позиционирует себя** как клиентоориентированная и работающая с обратной связью*» (8).

Выполнив анализ приведенных выше и других аналогичных предложений из корпуса, мы пришли к выводу, что в большинстве случаев ирреалис приводит к изменению контекстно-независимой тональности лексических единиц на противоположную, поэтому такие лексические индикаторы ирреалиса были добавлены в онтолексикон и соотнесены с концептом REV.

Еще одним способом фиксации тональности в ирреалисе, который мы использовали в нашем исследовании, является включение в онтолексикон модальных выражений и соотнесение их с соответствующим концептом онтологической ветки SENTIMENT. Например, в предложении «*Считаю если вы имеете смелость заявлять такие цены, то и отношение к людям, система **должна быть на высшем уровне***» (9) многокомпонентная лексема *должна быть на высшем уровне*, сигнализирующая об ирреалисе, указывает на отрицательное впечатление автора отзыва именно из-за модальности долженствования. Такие фразы целиком заносятся в онтолексикон и соотносятся с концептом NEG.

Отметим, что возможны случаи, когда влияние ирреалиса может быть выявлено только на основании анализа более широкого, чем предложение, контекста. Например, в корпусе встречается предложение «*Рекомендовали её как **первоклассного специалиста***» (10), в котором используется словосочетание *первоклассный специалист*, вне контекста имеющее положительную тональность. Однако анализ более широкого контекста

отзыва позволяет сделать вывод, что в данном случае автор весьма несогласен с такой оценкой врача, и тональность должна быть изменена на отрицательную. Тем не менее, остается вероятность, что в других отзывах такое выражение не будет иметь отрицательной окраски, и окончательное решение в каждом конкретном случае остается за экспертом. Поэтому такие лексемы в онтолексиконе связываются с двумя концептами NEG / NEGX и POS / POSX, подклассами онтологической ветки SENTIMENT, и в ходе автоматического аннотирования получают двойные метки вида ~NEG~POS. Для разрешения возникающей в таких случаях неоднозначности потребуется либо разработка формальных правил, либо привлечение эксперта-постредактора.

В отрицательных отзывах распространены выражения иронии и сарказма, учет которых при определении тональности текста представляет значительные трудности. Некоторые исследователи не случайно рассматривают иронию и сарказм как отдельную группу факторов, осложняющих анализ тональности [4], поскольку решение этой проблемы требует отдельных дополнительных исследований. На текущем этапе нашего проекта тонально окрашенные лексические единицы, часто используемые в рассматриваемом корпусе как в буквальном значении, так и иронично или саркастически, отнесены одновременно к двум полярным онтологическим концептам ветки SENTIMENT: POS и NEG или POSX и NEGX соответственно. Например, слово *спасибо*, нередко употребляемое пользователями саркастически, в нашем онтолексиконе, как и в предыдущем случае, соотнесено с концептами POS и NEG. Как и в описанном выше случае, для разрешения такой неоднозначности потребуется либо разработка формальных правил, либо привлечение эксперта-постредактора.

Еще одним контекстным фактором, обнаруженным в исследуемом корпусе, является оценочное слово в составе нейтральной многокомпонентной группы. Примером является употребление в текстах отзывов устойчивых речевых формул, в частности формул приветствия. Так, слово *добрый*, имеющее положительную тональность вне контекста, в словосочетании *добрый день* не является выражением мнения автора о медицинском центре, а представляет собой исключительно формулу вежливости и не влияет на оценку тональности высказывания: ее в равной степени используют как довольные, так и недовольные пациенты. Решением этой проблемы в нашем проекте

является включение подобных речевых формул в онтолексикон, лексемы которого соотнесены с концептом OTHER. Такая классификация позволяет не учитывать словарно-оценочные лексемы в составе нейтральной многокомпонентной группы при оценке тональности.

