

УДК 81'33

**Зиновьева Анастасия Юрьевна**  
кандидат филологических наук,  
доцент  
Южно-Уральский государственный  
университет  
г. Челябинск, Россия

**Anastasiia Zinoveva**  
PhD in Philology, Associate Professor  
South Ural State University  
Chelyabinsk, Russia  
*zinovevaai@susu.ru*

## ЭМОДЗИ В АНАЛИЗЕ ТОНАЛЬНОСТИ (на материале отзывов о медцентрах на русском языке)

Цель исследования – выявить особенности функционирования эмодзи в интернет-отзывах о медицинских центрах Челябинска и оценить их применимость в многоаспектном анализе тональности, в частности определить влияние эмодзи на вычисление показателя тональности. В качестве материала исследования использован корпус из 3247 отзывов пользователей электронного справочника 2GIS о 22 медицинских центрах города (по состоянию на март 2023 г.) общим объемом 200 265 словоупотреблений. В работе использованы количественные и качественные методы анализа: корпусный и контекстный анализ, лексико-онтологическое моделирование, сравнительный анализ. Инструментарий исследования включает базу данных эмодзи Emoji Sentiment Ranking, а также инструменты НОЦ «Лингво-инновационные технологии» ЮУрГУ для многоаспектного анализа тональности: онтологию предметной области, русскоязычный онтолексикон, платформу концептуального аннотирования, «золотой» корпус. Результаты исследования продемонстрировали, что в рассмотренном корпусе эмодзи используются для выражения различных градаций оценки и усиления тональности, для передачи оценок и (значительно реже) аспектов, могут дублировать тональность высказывания или выступать самостоятельным индикатором тональности.

*Ключевые слова: эмодзи; многоаспектный анализ тональности; тональная классификация эмодзи; интернет-отзывы; русский язык; медицинские услуги.*

## EMOJIS IN SENTIMENT ANALYSIS: A CASE STUDY IN RUSSIAN MEDICAL CENTER REVIEWS

The study aims to identify the functional characteristics of emojis in online reviews of medical centers in Chelyabinsk, and to assess their potential use in multi-aspect sentiment analysis—specifically, how emojis influence sentiment score calculations. The research material is a corpus of 3,247 user reviews extracted from the 2GIS electronic database, covering 22 city medical centers as of March 2023, with a total word count of 200,265. The methods employed in this study include quantitative and qualitative approaches, such as corpus and contextual analysis, lexical-ontological modeling, and comparative analysis. The research tools encompass the Emoji Sentiment Ranking database and multi-aspect sentiment analysis tools developed by the Innovation in Linguistics Research and Education Center at South Ural State University: a domain ontology, a Russian-language ontollexicon, a conceptual annotation platform, and a gold-standard corpus. The findings indicate that, within the corpus under analysis, emojis are used to express and intensify various sentiments, convey opinions or (much less frequently) evaluated aspects; they may reproduce the sentiment expressed by the text they accompany or act as a standalone sentiment indicator.

*Key words: emoji; multi-aspect sentiment analysis; emoji sentiment classification; online reviews; Russian; medical services.*

В настоящее время одним из наиболее перспективных направлений в области автоматической обработки текста является анализ тональности, поскольку он позволяет, среди прочего, извлекать из интернет-отзывов ценные сведения о мнениях потребителей тех или иных услуг. В этом контексте особое значение приобретают **эмодзи** (от яп. э ‘картинка’ + *модзи* ‘знак’) – «пиктограммы, используемые в интернет-коммуникации и изображающие лица, животных, распространенные предметы, знаки и природные явления» [1, с. 233], которые представляют собой значимые невербальные элементы, способные дополнять или модифицировать смысл исходного сообщения. Учет эмодзи в анализе тональности позволяет получать более точные результаты [2–4], особенно при обработке коротких сообщений, где эмодзи могут быть единственным индикатором настроения автора [5]. При этом в отличие от эмодзи (от англ. *emotion* ‘эмоция’ + *icon* ‘значок’) эмодзи могут быть использованы не только для передачи эмоций, но и для обозначения объектов, действий и идей, что потенциально позволяет извлекать из сообщений, помимо оценок и мнений, и сами оцениваемые аспекты.

