



<https://doi.org/10.22363/2687-0088-44439>

EDN: CDQHHV

Research article / Научная статья

RusEmoLex: Russian Emotion Lexicon

Polina V. IAROSHENKO^{1,2}  , Natalia V. LOUKACHEVITCH^{1,2} 

¹ *Lomonosov Moscow State University, Moscow, Russia*

² *Ivannikov Institute for System Programming of the Russian Academy of Sciences, Moscow, Russia*

 polina.iaroshenko@yandex.ru

Abstract

Emotion detection in natural language is an important area of computational linguistics. It involves the description, categorisation and systematization of emotion lexicon that can be further employed to enhance the performance of machine learning algorithms. The aim of the study is twofold: firstly, to provide a comprehensive description of the Russian-language resources pertaining to emotion vocabulary, and secondly, to create an emotion lexicon based on these resources. The study describes the following types of Russian-language sources: lexicographic resources (dictionaries and thesauri); data from the semantically tagged part of the Russian National Corpus; datasets consisting of texts labelled by emotion type; and word–emotion association lexicons. The methodology of creating RusEmoLex encompasses the following stages: the creation of initial lists of emotion words derived from available sources; the compilation of the lists; and the selection of lexis in RusEmoLex. A lexical unit has to meet the following criteria to be included in RusEmoLex: repeated occurrence in the sources (a lexical unit needs to be present in two or more sources); the quality of the sources (a lexical unit needs to be included either in a dictionary or a corpus resource and it should not be a translation); the class label (the lexical units have an emotional label established on the basis of the majority of sources). RusEmoLex is an open access resource which includes 1024 lexical units and can be used for linguistic studies and in the field of natural language processing.

Keywords: *emotions, emotion lexicon, semantics, sentiment analysis, machine learning, natural language processing*

For citation:

Iaroshenko, Polina V. & Natalia V. Loukachevitch. 2025. RusEmoLex: Russian Emotion Lexicon. *Russian Journal of Linguistics* 29 (3). 659–687. <https://doi.org/10.22363/2687-0088-44439>



RusEmoLex: эмоциональный лексикон для русского языка

П.В. ЯРОШЕНКО^{1,2}  , Н.В. ЛУКАШЕВИЧ^{1,2} 

¹Московский государственный университет имени М.В. Ломоносова, Москва, Россия

²Институт системного программирования имени В.П. Иванникова
Российской академии наук, Москва, Россия

 polina.iaroshenko@yandex.ru

Аннотация

Распознавание эмоций в тексте – одна из актуальных задач в области обработки естественного языка. Неотъемлемой частью этой задачи является описание и систематизация эмоциональной лексики, которая может служить в качестве маркера, указывающего на то, что в тексте выражается та или иная эмоция. На сегодня уже существует достаточно много англоязычных ресурсов эмоциональной лексики в открытом доступе, однако для других языков, в частности русского, размеченных данных значительно меньше. Цель работы – описать доступные для использования русскоязычные ресурсы эмоциональной лексики и представить созданный на их основе новый объединяющий эмоциональный лексикон – RusEmoLex (Russian Emotion Lexicon). Описаны следующие типы русскоязычных источников: лексикографические ресурсы (словари и тезаурусы); корпусные данные; данные, полученные в результате опросов (зафиксированные ассоциации слов с эмоциями); наборы данных, используемые в машинном обучении. Описана методика создания ресурса RusEmoLex на основе доступных русскоязычных источников. Методика включает следующие этапы: формирование исходных списков эмоциональной лексики на основе доступных источников; объединение списков; отбор лексики в RusEmoLex. Для включения в RusEmoLex лексическая единица должна удовлетворять следующим критериям: (1) количество источников, где зафиксирована лексическая единица (лексическая единица должна входить в два или более источников); (2) качество источников (как минимум один из источников, куда вошла лексическая единица, должен относиться к словарным или корпусным ресурсам и не являться переводным); (3) наличие метки класса (лексическая единица должна иметь установленный на основании большинства источников эмоциональный класс). Основным результатом исследования является эмоциональный лексикон RusEmoLex, включающий 1024 лексических единиц, размеченных по частеречной принадлежности; эмоциональному классу; категории источников, где встречается слово; количеству вхождений в источники. RusEmoLex находится в открытом доступе и может использоваться как для собственно лингвистических задач, так и в области обработки естественного языка.

Ключевые слова: эмоции, эмоциональная лексика, семантика, анализ тональности, машинное обучение, обработка естественного языка

Для цитирования:

Ярошенко П.В., Лукашевич Н.В. RusEmoLex: эмоциональный лексикон для русского языка. *Russian Journal of Linguistics*. 2025. Vol. 29. № 3. Р. 659–687. <https://doi.org/10.22363/2687-0088-44439>

1. Введение

Эмоции как объект исследования подразумевают междисциплинарность. Изучение собственно эмоций относится скорее к области психологии, в то же время выражение эмоций в языке активно исследуется в различных областях

лингвистической науки: в семантике (Гладкова 2010, Падучева 2004: 273–307, Иоанесян 2024, Gladkova 2022, Wierzbicka 2010 и др.), прагматике (Alba-Juez & Naugh 2025, Ларина 2015, 2019), и др., в том числе в кросс-культурном контексте (Ионова, Шаховский 2018, Gladkova 2010, Wierzbicka 1992, 1999, 2018 и др.). В трудах В. И. Шаховского (2008, 2009 и др.) были сформулированы теоретические основы лингвистики эмоций (эмотиологии) как отдельного направления языкознания. Лингвистика эмоций характеризуется полипарадигмальностью, то есть находится на пересечении целого ряда аспектов – когнитивного, прагматического, культурологического, текстового и др. (Шаховский 2008: 24). Одной из наиболее активно развивающихся областей в рамках лингвистики эмоций является изучение эмоций в коммуникативно-прагматическом и дискурсивном аспекте (Alba-Juez & Larina 2018, Larina & Ponton 2022, Mackenzie & Alba-Juez 2019, Zappettini et al. 2021). Этому направлению, в частности, посвящены специальные выпуски журнала *Russian Journal of Linguistics* (2015, № 1; 2018, Т. 22. № 1; 2021, Т. 25, № 3).

В компьютерной лингвистике значительное развитие получила задача анализа эмоций¹ (emotion analysis), которая подразумевает не только распознавание эмоций (emotion recognition), но также их понимание (emotional understanding) и даже имитацию. Большое количество исследований посвящено изучению эмоционального интеллекта больших языковых моделей, их способности воспроизводить эмоциональные реакции человека (см., например, Dalal et al. 2025).

Анализ эмоций имеет практическую ценность для многих приложений обработки естественного языка: создание персонализированных систем искусственного интеллекта и совершенствование диалоговых систем, распознавание эмоций в текстах отзывов или в комментариях пользователей социальных сетей (Cavicchio 2025: 5–8), классификация текстов в зависимости от наиболее ярко выраженной эмоций (Sboev et al. 2021) и др.

Описание и систематизация эмоциональной лексики могут способствовать улучшению качества работы систем искусственного интеллекта. Лексические единицы этого класса могут служить в качестве своеобразных маркеров, которые указывают на то, что в тексте выражается та или иная эмоция. Список эмоциональных слов или словосочетаний-маркеров может быть полезен при создании наборов данных для машинного обучения, а также может применяться для дообучения моделей. В области обработки естественного языка (natural language processing) такие списки зачастую называются лексиконами². В работе термин «лексикон» будет употребляться именно в этом значении.