Помимо рассмотренных выше факторов, анализ контекстных факторов в широком контексте исходного корпуса позволяет выделить еще один фактор: наличие в отзыве об одном объекте (клинике или аспекте) тонально окрашенной лексики, не относящейся к указанному объекту, что может быть проиллюстрировано на примере следующего отзыва: *«Была на приеме у терапевта. За прием отдала 1500, прием был 5 минут. Ничего так и не поняла, пошла в другой центр. Там осмотрели с ног до головы. Назначили правильное лечение, которое действует, и я иду на поправку»* (11). В этом отзыве о медицинском центре N отрицательное впечатление о данном центре выра-

жено только в первой половине отзыва (выделено жирным), в то время как вторая половина отзыва с описанием положительных впечатлений посвящена медицинскому центру M. Во избежание искажения результатов анализа тональности такие отзывы могут быть полностью или частично исключены из анализируемой выборки. Так, в приведенном примере фрагмент отзыва, в котором речь идет о другом медицинском центре, может быть автоматически отсечен по ключевой фразе *в другой центр*. Проблемы могут возникнуть и в тех случаях, когда в отзыве отсутствуют конкретные маркеры изменения темы или темы изменяются слишком часто, — такие отзывы должны быть вручную проверены и исключены экспертом.

Выявленные контекстные факторы с указанием соответствующих им онтологических концептов и примеров лексических единиц, реализующих данные факторы в корпусе, систематизированы в табл. 1.

Таблица 1

Table 1

Контекстные факторы в корпусе отзывов о медицинских услугах

Тип контекстного фактора	Онтологический концепт (значение)	Примеры единиц онтолексикона
Неспецифический контекстный фактор	MAG (усиление тона)	в особенности хочется отметить, нет никакого сомнения
	ANT (ослабление тона)	вроде, вряд ли, иногда, кое-где, лишь, не столь важно
	REV (изменение тона на противоположный)	без, далеко не, до сегодняшнего дня, не
	POS / NEG / POSX / NEGX + ASPECT (хорошо / плохо / отлично / отвратительно + подкласс концепта ASPECT)	не врач, не берут трубку, не буду сюда больше обращаться
Ирреалис	REV (изменение тона на противоположный)	а я думала, где тут, ладно бы, научите, позиционируют себя, следите лучше за
	POS / NEG / POSX / NEGX + ASPECT (хорошо / плохо / отлично / отвратительно + подкласс концепта ASPECT)	должен быть выбор, должна быть на высшем уровне, не должны занимать парковки
Ирония и сарказм	POS + NEG, POSX + NEGX (комбинация значений хорошо / плохо / отлично / отвратительно)	браво, замечательное руководство, классный специалист, круто, особенно повезло, роскошное место, спасибо
Оценочное слово в составе нейтральной многокомпонентной группы	O (прочее)	дал добро, добрый день, добрый вечер, скорая помощь, уважаемое руководство
Тональная лексика, не относящаяся к объекту отзыва	POS / NEG / POSX / NEGX (хорошо / плохо / отлично / отвратительно)	иду на поправку, правильное лечение, стало еще хуже

- (1) {Порядовала}~POS {лишь}~ANT девушка-администратор.
- (2) Медицинский центр Лотос {всегда}~MAG отличает {профессионализм медицинского и административного персонала}~S~POS.
- (3) {До сегодняшнего дня}~REV {претензий к качеству предоставляемых услуг не было}~Q~POS.
- (4) Бывают {недочеты}~NEG со стороны работы администраторов, но это {не критично}~ANT.
- (5) Это {не врач}~CH~NEGX!
- (6) Как в поликлинику хрен дозвонишься, {а я думала}~REV {частная клиника высокого профиля}~CP~POS.
- (7) {Где тут}~REV {профессиональные врачи}~Q~POS?
- (8) И это при том, что компания {позиционирует себя}~REV как {клиентоориентированная}~POS и {работающая с обратной связью}~COM~POS.
- (9) Рекомендовали её, как {первоклассного специалиста}~Q~NEG~POSX.
- (10) Считаю если вы имеете смелость заявлять такие цены, то и отношение к людям,
- (11) Была на приеме у терапевта. За прием отдала 1500, прием был 5 минут. {Ничего}~NEG так и не поняла, {пошла в другой центр}~CP~NEG. [Там осмотрели с ног до головы. Назначили {правильное лечение}~POS, которое действует, и я {иду на поправку}~Q~POS.]

Рис. 1. Автоматическая аннотация фрагментов исходного корпуса на основе обновленного е-лексикона

Fig. 1. Automatic annotation of fragments of the original corpus based on the updated e-lexicon

Выделенные и систематизированные в результате описанного выше корпусного анализа глоссы контекстных факторов внесены в онтолексикон платформы концептуального аннотирования, использующей коды связанных с лексемами онтологических концептов в качестве меток аннотации, ориентированной на аспектно-тональный анализ, которая может быть использована как для формулировки правил автоматического постредактирования, так и непосредственно для вычисления тональности текстов. На рис. 1 показана автоматическая аннотация приведенных выше примеров на основе обновленного онтолексикона, более корректно учитывающего контекстные факторы.