**Цель** настоящей работы – выявить особенности функционирования эмодзи в корпусе интернет-отзывов об услугах частных медицинских центров города Челябинска на русском языке и оценить степень их применимости в многоаспектном анализе тональности, в частности определить их влияние на вычисление показателя тональности фрагментов данного корпуса.

Работа выполнена в рамках проекта, посвященного многоаспектному анализу тональности отзывов о медицинских центрах Челябинска, который реализуется в НОЦ «Лингво-инновационные технологии» Южно-Уральского государственного университета. В проекте используются специально разработанные предметно-ориентированные лингвистические ресурсы, часть которых задействована в настоящем исследовании, а именно:

1) **исходный корпус**: 3247 интернет-отзывов пользователей электронного справочника 2GIS [6] о 22 медицинских центрах города (по состоянию на март 2023 г.) общим объемом 200 265 словоупотреблений;

2) **онтология предметной области**, ориентированная на многоаспектный анализ тональности, которая содержит четыре ветви концептов: ASPECT (аспект оценки: например, Q – качество оказания услуг и результат лечения, INT – интерьер медицинского центра, его дизайн и оснащение), SENTIMENT (тональность: POS – положительная, POSX – крайне положительная, NEG – отрицательная, NEGX – крайне отрицательная), MODIFIER (модификатор тональности: ANT – ослабление, MAG – усиление, REV – изменение на противоположную), OTHER;

3) **исходный онтолексикон** – список релевантных лексических единиц, извлеченных из корпуса, соотнесенных с концептами онтологии;

4) **платформа концептуального аннотирования** [7]: теггер на основе лексико-онтологической базы знаний предметной области, который обеспечивает автоматическую разметку текста аспектными и тональными метками с возможностью последующего постредактирования разметки пользователем;

5) **размеченный корпус** отзывов о медицинских центрах с эталонной аспектно-тональной разметкой («золотой» корпус) без учета эмодзи (подробнее о разработке корпуса см. [8]).

Исследование было проведено в четыре этапа.

На **первом этапе** из исходного корпуса были извлечены все эмодзи с левым и правым контекстом в размере предложения. Каждому эмодзи были поставлены в соответствие код и название по стандарту Unicode [9]; например, эмодзи 👍 соответствует код U+1F44D и название «thumbs up». Эмодзи, обозначающие людей или части тела, как правило, представлены в нескольких вариантах, которые различаются цветом кожи и (или) половой принадлежностью изображенных людей; в этом случае они имеют дополнительные коды. При подсчете частоты встречаемости эмодзи с общим основным кодом были объединены, поскольку указанные различия представляются нерелевантным для поставленной задачи.

На **втором этапе** стояла задача классифицировать каждый эмодзи. В контексте анализа тональности эмодзи традиционно подразделяют на положительные, отрицательные и нейтральные, однако возможна более подробная классификация, если это продиктовано задачами исследования [10]. В нашем случае за основу классификации была взята описанная выше онтология. Кроме того, для определения тональности эмодзи в выделенных фрагментах отзывов применялись база данных Emoji Sentiment Ranking [11] и контекстный анализ эмодзи на основе данных «золотого» корпуса, что позволило отнести каждый эмодзи к одному или нескольким концептам онтологии, принадлежащим ветвям ASPECT, SENTIMENT, MODIFIER и OTHER.

На **третьем этапе** исследования была произведена лексикализация эмодзи в выделенных фрагментах исходного корпуса, т. е. каждый эмодзи был заменен на его название по Unicode; также соответственно был обновлен онтолексикон. Далее фрагменты исходного корпуса были повторно размечены на платформе концептуального аннотирования с обновленной лексико-онтологической базой знаний и подвергнуты последующему постредактированию для снятия неоднозначности меток, в результате чего был создан «золотой» корпус с учетом эмодзи.