¹ Термин «анализ эмоций» в данной статье употребляется в том же значении, что и англоязычный аналог, активно используемый в исследованиях, посвященных обработке естественного языка, — «emotion analysis» (см., например, Plaza-del-Arco et al. 2024). Под анализом эмоций понимается изучение различных способов отображения эмоций в языке.

² См., например, (Cavicchio 2025: 34), где эмоциональный лексикон определяется как список слов, выражений или закономерностей, позволяющих уловить различные оттенки эмоционального спектра.

Как и во многих других областях обработки естественного языка, в сфере анализа эмоций наблюдается активное преобладание англоязычных ресурсов (см., например, Mohammad 2023, Buechel et al. 2020). Для других языков, в том числе и для русского, размеченных данных значительно меньше. Для русского языка существует ряд ресурсов, разнородных как по целям и методике их создания, так и по формату. Можно выделить следующие типы источников:

- лексикографические ресурсы (словари и тезаурусы); корпусные данные;
- данные, полученные в результате опросов (зафиксированные ассоциации слов с эмоциями);
- наборы данных, созданные для использования в машинном обучении.

Предварительный анализ ресурсов показал, что лексика в разных источниках сильно различается, соответственно, слов, которые бы встречались одновременно в большом количестве источников, не так много. Кроме того, не все источники можно считать в равной степени надежными, многие требуют дополнительной проверки.

Целью работы является описание доступных для использования русскоязычных ресурсов эмоциональной лексики и создание на их основе нового объединяющего эмоционального лексикона – RusEmoLex (Russian Emotion Lexicon).

Первый раздел статьи посвящен проблеме описания эмоциональной лексики. Во втором разделе представлены доступные русскоязычные источники эмоциональной лексики. В третьем разделе дана методика создания объединяющего лексикона RusEmoLex на основе описанных источников. В четвертом разделе описана структура RusEmoLex и представлены фрагменты из него. Пятый раздел посвящен обсуждению результатов и дальнейших перспектив исследования.

2. Эмоциональная лексика: к проблеме описания семантического класса

Языковые средства выражения эмоций разнообразны. Эмоциональный компонент может быть реализован в единицах различных языковых уровней: фонетического, морфологического, лексического, синтаксического (Ионова 2023). В рамках нашего исследования будет рассматриваться только уровень лексики. Описание семантического класса эмоциональной лексики³ представляется достаточно трудной задачей, об этом упоминается, в частности, в работе (Апресян 1995: 366–373), где подсистема характеризуется как одна из наиболее сложно организованных.

Не вполне очевидно, каким образом очертить границу для класса эмоциональной лексики, поскольку слова могут непосредственно указывать на

³ Для обозначения этого семантического класса могут использоваться различные термины: «эмотивная лексика», «аффективная лексика» или же «эмоциональная лексика». В настоящей статье здесь и далее этот класс будет обозначаться как «эмоциональная лексика».

эмоции, называть их (*радость, злость*) или же быть связанными с эмоциями более опосредованно. Апресян (1995) разделяет базовую эмоциональную лексику (*беспокоиться, страх*) и лексические единицы, которые не являются обозначениями эмоции в собственном смысле, однако включают в свое значение указание на эмоциональное состояние субъекта (*любоваться* – смотреть, испытывая при этом определенную эмоцию), либо метафорические обозначения физических симптомов эмоции (*засиять* – например, от радости) (Апресян 1995: 366–373).

В.Ю. Апресян (2010) выделяет четыре подтипа языковых средств выражения эмоций:

- клишированные эмоциональные междометия (например, *ого* для удивления);
- свободные контекстуально-мотивированные способы выражения (например, ввод прямой речи глаголом эмоции: *удивилась она, обрадовался он*);
- специализированные слова, закрепленные за той или иной эмоцией (например, *потрясающе* для восторга);
- устойчивые конструкции с прямым указанием на эмоцию (например, *мне грустно*).

Шаховский (2009) отмечает, что эмоциональную лексику можно разделить по следующим группам в зависимости от способа выражения эмоции:

- прямая номинация эмоции (непосредственное указание на эмоцию – *радость, гнев*);
- описание эмоции (жесты, взгляд, характеристики речи и голоса – то есть указание на «симптомы» эмоции);
- непосредственное выражение эмоции (междометия, инвективная лексика и др.).

Место эмоционального компонента в семантике слова также может быть различно. Шаховский (2008) вводит систему терминов для описания трех типов статуса эмоционального семантического компонента – *аффектив* (собственно эмотивность), *коннотатив* (эмотивность как одна из реализаций семантики слова), *потенциатив* (ситуативная эмотивность) (Шаховский 2008: 53–127).

В работах, относящихся к области компьютерной лингвистики, как правило, предлагается несколько более упрощенный подход к описанию эмоциональной лексики. Это объясняется тем, что зачастую задачи компьютерной лингвистики подразумевают обработку значительного количества данных и по крайней мере частичную автоматизацию многих процессов. Например, в (Mohammad 2023) выделяются две категории эмоциональных слов: слова, которые включают эмоциональный компонент в ядро значения (*core meaning*), а также те, что связаны с эмоциями лишь посредством коннотаций (например, слово *вор* может ассоциироваться со злостью, а слово *солнце* с радостью). Обе категории автор рассматривает как релевантные для формирования

эмоциональных лексиконов, которые в дальнейшем можно использовать для различных задач компьютерной лингвистики.

В настоящей работе эмоциональная лексика так же понимается достаточно широко, поскольку доступные источники на русском языке включают эмоциональные слова различных типов – как прямые указания на эмоциональные состояния (*радость, злость*), так и лексику, связанную с эмоцией лишь ассоциативно (*праздник* для класса «радость», *мошенник* для класса «злость»).

Отдельную сложность представляет собой деление эмоциональных слов на классы по типу выражаемой эмоции. При анализе эмоциональной лексики исследователи, как правило, опираются на модели классификации эмоций, разработанные в психологии (например, Izard 1991). Существует большое количество подходов к классификации эмоций, при этом в области обработки естественного языка предпочтение, как правило, отдается работам Пола Экмана (Ekman 1992) и Роберта Плутчика (Plutchik 1980). Например, в работах (Wang et al. 2020, Park et al. 2020) для классификации текстов по эмоции были использованы «эмоциональные векторные представления» (emotional embeddings) на основе колеса эмоций Роберта Плутчика.

Деление эмоций на классы имеет практическую значимость для задач обработки естественного языка. Как правило, распознавание эмоций в тексте подразумевает и их классификацию, требуется не только выявить, выражается ли в тексте та или иная эмоция, но и определить, какая именно это эмоция. Анализ эмоций (emotion analysis) традиционно подразумевает классификацию на большее количество категорий, нежели анализ тональности (sentiment analysis), где в случае классической постановки задачи присутствуют три категории тональности – позитивная, негативная или нейтральная. Например, один из наиболее известных наборов данных, используемых для классификации текстов по типу эмоции – GoEmotions (Demszky et al. 2020), включает 27 эмоциональных классов.