Заключение

В настоящей работе представлен анализ контекстных факторов, которые влияют на интерпретацию оценки, выражаемой тональными лексическими единицами, при анализе тональности

отзывов о медицинских услугах на основе словаря тональной лексики. Проведенное исследование позволило увеличить покрываемость аннотационного словаря посредством отнесения выделенных лексических единиц, в виде которых контекстные факторы реализованы в исследуемом текстовом материале, в соответствующие концептуальные классы предметной области, что в свою очередь обеспечивает более корректное автоматическое аннотирование отзывов и сокращает время и усилия на их постредактирование.

Полученные результаты открывают горизонты для дальнейших исследований: в частности, планируется более тщательный анализ таких факторов, как ирония и сарказм, а также разработка формальных лингвистических правил для платформы концептуального аннотирования, которые позволят корректно учитывать наличие выявленных контекстных факторов в аннотируемых текстах.

Список источников

1. Бабина О. И. Модель регрессионного анализа тональности текста для оценки уровня удовлетворенности клиента // Вестник Южно-Уральского государственного университета. Серия: Лингвистика. 2024. Т. 21, № 4. С. 63–70.

2. Кузнецова Е. С., Лукашевич Н. В., Четверкин И. И. Тестирование правил для системы анализа тональности // Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегодной международной конференции «Диалог» (2013), Бекасово, 29 мая – 2 июня 2013 года. Вып. 12 (19). Том 2. Доклады специальных секций. М. : Изд-во РГГУ, 2013. С. 71–78.
3. Кулагин Д. И. Открытый тональный словарь русского языка КартаСловСент // Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегодной международной конференции «Диалог» (2021), Москва, 16–19 июня 2021 года. Вып. 20. М. : Изд-во РГГУ, 2021. С. 1106–1119.
4. Лукашевич Н. В. Автоматический анализ тональности текстов: проблемы и методы // Интеллектуальные системы. Теория и приложения. 2022. Т. 26, № 1. С. 50–61.
5. Полозов И. К., Волкова И. А. Применение технологии Word2vec в задаче выделения инверторов тональности // МНИЖ. 2020. № 4–1 (94). С. 36–39.
6. Самигулин Т. Р., Джурабаев А. Э. У. Анализ тональности текста методами машинного обучения // Научный результат. Информационные технологии. 2021. Т. 6, № 1. С. 55–62.
7. Шереметьева С. О., Бабина О. И. Платформа для концептуального аннотирования многоязычных текстов // Вестник ЮУрГУ. Серия «Лингвистика». 2020. Т. 17, № 4. С. 53–60.
8. Abdulla N., Ahmed N. A., Shebab M. A., Al-Ayyoub M., Al-Kabi M., Al-Rifai S. Towards improving the lexicon-based approach for Arabic sentiment analysis // International Journal of Information Technology and Web Engineering. 2014. Vol. 9 (3). P. 55–71.
9. Alfrjani R., Osman T., Cosma G. A hybrid semantic knowledgebase-machine learning approach for opinion mining // Data & Knowledge Engineering. 2019. Vol. 121. P. 88–108.
10. Appel O., Chiclana F., Carter J., Fujita H. Successes and challenges in developing a hybrid approach to sentiment analysis // Applied Intelligence. 2017. Vol. 48. P. 1176–1188.
11. Blinov P. D., Klekovkina M. V., Kotelnikov E. V., Pestov O. A. Research of lexical approach and machine learning methods for sentiment analysis // Computational Linguistics and Intellectual Technologies. 2013. Vol. 2 (12). P. 48–58.
12. Gupta I., Joshi N. Enhanced Twitter sentiment analysis using hybrid approach and by accounting local contextual semantic // Journal of Intelligent Systems. 2019. Vol. 29 (1). P. 1611–1625.
13. Polanyi L., Zaenen A. Contextual valence shifters // Computing Attitude and Affect in Text: Theory and Applications. The Information Retrieval Series. Dordrecht: Springer, 2006. Vol. 20. P. 1–10.
14. Sebastiani F. Machine learning in automated text categorization // ACM Computing Surveys. 2002. Vol. 34. P. 1–47.
15. Sheremetyeva S. O., Babina O. I. On automated creation of gold-standard corpus for multi-aspect sentiment analysis // Proceedings of the of the International Conference “Internet and Modern Society”. Saint-Petersburg: ITMO, 2024 (в печати).