На **четвертом этапе** были рассчитаны и сопоставлены друг с другом показатели тональности каждого фрагмента «золотого» корпуса без учета и с учетом эмодзи. Показатель тональности (S) рассчитывался по формуле:

$$S = \frac{1}{n} \sum_{i=1}^n c_i m_i + 3, (1)$$

где  $n$  – количество тонально окрашенных единиц в тексте;

$c_i$  – коэффициент тональности  $i$ -й единицы;

$m_i$  – коэффициент, учитывающий влияние модификатора.

Примечание: коэффициент  $c_i$  принимает следующие значения: POSX = 2, POS = 1, NEG = -1, NEGX = -2. Коэффициент  $m_i$  равен 2 при воздействии на тонально окрашенную единицу модификатора MAG (кроме случаев, когда такая единица имеет тональность POSX или NEGX, для которых  $m_i = 1$ ); 0,5 при воздействии модификатора ANT и 1 в остальных случаях. Показатель тональности  $S$  принимает значения в диапазоне от 1 до 5, где 1 соответствует максимально отрицательной тональности, а 5 – максимально положительной. Если  $n = 0$ , показатель тональности не вычисляется.

По результатам первого этапа исследования в рассматриваемом корпусе обнаружено 286 фрагментов, содержащих от одного до десяти эмодзи, причем в 285 случаях эмодзи иллюстрируют левый контекст и лишь в одном – правый. Общее количество эмодзи в корпусе составило 402. Среди них выделено 78 уникальных эмодзи (при этом эмодзи, отличающиеся только цветом кожи или половой принадлежностью изображенных людей, но имеющие общий код по стандарту Unicode, учитывались как один символ), 28 из которых встречаются в корпусе только один раз. Десять наиболее частых эмодзи представлены в Таблице 1.

Таблица 1

Частотный рейтинг десяти наиболее употребимых эмодзи  
в исследуемом корпусе отзывов о медицинских центрах

№ п/п	Эмодзи	Код по Unicode	Название по Unicode	Частота
1		U+1F44D	thumbs up	72
2		U+2764	red heart	29
3		U+1F64F	folded hands	25
4		U+1F44E	thumbs down	16
5		U+1F621	enraged face	15
6		U+1F60A	smiling face with smiling eyes	15
7		U+263A	smiling face	14
8		U+1F926	person facepalming	13
9		U+1F601	beaming face with smiling eyes	11
10		U+1F490	bouquet	11

На втором этапе эмодзи были распределены по девяти концептам онтологии, среди которых представлены как концепты, обозначающие тональность и ее модификаторы (POS, POSX, NEG, NEGX, MAG), так и концепты ветвей ASPECT и OTHER. При этом эмодзи, ослабляющие положительную или отрицательную тональность высказывания (ANT), а также изменяющие тональность на противоположную (REV), в корпусе не выявлены. Лишь три эмодзи были отнесены к концептам ветви ASPECT, что свидетельствует о предпочтении авторов отзывов выражать оцениваемые аспекты вербально.

Эмодзи, отнесенные к концепту OTHER, не обладают тональностью и не изменяют тональность высказывания вообще или в отдельных случаях употребления в корпусе, либо обозначаемая ими эмоция относится к самому автору, а не к аспектам работы медицинского центра. Например, в предложении *Мы обошли несметное количество врачей, которые разводили руки* 🙄 (примеры приводятся по исходному корпусу в авторской редакции) эмодзи 🙄 иллюстрирует выражение *разводили руки* и является нетональным, а в предложении *Хочется отметить медсестру из процедурной, (к огромному своему стыду не знаю имени 😞)* эмодзи 😞 передает сожаление автора по поводу незнания имени медсестры и не относится к оценке медицинского центра.