В данной работе рассматривается эмоциональная лексика, относящаяся к базовым эмоциональным категориям: радость, злость, грусть, страх, удивление. Такое решение обосновывается двумя факторами: во-первых, эти классы эмоций наиболее последовательно представлены в доступных нам русскоязычных источниках эмоциональной лексики; во-вторых, именно эти классы эмоций являются наиболее распространенными в исследованиях, относящихся к области обработки естественного языка, согласно статистике, данной в обзоре (Plaza-del-Arco et al. 2024).

Таким образом, класс эмоциональной лексики достаточно широк, лексику можно классифицировать по большому количеству оснований. Определение границ этого тематического класса представляет собой отдельную исследовательскую задачу. В рамках данной статьи предлагается методика создания эмоционального лексикона на основе существующих русскоязычных ресурсов, которая не предполагает отбора лексики вручную.

3. Русскоязычные ресурсы эмоциональной лексики

В данном разделе будут представлены русскоязычные источники эмоциональной лексики. Первая часть раздела посвящена русскоязычным словарям и тезаурусам: «Русский семантический словарь» под ред. Н.Ю. Шведовой; «Алфавит эмоций: тезаурус эмотивной лексики» Л.Г. Бабенко; лексикон оценочных слов и выражений русского языка «РусСентиЛекс» (Loukachevitch & Levchik 2016); а также данным Национального корпуса русского языка (НКРЯ). Во второй части речь пойдет о списках, где каждому слову присвоен эмоциональный рейтинг на основе оценки респондентов: ресурс NRC EmoLex (Mohammad, Turney 2013), база данных ENRuN (Сысоева, Люсин 2024). Третья часть посвящена наборам данных, содержащим тексты, классифицированные по категориям в зависимости от типа эмоции, которая наиболее явно выражена в тексте.

3.1. Словарные и корпусные данные

Наиболее традиционные с лингвистической точки зрения источники для формирования эмоционального лексикона – лексикографические ресурсы и корпусные данные.

«Русский семантический словарь» (ред. Н.Ю. Шведова) (Шведова 1998) представляет собой систематизированный по классам слов толковый словарь русской общеупотребительной лексики. Словарь содержит только имена существительные. Раздел, представляющий интерес для составления эмоционального лексикона, – «Сами чувства, эмоциональные состояния» – включает 235 лексических единиц. Характеризуя стилистику лексики, представленной в источнике, отметим, что раздел включает 12 единиц с пометой «разг.» (*грустинка, кайф*); 8 с пометой «книжн.» (*ненавистничество, нега*), 9 с пометой «устар.» (*сплин, хлад*).

«Алфавит эмоций: тезаурус эмотивной лексики» Л.Г. Бабенко (Бабенко 2022) – специализированный тезаурус списочного типа, где слова разделены на группы в зависимости от того, к какой эмоции они относятся. Всего тезаурус содержит 11 210 лексических единиц различной частеречной принадлежности. Внутри групп лексика дополнительно делится на подгруппы: *эмоциональное состояние; эмоциональное отношение; становление эмоционального состояния и отношения; эмоциональное воздействие; внешнее выражение эмоций; эмоциональная характеристика; эмоциональное качество; человек как средоточие и носитель эмоций*. При этом следует отметить, что эмоциональная лексика понимается автором достаточно широко: выделяются 39 базовых эмоций. Наряду с классическими эмоциями (например, *грусть, страх, удивление*) представлены не вполне стандартные для этого семантического поля категории лексики, например, *настойчивость* или *искренность*, скорее обозначающие личностные характеристики, нежели эмоциональное состояние или отношение. Слова, разделенные на классы,

снабжаются лексикографическими пометами. В том числе в словаре представлена и разговорно-сниженная лексика (*балдеть, везуха, дрейфить*).

Лексикон оценочных слов и выражений русского языка «РуСентиЛекс» (Loukachevitch, Levchik 2016) включает более 12 тысяч единиц. Эмоциональная лексика не делится на классы, а объединяется под общим тегом «feeling», который включает 1760 единиц. Присутствует разделение слов и выражений по тональности: положительная – «positive» (*ароматный, блаженство, веселье*), негативная – «negative» (*вешать нос, испуганный, беспокойство*), смешанная – «positive/negative» (*волнение, жаждать*).

Данные Национального корпуса русского языка (НКРЯ). В настоящий момент в НКРЯ (Савчук и др. 2024) для той части корпуса, которая имеет семантическую разметку, доступна функция поиска по лексико-семантическим признакам. Можно сделать выгрузку лексических единиц с семантической пометой «эмоция». Как и в случае с РуСентиЛекс, слова имеют общий тег и не размечены по классам. Выгрузка включает глаголы (299 единиц) и имена существительные (147 единиц).

Безусловное преимущество словарных и корпусных данных заключается в том, что в них, как правило, включена лексика, формирующая ядро семантического класса эмоций – непосредственное указание на эмоциональное состояние (*огорчиться, радость, злой*), описание его характерных проявлений (*нахмуриться, заплакать, улыбнуться*). При этом доминирует пласт лексики, относящийся к литературному языку, в то время как сленговая лексика или междометия охвачены менее полно, что представляется совершенно естественным для такого типа ресурсов. Однако именно в контексте направления распознавания эмоций доминанту литературного языка можно расценивать и как недостаток, поскольку разговорная лексика может оказаться важна, например, при анализе текстов из социальных сетей. Еще одним недочетом может считаться то, что не всегда эмоциональная лексика разделена на классы, а в случаях, когда разделена, то зачастую используются не общепринятые, упрощенные модели деления на категории, характерные для компьютерной лингвистики. Таким образом, использовать словарные и корпусные ресурсы без дополнительной обработки не представляется возможным.

3.2. Эмоциональная оценка слов-стимулов

Помимо традиционных лингвистических ресурсов можно использовать данные опросов, где респондентам предлагается соотнести слово с какой-либо эмоцией.

«NRC Emotion Lexicon» (Mohammad, Turney 2013) – англоязычный ресурс, созданный с помощью краудсорсинга, на данный момент включает более 14 000 единиц различной частеречной принадлежности. Слова были размечены по 8 базовым эмоциям по модели (Plutchik 1980): классы «злость», «страх», «предвкушение», «доверие», «грусть», «радость», «отвращение».

Кроме того, слова оценивались респондентами по тональности как позитивные или негативные. Оценка по эмоциям и тональности производилась по бинарной шкале 0/1. Одно слово могло помещаться в несколько классов или не входить ни в один из предложенных, если получит нулевые оценки по всем категориям. Этот ресурс включает как непосредственные указания на эмоции или их проявления (*шок* для класса «удивление», *рыдать* для класса «грусть»), так и коннотации (*кораблекрушение* для класса «грусть», *герпес* для класса «отвращение»).

«NRC Emotion Lexicon» был автоматически переведён на 108 языков, в том числе и на русский. Из-за специфики автоматического перевода в русскоязычном варианте лексикона могут встречаться одинаковые слова, которые при этом имеют различные оценки. Так, в оригинальной версии респондентам для разметки по эмоциям предлагались три различных слова – *merciless*, *ruthless*, *unrelenting*, на русский все они были переведены одинаково – *безжалостный*. Соответствующий фрагмент лексикона представлен в табл. 1.