References

1. Babina OI. Regression-based sentiment analysis model for predicting customer satisfaction. *Bulletin of the South Ural State University. Ser. Linguistics*. 2024;21(4):63-70. (In Russ.).
2. Kuznetsova ES, Loukachevitch NV, Chetviorkin II. Testing rules for a sentiment analysis system. *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference “Dialogue”*. 2013;12(2):71-78. (In Russ.).
3. Kulagin DI. Publicly available sentiment dictionary for the Russian language KartaSlovSent. *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference “Dialogue”*. 2021;20:1106-1119. (In Russ.).
4. Loukachevitch NV. Automatic sentiment analysis of texts: problems and methods. *Intelligent Systems. Theory and Applications*. 2022;26(1):50-61. (In Russ.).
5. Polozov IK, Volkova IA. Applying Word2vec technology to shifter extraction task. *MNIZh = International Research Journal*. 2020;4-1(94):36-39. (In Russ.).
6. Samigulin TR, Djurabaev AEU. Sentiment analysis of text by machine learning methods. *Research Result. Information Technologies*. 2021;6(1):55-62. (In Russ.).
7. Sheremetyeva SO, Babina OI. A platform for knowledge assisted conceptual annotation of multilingual texts. *Bulletin of the South Ural State University. Ser. Linguistics*. 2020;17(4):53-60. (In Russ.).

8. Abdulla N, Ahmed NA, Shebab MA, Al-Ayyoub M, Al-Kabi M, Al-Rifai S. Towards improving the lexicon-based approach for Arabic sentiment analysis. *International Journal of Information Technology and Web Engineering*. 2014;9(3):55-71.
9. Alfrjani R, Osman T, Cosma G. A hybrid semantic knowledgebase-machine learning approach for opinion mining. *Data & Knowledge Engineering*. 2019;121:88-108.
10. Appel O, Chiclana F, Carter J, Fujita H. Successes and challenges in developing a hybrid approach to sentiment analysis. *Applied Intelligence*. 2017;48:1176-1188.
11. Blinov PD, Klekovkina MV, Kotelnikov EV, Pestov OA. Research of lexical approach and machine learning methods for sentiment analysis. *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue"*. 2013;2(12):48-58.
12. Gupta I, Joshi N. Enhanced twitter sentiment analysis using hybrid approach and by accounting local contextual semantic. *Journal of Intelligent Systems*. 2019;29(1):1611-1625.
13. Polanyi L, Zaenen A. Contextual valence shifters. *Computing Attitude and Affect in Text: Theory and Applications. The Information Retrieval Series. Dordrecht: Springer*; 2006;20:1-10.
14. Sebastiani F. Machine learning in automated text categorization. *ACM Computing Surveys*. 2002;34:1-47.
15. Sheremetyeva SO, Babina OI. On automated creation of gold-standard corpus for multi-aspect sentiment analysis. Proceedings of the of the International Conference "Internet and Modern Society". Saint-Petersburg: ITMO; 2024 (in print).

Информация об авторах

А. Ю. Зиновьева — кандидат филологических наук, доцент кафедры лингвистики и перевода.

С. О. Шереметьева — доктор филологических наук, доцент, профессор кафедры лингвистики и перевода.

Information about the authors

A. Yu. Zinoveva — Cand. Sci. (Philology), Associate Professor, Department of Linguistics and Translation Studies.

S. O. Sheremetyeva — Dr. Sci. (Philology), Professor, Department of Linguistics and Translation Studies.

Статья поступила в редакцию 09.02.2025; одобрена после рецензирования 24.02.2025; принята к публикации 22.05.2026.

The article was submitted 09.02.2025; approved after reviewing 24.02.2025; accepted for publication 22.05.2026.

Вклад авторов: оба автора сделали эквивалентный вклад в подготовку публикации.
Авторы заявляют об отсутствии конфликта интересов.

The authors declare no conflicts of interests.
Contribution of the authors: the authors contributed equally to this article.