Некоторые эмодзи проявляют неоднозначность в корпусе, вследствие чего они были отнесены к нескольким концептам. Преимущественно это традиционно положительные эмодзи (👍, 😊, 🙌, 😄 и т. п.), которые в отдельных контекстах выражают иронию и сарказм. Например, в предложении *Краснуха, которая прошла после одной таблетки супрастина* 🙌👍 оба эмодзи несут иронический подтекст, отражающий сомнение автора по поводу поставленного диагноза.

Повторение эмодзи усиливает их тональность, поэтому группы из двух и более одинаковых эмодзи, имеющих тональность POS и NEG, отнесены соответственно к концептам POSX и NEGX. Частный случай – эмодзи ★, который, по данным *Emoji Sentiment Ranking*, преимущественно позитивен либо нейтрален, однако его также используют для выражения оценки по пятибалльной шкале. Соответственно, в таком контексте ★ соответствует концепту NEGX, ★★ – NEG, ★★★★★ – POS, ★★★★★★ – POSX (случай с тремя звездами потребует дополнительного контекстного анализа, поскольку оценка «три» по пятибалльной шкале может быть как нейтральной, так и отрицательной).

Итоговое распределение эмодзи по концептам приведено в Таблице 2.

Таблица 2

Распределение эмодзи по концептам онтологии

Концепт	Описание концепта	Эмодзи
POS	Положительная тональность; базовые положительные эмоции (радость, одобрение, благодарность)	
POSX	Крайне положительная тональность; более интенсивные положительные эмоции (восторг, восхищение, глубокое удовлетворение)	
NEG	Отрицательная тональность; базовые отрицательные эмоции (грусть, разочарование, злость, тревога); ирония и сарказм	
NEGX	Крайне отрицательная тональность; более интенсивные отрицательные эмоции (отчаяние, ярость, ужас); насмешка	
MAG	Усиление тональности высказывания	
OTHER	Эмодзи, не имеющие тональности и не изменяющие тональность высказывания либо не относящиеся к оценке медицинского центра	
INT	Интерьер медцентра, его дизайн, оснащение	
Q	Качество оказания услуг, результат лечения	

Результаты сравнительного анализа показателей тональности фрагментов в выборках «золотого» корпуса с учетом и без учета эмодзи приведены на Рисунке 1. На графике представлено распределение показателей тональности 242 из 286 фрагментов. Большая часть точек расположена вблизи пунктирной линии – это свидетельствует о том, что присутствие эмодзи не оказывает существенного влияния на тональность фрагмента. В тех случаях, когда влияние имеется, оно носит незначительный характер.