Таблица 1. Пример из версии NRC Emotion Lexicon с переводом на русский язык
Table 1. Example from NRC Emotion Lexicon translated into Russian

English Word	Anger	Anticipation	Disgust	Fear	Joy	Negative	Positive	Sadness	Surprise	Trust	Russian Word
merciless	0	0	0	1	0	1	0	0	0	0	безжалостный
ruthless	1	0	1	1	0	1	0	0	0	0	безжалостный
unrelenting	0	0	0	0	0	0	0	0	0	0	безжалостный

Встречаются и явные ошибки в переводе, особенно разговорной лексики, которая представляют проблему при использовании автоматического перевода (Озюменко, Ларина 2025). Например, слово *john* переведено как имя собственное *Джон*, хотя это слово имеет значение ‘туалет’. У респондентов слово *john* ассоциируется с классами «disgust» и «negative».

Таким образом, автоматически переведенные данные следует использовать с осторожностью: обращать внимание на дубликаты и погрешности перевода.

База данных ENRuN (Emotional Norms for Russian Nouns) (Сысоева, Люсин 2024) содержит эмоциональные оценки 1800 существительных русского языка. Для каждого слова представлены средние значения, стандартные отклонения, минимальные и максимальные оценки по каждому параметру, а также количество человек, оценивших данное слово. Для нашего исследования интерес представляет прежде всего категориальная оценка существительных, то есть степень связи той или иной лексической единицы со следующими классами эмоций: «радость», «грусть», «злость», «страх», «отвращение». Степень эмоциональной окраски оценивалась респондентами по

шкале от 0 до 5. Ресурс включает как прямые номинации эмоций (*веселье, уныние*) или те или иные способы выражения эмоций (*улыбка, слёзы*), так и слова, указывающие на предметы или явления действительности, связанные с той или иной эмоцией исключительно ассоциативно (*виселица* для класса «страх»; *королева* для класса «радость»).

3.3. Наборы данных для автоматического распознавания эмоций в текстах

Еще одним ресурсом могут послужить размеченные наборы данных (датасеты), используемые для классификации текстов по эмоциям⁴. Сами по себе эти ресурсы не содержат готового списка эмоциональной лексики. Тем не менее, эти ресурсы представляют большой интерес для исследователей из области обработки естественного языка. Кроме того, на основе этих наборов данных можно получить список эмоциональной лексики. Наш подход к получению списка размеченных эмоциональных слов путем обучения линейных классификаторов представлен в п. 4.1 данной статьи. Однако прежде всего рассмотрим доступные данные.

В открытом доступе представлены следующие датасеты, содержащие размеченные по классам эмоций тексты на русском языке: CEDR (Corpus for Emotions Detecting in Russian-language text sentences of different social sources) (Sboev et al. 2021), доступная часть набора данных AIST (Kazyulina et al. 2021), русскоязычная часть набора данных для соревнования SemEval-2025⁵ (Muhammad et al. 2025); датасет GoEmotions (Demszky et al. 2020), автоматически переведенный на русский язык.

В датасетах представлены в основном короткие тексты в формате комментария из социальных сетей. Тексты обладают характерной спецификой: содержат большое количество разговорной, сленговой лексики, эмотиконов, хэштегов, встречаются также орфографические ошибки или опечатки. Примеры текстов с меткой класса «Грусть» из обучающей выборки набора данных SemEval-2025 (Muhammad et al. 2025). В примере (1) используется англицизм, в примере (2) – опечатка и эмотикон.

(1) *Сижу и жру пельмени, пока гости не пришли. Худший НГ эвер.*

(2) *Вером редко когда хорошее настроение =(((*

В табл. 2 дана общая информация о наборах данных.

⁴ Существуют также наборы данных, нацеленные на задачу распознавания эмоций в разговорной речи, например, DUSHA (Kondratenko et al. 2022), однако в данной работе они не принимались во внимание.

⁵ Ежегодные соревнования по семантическому анализу, которые проводятся в рамках международного исследовательского семинара Semantic Evaluation (<https://semeval.github.io/>).

Таблица 2. Наборы данных для классификации текстов по эмоциям на русском языке
Table 2. Datasets for Emotion Classification in Russian

Набор данных	Классы эмоций	Количество размеченных текстов
CEDR	радость, грусть, удивление, страх, злость	9 410
AIST	радость, грусть, злость, неуверенность, нейтральность	50 750
SemEval-2025	радость, грусть, удивление, страх, злость, отвращение	2 878
ruGoEmotions	27 эмоций + нейтральность	58 000

4. Методика исследования

Для создания RusEmoLex на основе каждого доступного источника были сформированы исходные списки эмоциональной лексики. Затем списки, полученные на основе каждого источника, были объединены. Далее из полученного объединенного списка по ряду критериев производился отбор лексики в RusEmoLex.

Для включения в RusEmoLex лексическая единица должна удовлетворять следующим критериям:

- **количество источников:** лексическая единица должна входить в два или более источников;
- **качество источников:** как минимум один из источников, куда вошла лексическая единица, должен относиться к словарным или корпусным ресурсам и не являться переводным;
- **наличие метки класса:** лексическая единица должна иметь установленный на основании большинства источников эмоциональный класс.

Предлагаемая методика нацелена на то, чтобы формализовать процедуру отбора лексики, сделав ее воспроизводимой и менее зависимой от субъективных суждений разметчика. Далее рассмотрим каждый этап более подробно.

4.1. Отбор лексики из каждого источника

На основе доступных источников были составлены списки эмоциональных слов. Для разных источников методика отбора слов в RusEmoLex различается, поскольку и сами исходные ресурсы отличаются с точки зрения формата и области применения. Лексикографические и корпусные данные требовали минимальной обработки, в то время как формирование списка слов на основе наборов данных, используемых для машинного обучения, потребовало больше этапов обработки.

В список на основе Русского семантического словаря были включены все 235 единиц из раздела «Сами чувства, эмоциональные состояния». Классы лексики представлены в табл. 3.

**Таблица 3. Список эмоциональных слов, созданный
на основе Русского семантического словаря**
Table 3. Emotion word list from The Russian Semantic Dictionary

Класс	Количество единиц
обида, досада, недовольство	23
отвращение, презрение, зависть, неприязнь	28
злоба, гнев, рассерженность	27
сочувствие, жалость	7
страдание, раскаяние; смущение, стыд	22
отчаяние, печаль, уныние	34
восторг, радость, веселье; довольство, умиротворение	30
страх, тревога; нетерпение	22
любовь, нежность, расположение, привязанность; благодарность	42
Всего	235

В список эмоциональных слов на основе ресурса «Алфавит эмоций: тезаурус эмотивной лексики» (Бабенко 2022) были отобраны только те группы слов, относящиеся к пяти базовым эмоциям, которые находятся в центре внимания в данной статье. Во-первых, были выбраны одноименные категории – радость, грусть, злость, страх, удивление. Во-вторых, были также включены категории «горе» (можно интерпретировать как сильную грусть) и «счастье» (можно интерпретировать как сильную радость). Список включает 3241 слово. Классы лексики, включенной в список, даны в табл. 4.

**Таблица 4. Список эмоциональных слов, созданный
на основе ресурса «Алфавит эмоций: тезаурус эмотивной лексики»**
Table 4. Emotion word list from “Alphabet of Emotions: Thesaurus of Emotive Lexis”

Класс	Количество единиц
радость	549
счастье	263
грусть	624
горе	742
злость	451
страх	293
удивление	319
Всего	3241

В списки слов, созданных на основе источников РуСентиЛекс и НКРЯ, где эмоции не делятся на классы, были включены все слова, которым присвоены теги «feeling» и «эмоция» соответственно. Список эмоциональных слов и выражений на основе РуСентиЛекс составил 1760 единиц. Список слов на основе НКРЯ составил 466 единиц.

Для формирования списка эмоциональных слов на основе NRC Emotion Lexicon отбирались такие слова, которые по целевому классу (все категории,

кроме «anticipation» и «trust») имели отметку 1. Количественные данные по списку представлены в табл. 5.

Таблица 5. Список эмоциональных слов, созданный на основе русскоязычной версии NRC Emotion Lexicon
Table 5. Emotion word list from NRC Emotion Lexicon translated into Russian

Класс	Количество единиц
радость	687
страх	1474
удивление	532
отвращение	1056
грусть	1187
злость	1245
Всего	6181

Для формирования списка слов на основе базы данных ENRuN были отобраны такие существительные, которые получили оценку респондентов от 2, 5 (то есть половина или более по шкале от 0 до 5) по целевой эмоции. Список включает 808 лексических единиц. Количественные данные по списку представлены в таблице 6.

Таблица 6. Список эмоциональных слов, созданный на основе ENRuN
Table 6. Emotion word list from ENRuN

Класс	Количество единиц
радость	317
страх	161
отвращение	131
печаль	111
злость	88
Всего	808

Для создания списка эмоциональной лексики на основе датасетов CEDR (Sboev et al. 2021), AIST (Kazyulina et al. 2021), SemEval-2025 (Muhammad et al. 2025), GoEmotions (Demszky et al. 2020) был сформирован общий набор данных, включающий лемматизированные тексты из всех датасетов и метки классов.

Далее для обработки данных были использованы методы машинного обучения, которые позволяют оценить влияние каждого отдельного слова на определение общего эмоционального содержания текста. Оценка реализуется путем автоматического подбора коэффициентов к частотным характеристикам слов на обучающей выборке. На общем наборе данных были обучены два линейных классификатора – логистическая регрессия и метод опорных векторов. После обучения моделей можно получить информацию о весе каждого используемого признака, в данном случае в роли признаков

выступают слова из текстов датасета. Например, три слова-признака с наибольшим весом для класса «Злость» в датасете CEDR при обучении модели-классификатора (логистическая регрессия): *злость* 5.068; *злиться* 4.963; *злой* 4.760. Именно эти слова оказались наиболее значимы для определения текста в класс «Злость».

На основе двух классификаторов были получены два списка наиболее значимых слов-признаков. Затем для дополнительной валидации полученные списки слов были объединены (с учетом общих элементов), а слова-признаки упорядочены в соответствии с усредненным рангом. В результате получился список, включающий 1422 слова.

Три наиболее значимых слова-признака (исключая обценную лексику, которая оказалась характерна для класса «Злость») для каждой эмоции представлены в табл. 7.

Таблица 7. Примеры из списка эмоциональных слов, полученных из датасетов
Table 7. Examples from dataset-based list

Метка класса	Слова-признаки
радость	рад, счастливый, радость
страх	бояться, пугать, страшный
злость	(...), бесить, ненавидеть, (...), (...), злость
удивление	удивить, недоумение, шокировать
грусть	печальный, грустный, сожаление

Полученный на основе датасетов список представляет интерес, так как включает ряд разговорных и сленговых слов (*пасиб, лапочка*) и междометий (*ахахах, вау, оу*), которые не встречались в описанных выше источниках. Однако список нельзя считать готовым к использованию, поскольку данные можно охарактеризовать как зашумленные, среди слов-признаков с достаточно высоким весом встречаются, например, имена собственные (*Владислав, Кристина*), фрагменты устойчивых словосочетаний (слово *рождение* из-за того, что в текстах датасетов встречались поздравления с днем рождения).

**4.2. Создание единого списка эмоциональной лексики
на основе всех источников**

На основе доступных русскоязычных источников было составлено семь списков слов, которые затем были объединены. В итоге был получен список из 7937 уникальных лемм. Количественное распределение слов представлено в табл. 8.

Следует отметить, что уровень пересечения слов в списках из различных источников достаточно низкий, имеются серьезные различия как в количественном, так и в качественном аспекте. Самый большой исходный список слов был сформирован на базе NRC EmoLex (6 181 единиц), а самый маленький – на основе Русского семантического словаря (235 единиц).

Различен и лексический состав списков, поскольку сами ресурсы также сильно расходятся с точки зрения цели и методики их создания.

Таблица 8. Количество слов в списках на основе различных источников
Table 8. Number of words in lists based on different sources

Источник	Количество единиц в списке
Русский семантический словарь	235
Алфавит эмоций: тезаурус эмотивной лексики	3 241
НКРЯ	466
РуСентиЛекс	1 760
NRC Emotion Lexicon	6 181
ENRuN	808
Наборы данных для машинного обучения	1 422
Общее количество уникальных лемм	7 937

На рис. 1 представлена матрица пересечения источников (степень совпадения между источниками дана в процентах).

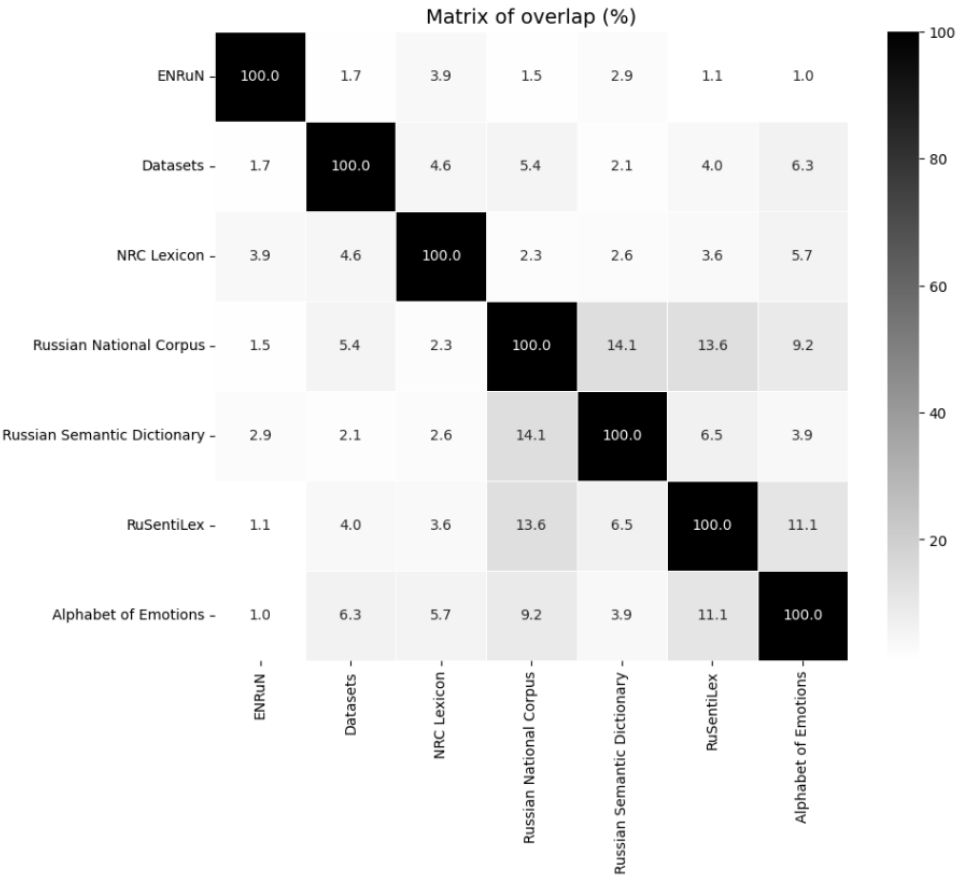


Рис. 1. Матрица пересечения
Figure 1. Matrix of overlap

Для создания эмоционального лексикона объединенный список слов прошел несколько этапов обработки, которые будут описаны в следующем разделе.