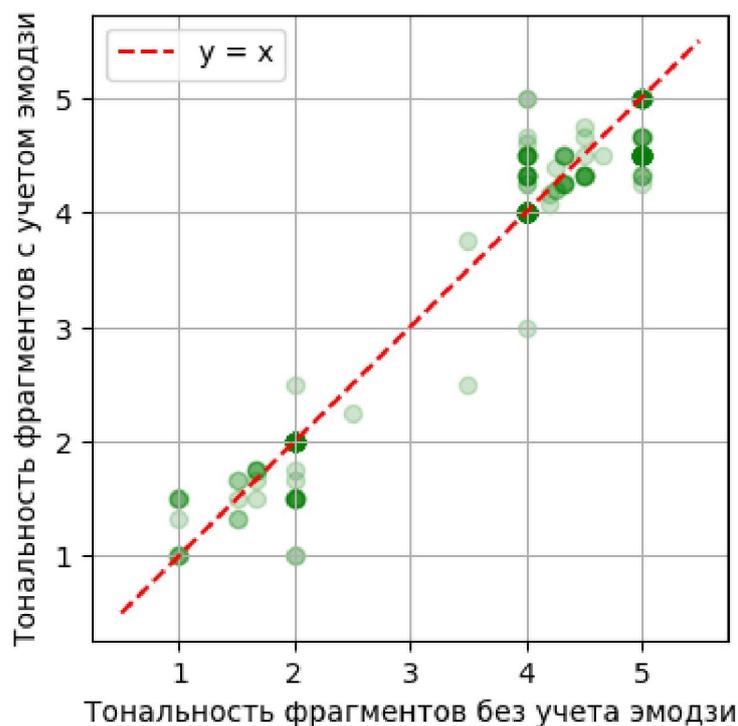


Рис. 1. Сравнение тональности фрагментов корпуса с учетом и без учета эмодзи

Показатель тональности 44 из 286 фрагментов изначально не рассчитывался из-за отсутствия тонально окрашенных лексических единиц, однако включение эмодзи позволило определить тональность еще 37 из них (оставшиеся семь фрагментов не имеют тональности, либо она нейтральна). Интересно, что в пяти фрагментах эмодзи выполняют функцию предиката и являются неотъемлемой частью высказывания, без которой невозможно определить его тональность; например: *Лор 🙄*; *Питание тоже 🙄*; *медсестры просто 🔥🔥🔥*. В остальных 32 фрагментах эмодзи выступают в роли дополнительного, факультативного элемента, уточняющего эмоциональный оттенок высказывания. Без эмодзи данные фрагменты остаются понятными и сохраняют базовую тональность, которую, однако, затруднительно определить с использованием тонального лексикона, поскольку оценка ситуации не выражена явно на уровне лексем. Примером является фрагмент *Мы нашли своего гастроэнтеролога 😊*, где ни одно из слов не обладает тональной окраской, однако совокупность слов формирует позитивную идею, подтверждаемую положительным эмодзи.

Влияние эмодзи на определение анализа тональности текстовых фрагментов можно проиллюстрировать на примере самого распространенного в корпусе символа 👍 *thumbs up* ‘палец вверх’. По данным *Emoji Sentiment Ranking*, эмодзи 👍 в большей мере положителен: его показатель тональности

в диапазоне  $[-1 \dots 1]$  равен 0,521. В рассматриваемом корпусе данный эмодзи, который встречается 72 раза в 57 фрагментах, был отнесен к концептам POS, NEG, OTHER или POSX (в группах от двух эмодзи), что отражено на Рисунке 2.

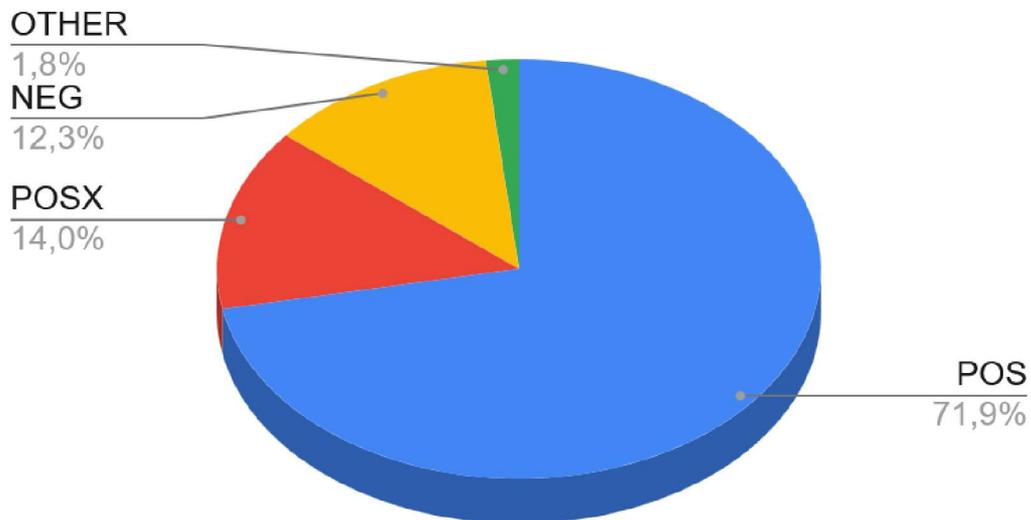


Рис. 2. Варианты тональности эмодзи 👍 *thumbs up* ‘палец вверх’ в корпусе

Относительно влияния эмодзи 👍 на определение показателя тональности следует отметить, в 29 из 55 фрагментов показатель оставался неизменным; в 8 – изменялся менее чем на 0,5; в 14 – на 0,5; в 1 – на 0,67. В четырех фрагментах вычисление показателя тональности было возможно исключительно с учетом эмодзи. В одном случае включение эмодзи в расчет не повлияло на результат, поскольку эмодзи 👍 не обладал тональностью и иллюстрировал слово *ок*, использованное пациентом для описания своего состояния до обращения к врачу.