4.3. Отбор лексики в RusEmoLex

Количество источников. Как было сказано выше, лексических единиц, которые входили бы в большое количество источников, не так много. В исходном списке слов видим, что в один источник входит 6663 слов; в два источника – 887; в три источника – 251; в четыре источника – 93; в пять источников – 20; в шесть источников – 20; в 7 источников – 3.

Для дальнейшей обработки были отобраны только такие слова, которые входят как минимум в два источника, что позволило отсеять наименее релевантную лексику. Таким образом, количество слов-кандидатов сократилось с 7937 до 1274.

Качество источников. Источники, на основе которых составлялся список слов-кандидатов, были созданы для различных целей. Так, при формировании лексикографических ресурсов специалистами отбирались слова, непосредственно связанные с эмоциями. Совершенно другой принцип лежит в основе формирования списков слов-стимулов для определения их эмоциональной окраски. Например, (Сысоева, Люсин 2024) в ходе разработки базы данных ENRuN список существительных для опроса респондентов составляли на основе частотного словаря русского языка (Ляшевская, Шаров 2009), отбирая лексику оттуда по определенным критериям (например, длина слова). Таким образом, в источниках ENRuN и NRC Emotion Lexicon лексика разнообразная: это могут быть как указания на эмоции и их проявления (то, что представляет для нас основной интерес в рамках исследования) или же слова, которые связаны с эмоциями лишь ассоциативно. Напомним также, что NRC Emotion Lexicon был переведен на русский язык автоматически, поэтому содержит неточности.

Списки слов-признаков, полученные в результате обучения линейных классификаторов на наборах данных, ориентированных на распознавание эмоций в тексте, характеризуются высоким уровнем шума: наряду с целевой лексикой туда могли попасть и случайные слова, не имеющие отношения к эмоциям. Кроме того, один из датасетов (RuGoEmotions) исходно является англоязычным, что также может негативно сказываться на качестве слов-признаков.

Таким образом, было принято решение разделить источники на категории А и В в зависимости от их надежности. В категорию А были включены словари и данные корпуса, в категорию В – данные опросов про ассоциации слов с эмоциями (NRC Emotion Lexicon, ENRuN) и наборы данных для машинного обучения.

Лексические единицы в списке слов-кандидатов в результате были распределены на 3 группы: группа слов А – входят только в источники категории

А; группа слов В – входят только в источники категории В; группа слов АВ – входят в источники обеих категорий. Количественное распределение представлено в табл. 9.

Таблица 9. Распределение слов по категории источников
Table 9. Distribution of words by source category

Группа по категории источника	Количество лексических единиц в группе
А	477
АВ	595
В	202

Принимая во внимание качественное различие источников, в список слов-кандидатов для дальнейшего рассмотрения были включены только такие лексические единицы, которые есть хотя бы в одном источнике категории А – то есть слова групп А и АВ.

Таким образом, количество слов-кандидатов сократилось с 1274 до 1072.

Присвоение меток классов. Несмотря на то, что перечень эмоций, интересующих нас в рамках исследования, был определен заранее, при классификации отобранных слов возник ряд сложностей. Во-первых, некоторые источники высокой степени надежности (категория А) не предполагают деление на классы – в выгрузке из НКРЯ классы отсутствуют, в РуСентиЛекс есть только деление по тональности (позитивная, негативная, смешанная). Во-вторых, классы в источниках различаются.

Были составлены правила по соотношению классов из различных А-источников. Сведение более сложных описаний классов (как, например, в Русском семантическом словаре) к более базовым предполагает некоторое упрощение, однако представляется адекватным задачам настоящего исследования.

Рассмотрим принятое соотношение классов в табл. 10.

Таблица 10. Соотношение классов в различных источниках
Table 10. Class correspondence in different sources

Источник	Класс источника	Базовый класс
Русский семантический словарь	обида, досада, недовольство	Злость
Русский семантический словарь	отвращение, презрение, зависть, неприязнь	Злость
Русский семантический словарь	злоба, гнев, рассерженность	Злость
Русский семантический словарь	сочувствие, жалость	Грусть
Русский семантический словарь	страдание, раскаяние; смущение, стыд	Грусть
Русский семантический словарь	отчаяние, печаль, уныние	Грусть
Русский семантический словарь	страх, тревога; нетерпение	Страх
Русский семантический словарь	восторг, радость, веселье; довольство, умиротворение	Радость

Источник	Класс источника	Базовый класс
Русский семантический словарь	любовь, нежность, расположение, привязанность; благодарность	Радость
Алфавит эмоций	Злость	Злость
Алфавит эмоций	Удивление	Удивление
Алфавит эмоций	Радость	Радость
Алфавит эмоций	Счастье	Радость
Алфавит эмоций	Страх	Страх
Алфавит эмоций	Грусть	Грусть
Алфавит эмоций	Горе	Грусть
РусСентиЛекс	positive	Радость

Наиболее проблемным с точки зрения соотношения классов оказался РусСентиЛекс, в особенности метка «negative», так как среди пяти рассматриваемых эмоций три являются негативными: метка «negative» может соответствовать любому из трех классов – «грусть», «злость» или «страх». По этой причине для РусСентиЛекс засчитывалась только метка «positive» как «радость».

После проведенной унификации классов разметка единиц производилась по следующим принципам:

1. Для единиц из группы А метка класса выбиралась в соответствии с наиболее частотным классом по всем источникам.

2. Для единиц из группы АВ при определении класса мы также ориентировались исключительно на метки источников категории А.

Рассмотрим пример в табл. 11. Слово *удовлетворение* встречается в 4 источниках категории А. НКРЯ не содержит меток классов. В остальных источниках представлены классы, сводимые к категории «Радость». Таким образом, слову присваивается итоговая метка «Радость».

Таблица 11. Пример определения итоговой метки класса для RusEmoLex
Table 11. Example of defining the resulting class label for RusEmoLex

Лексическая единица	НКРЯ	Русский семантический словарь	РусСентиЛекс	Алфавит эмоций
удовлетворение	-	восторг, радость, веселье; довольство, умиротворение	positive	Счастье

При невозможности определить метку класса единица не включалась в итоговый рекомендованный список эмоциональной лексики.

Невозможность присвоения метки класса возникала, например, если слово входило только в источники категории А, не подразумевающие деления на классы (НКРЯ, РусСентиЛекс). В ряде случаев выявлялись противоречия между источниками: лексической единице присваивались разные классы. При этом отметим, что возникающие в источниках противоречия могли быть

связаны со свойствами обозначаемой данной лексической единицей эмоции – в частности, с ее сложным, неоднозначным характером. Например, *страсть* в Русском семантическом словаре относится к классу «Любовь, нежность, расположение, привязанность; благодарность» (соответствует классу «радость»), а согласно (Бабенко 2022) – к классу «страх». Такой результат представляется вполне релевантным, *страсть* можно охарактеризовать как сложную, смешанную эмоцию⁶. Рассмотрение подобных случаев не входит в задачи данной работы, однако представляет отдельный интерес и требует дальнейшего исследования.

Расширение списка слов за счет однокоренных слов. Полученный лексикон дополнялся с помощью поиска однокоренных слов: по каждой лексической единице, которая уже была включена в список, проводилась проверка на наличие однокоренных слов среди тех единиц, которые по тем или иным причинам в список отобраны не были (то есть среди единиц, которые не получили метку класса, а также слов, которые встречались исключительно в источниках типа В). Данный этап позволил сделать процедуру формирования эмоционального лексикона более последовательной.

Рассмотрим примеры. Слову *беспокойство* на основе анализа источников была присвоена метка «Страх», в то же время для слова *беспокойный* метка класса не была назначена, это слово относится к группе АВ и входит в два источника – РуСентиЛекс (категория А) и NRC EmoLex (категория В) – метка класса присваивается по источнику категории А, но в данном случае это не представляется возможным, так как в РуСентиЛекс это слово отмечено как «negative», соответственно нельзя определить, какая именно категория подразумевается – страх, грусть или злость. Однако мы смогли присвоить метку класса слову *беспокойство*, это позволяет нам распространить ее и на однокоренные слова, частности на слово *беспокойный*.

При проверке по однокоренным словам во внимание принимались также и аффиксы. В том случае, если аффикс значительно модифицировал семантику слова, метка класса по аналогии с однокоренным словом не присваивалась. Прежде всего, это приставки со значением отрицания: при наличии такой приставки в одном из слов однокоренные слова могут относиться к различным классам сообразно их значению. Например, *нелюбовь* – класс «Злость», *любить* – класс «Радость». Рассмотрим другой пример, где приставка кардинальным образом не меняет значение слова: глаголу *скупать* на основании источников была присвоена метка класса «Грусть», для глагола *заскупать* метку не удалось определить по источникам, но она была присвоена по аналогии с однокоренным словом *скупать*.

⁶ Более подробно про смешанные эмоции как объект лингвистического исследования см. (Ионова, Штеба 2019).

В ходе анализа списка были выявлены такие случаи, когда однокоренные слова имели различные метки классов и при отсутствии аффиксов, существенно влияющих на значение. Рассмотрим группу однокоренных слов, приведенную в табл. 12.

Таблица 12. Пример с однокоренными словами
Table 12. Example of the word family

Лексическая единица	Метка класса на основании источников
<i>обида</i>	Грусть
<i>обидеть</i>	Не присвоена
<i>обидеться</i>	Грусть
<i>обидный</i>	Грусть
<i>обижаться</i>	Грусть
<i>обиженность</i>	Злость
<i>разобидеться</i>	Не присвоена

Двум единицам из группы метку класса на основании источников присвоить не удалось. Другие однокоренные слова на основании источников имеют различные метки: «Злость» и «Грусть». В таких случаях итоговая метка класса определялась по большинству. В данном примере мы видим одно слово с меткой «Злость» и четыре слова с меткой «Грусть». Итоговой меткой для всех слов этой группы будет «Грусть», поскольку именно «Грусть» зафиксирована в большинстве источников. В том случае, если определить метку по большинству не представлялось возможным, – например, они делились поровну между двумя классами – слова исключались из лексикона из-за наличия неустранимого в рамках нашего подхода противоречия.

После проведения всех описанных этапов мы получили список из 1024 единиц, размеченных по классам эмоций.

5. Результаты исследования

Основным результатом исследования является полученный русскоязычный эмоциональный лексикон RusEmoLex, который включает 1024 лексические единицы из источников различного типа.

По каждому слову в RusEmoLex доступна следующая информация: (1) частеречная принадлежность; (2) эмоциональный класс; (3) категория источников, где встречается слово; (4) количество вхождений в источники; (5) перечисление названий исходных источников, где зафиксировано слово.

Такая структура RusEmoLex позволяет в зависимости от целей и задач исследования отфильтровать и упорядочить лексические единицы по нужному критерию – например, если требуется рассмотреть слова, вошедшие в наибольшее количество источников, или лексику определенного эмоционального класса. По умолчанию слова в лексиконе упорядочены по алфавиту. Структура RusEmoLex проиллюстрирована примером в табл. 13.

Таблица 13. Структура RusEmoLex
Table 13. Structure of RusEmoLex

Слово	Часть речи	Класс	Категория источников	Количество вхождений	Источники
<i>веселье</i>	Существительное	радость	AB	6	ENRuN, NRC EmoLex, НКРЯ, Русский семантический словарь, РуСентиЛекс, Алфавит эмоций

Рассмотрим ключевые характеристики слов, которые вошли в состав RusEmoLex. С точки зрения частеречной принадлежности наиболее частотны оказались глаголы и существительные. В лексикон попало имя существительное с предлогом – *без ума*, остальные же единицы состоят из одного элемента. Выражение *без ума* встречается в двух источниках – РуСентиЛекс и NRC EmoLex. РуСентиЛекс, как было сказано выше, включает не только отдельные слова, но и выражения. NRC EmoLex является переводным ресурсом: то, что было автоматически переведено как *без ума*, в оригинале является именем прилагательным *mad*. В табл. 14 представлены количественные данные по всем частям речи, которые вошли в лексикон.

Таблица 14. Частеречное распределение слов в лексиконе
Table 14. PoS distribution in the lexicon

Часть речи	Количество единиц
Глагол	473
Имя существительное	336
Имя прилагательное	189
Наречие	25
Имя существительное с предлогом	1

Распределение слов по классам эмоций представлено на рис. 2. Классы не сбалансированные, такой результат является закономерным, так как в исходных источниках классы, как правило, также не были сбалансированными.

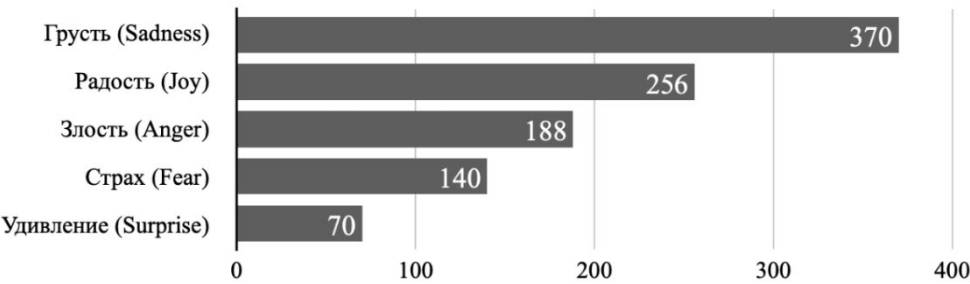


Рис. 2. Распределение слов в лексиконе по эмоциональным классам
Figure 2. Emotion classes distribution in the lexicon

Характеристика слов в лексиконе по категории источников дана в табл. 15. Наименее представлены в RusEmoLex слова из источников категории В, так как они могли попасть в итоговый список только на этапе проверки по однокоренным.

Таблица 15. Распределение слов в лексиконе по категории источников
Table 15. Source category distribution in the lexicon

Категория источника	Количество единиц
А	451
АВ	557
В	16

Характеристика слов по количеству вхождений в различные источники дана в табл. 16.