Итак, проведенный анализ корпуса отзывов о медицинских центрах позволил выявить несколько ключевых аспектов функционирования эмодзи и их влияния на определение тональности текста в рамках данной предметной области. Во-первых, проанализированные эмодзи служат средством выражения различных градаций оценки, включая положительную, крайне положительную, отрицательную и крайне отрицательную, а также способны усиливать тональность высказывания. Во-вторых, эмодзи могут применяться в многоаспектном анализе тональности не только для выявления оценок и мнений, но и для извлечения оцениваемых аспектов, однако в исследуемом корпусе наблюдается тенденция к вербальному выражению аспектов, эмодзи же преимущественно используются для передачи оценок. Наконец, преобладающая функция эмодзи заключается в дублировании тональности тексто-

вого фрагмента, однако они могут также выступать в качестве единственного индикатора тональности коротких фрагментов, что согласуется с результатами предыдущих исследований в данной области.

## ЛИТЕРАТУРА

1. Словарь языка интернета.ru / Под. ред. М. А. Кронгауза. М.: АСТ-ПРЕСС КНИГА, 2016. 288 с.
2. Sentiment of Emojis [Electronic resource] / P. Kralj Novak, J. Smailović, B. Sluban, I. Mozetič // PLoS One. 2015. Vol. 10 (12). DOI: <https://doi.org/10.1371/journal.pone.0144296>.
3. EMFSA: Emoji-Based Multifeature Fusion Sentiment Analysis [Electronic resource] / H. Tang, W. Tang, D. Zhu, Sh. Wang, Y. Wang, L. Wang // PLoS One. 2024. Vol. 19 (9). DOI: <https://doi.org/10.1371/journal.pone.0310715>.
4. Alawneh H., Hasasneh A., Maree M. On the Utilization of Emoji Encoding and Data Preprocessing with a Combined CNN-LSTM Framework for Arabic Sentiment Analysis // Modelling. 2024. Vol. 5 (4). P. 1469–1489.
5. Emoji-Integrated Polyseme Probabilistic Analysis Model: Sentiment Analysis of Short Review Texts on Library Service Quality [Electronic resource] / X. Zheng, W. Chen, H. Zhou, Zhe Li, T. Zhang, Qi Yuan // ИИТА. 2022. P. 313–322. DOI: <https://doi.org/10.18280/ts.390133>.
6. Карта Челябинска: улица, дома и организации города – 2GIS [Электронный ресурс]. URL: <https://2gis.ru/chelyabinsk> (дата обращения: 11.07.2025).
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. Sheremetyeva S. O., Babina O. I. On Automated Creation of Gold-Standard Corpus for Multi-Aspect Sentiment Analysis // Internet and Modern Society. Human-Computer Communication. 2025 (in print).
9. Full Emoji List, v16.0 [Electronic resource]. URL: <https://unicode.org/emoji/charts/full-emoji-list.html> (date of access: 11.07.2025).
10. Hakami S. A. A., Hendley R., Smith Ph. Emoji Sentiment Roles for Sentiment Analysis: A Case Study in Arabic Texts // Proceedings of the The Seventh Arabic Natural Language Processing Workshop (WANLP). 2022. P. 346–355.
11. Emoji Sentiment Ranking v1.0 [Electronic resource]. URL: [https://kt.ijs.si/data/Emoji\\_sentiment\\_ranking/index.html](https://kt.ijs.si/data/Emoji_sentiment_ranking/index.html) (date of access: 11.07.2025).