Таблица 16. Распределение слов в лексиконе по количеству вхождений в источники
Table 16. Emotion words at the intersection of resources

Количество вхождений в источники	Количество единиц
2	653
3	239
4	90
5	19
6	20
7	3

Рассмотрим фрагмент лексикона, где представлены слова, вошедшие в наибольшее количество источников (табл. 17).

Таблица 17. Фрагмент RusEmoLex
Table 17. Fragment of RusEmoLex, which includes the lexical units found in the largest number of sources

Слово	Часть речи	Класс	Категория источников	Количество вхождений
<i>сожаление</i>	существительное	Грусть	АВ	7
<i>удовольствие</i>	существительное	Радость	АВ	7
<i>ужас</i>	существительное	Страх	АВ	7
<i>веселье</i>	существительное	Радость	АВ	6
<i>восторг</i>	существительное	Радость	АВ	6
<i>горе</i>	существительное	Грусть	АВ	6
<i>грусть</i>	существительное	Грусть	АВ	6
<i>злоба</i>	существительное	Злость	АВ	6
<i>злость</i>	существительное	Злость	АВ	6
<i>испуг</i>	существительное	Страх	АВ	6
<i>кошмар</i>	существительное	Страх	АВ	6
<i>ликование</i>	существительное	Радость	АВ	6
<i>отчаяние</i>	существительное	Грусть	АВ	6

Слово	Часть речи	Класс	Категория источников	Количество вхождений
<i>паника</i>	существительное	Страх	АВ	6
<i>печаль</i>	существительное	Грусть	АВ	6
<i>радость</i>	существительное	Радость	АВ	6
<i>скорбь</i>	существительное	Грусть	АВ	6
<i>скука</i>	существительное	Грусть	АВ	6
<i>страх</i>	существительное	Страх	АВ	6
<i>тоска</i>	существительное	Грусть	АВ	6
<i>тревога</i>	существительное	Страх	АВ	6
<i>хандра</i>	существительное	Грусть	АВ	6
<i>ярость</i>	существительное	Злость	АВ	6

Примечательно, что слова, которые встречаются во всех семи источниках (*удовольствие, сожаление, ужас*), нельзя охарактеризовать как наиболее типичные и репрезентативные. На пересечение всех семи множеств не попали, например, базовые обозначения эмоций (*грусть, страх, злость* – эти слова вошли только в шесть источников). Отметим, что схожая тенденция была зафиксирована в работе (Котельников 2020), где рассматривались словари оценочной лексики для английского и русского языков. Так, для шестнадцати англоязычных словарей общими оказались слова *pretty, hell, hurt, sick*. В это множество не попали слова *good* и *bad*. Пересечение одиннадцати русскоязычных ресурсов дало пустое множество.

Таким образом, количество вхождений в источники не является единственным определяющим критерием для характеристики лексической единицы относительно ее принадлежности к ядру или периферии того или иного тематического класса – в нашем случае, класса эмоциональной лексики.

По стилистическим характеристикам слова, представленные в лексиконе, не размечались. Однако можно сказать, что лексический состав достаточно разнообразен: встречается как книжная или устаревшая лексика (*возликовать, кручина*), так и более сниженная, разговорная (*дрейфить, дуться*). Тем не менее, поскольку при создании лексикона мы ориентировались на более надежные источники (категория А), разговорная лексика представлена не очень широко, местоимения не представлены совсем, хотя они присутствовали в списке слов, созданном на основе наборов данных для машинного обучения.

6. Дискуссия

В статье представлен новый ресурс – RusEmoLex, созданный на основе доступных русскоязычных источников. Появление нового ресурса на русском языке особенно значимо в условиях доминирования в области обработки естественного языка англоязычных источников.

К преимуществам полученного ресурса можно отнести следующие пункты. Во-первых, лексикон содержит слова из различных источников,

которые теперь представлены в едином формате, что обеспечивает удобство использование нового ресурса. Во-вторых, преимуществом описанной в работе методики составления лексикона является ее формализованный характер. Такой подход позволяет повысить уровень объективности при отборе и классификации лексических единиц, поскольку мы опираемся не на один, а сразу на несколько ресурсов. Кроме того, предложенная в статье методика создания лексикона эмоциональных слов для русского языка в дальнейшем может применяться и для источников на других языках.

Полученный список эмоциональных слов нельзя считать окончательным. В дальнейшем он может быть дополнен новыми лексическими единицами и расширен за счет добавления новых классов эмоций (например, *отвращение, любовь, стыд*). Словарь может быть также пополнен путем выявления эмоциональной лексики в массивах текстов с помощью векторных представлений (эмбедингов). Однако следует отметить, что векторные представления слов зависят от корпуса, на котором они считаются, от применяемых моделей и их параметров, а также от особенностей контекстов в корпусе, включая вхождение слова в устойчивые словосочетания, это может приводить к ошибочным результатам (Wang 2020).

Тем не менее, на текущем этапе ресурс может использоваться в различных исследовательских задачах. Например, при создании новых русскоязычных наборов размеченных данных для анализа эмоций, а также при работе с большими языковыми моделями в рамках исследований по оценке понимания эмоций (emotional understanding) на материале русского языка.

7. Заключение

В данной статье были описаны русскоязычные ресурсы эмоциональной лексики, находящиеся в открытом доступе. Ресурсы были охарактеризованы по форме представления материала: лексикографические ресурсы (словари и тезаурусы); корпусные данные; данные, полученные в результате опросов (зафиксированные ассоциации слов с эмоциями); наборы данных, используемые в машинном обучении. Также источники были описаны с точки зрения лексического состава – частеречной принадлежности слов, стилистических характеристик и др.

Представлена методика создания нового русскоязычного ресурса RusEmoLex – объединенного списка эмоциональных слов, размеченных по частеречной принадлежности; эмоциональному классу; категории источников, где фиксируется слово; количеству вхождений в источники. Было показано, что RusEmoLex может использоваться как для собственно лингвистических задач, так и в целях обработки естественного языка. Текущая версия RusEmoLex включает 1024 лексических единицы, однако она не является окончательной, ресурс может быть расширен и дополнен путем включения новых источников или расширения списка классов эмоциональной лексики. Предложенная методика позволяет сделать процедуру отбора лексики более

объективной, а также в дальнейшем может быть использована и на материале других языков. RusEmoLex находится в открытом доступе⁷.

Информация о финансировании:

Данная работа была поддержана грантом, предоставленным Министерством экономического развития Российской Федерации в соответствии с Соглашением о предоставлении из федерального бюджета гранта в форме субсидии федеральному государственному бюджетному учреждению науки Институт системного программирования им. В.П. Иванникова Российской академии наук от 20 июня 2025 г. № 139-15-2025-011. Идентификатор соглашения 000000Ц313925P4G0002.

Благодарности:

Авторы благодарят аспирантку ВМК МГУ Петрову А.А. за содействие при обработке материала.

Участие авторов:

Ярошенко П.В. – сбор и обработка материалов; анализ полученных данных; написание текста; **Лукашевич Н.В.** – концепция и дизайн исследования; написание текста.

Список литературы / References

- Апресян В.Ю. Речевые стратегии выражения эмоций в русском языке // *Русский язык в научном освещении*. 2010. № 2. С. 26–57. [Apresyan, Valetina. Yu. 2010. Rechevye strategii vyrazheniya emotsii v russkom yazyke (Speech strategies for expressing emotions in Russian). *Russkiy yazyk v nauchnom osveshchenii* 2 (20). 26–57. (In Russ.)].
- Апресян Ю.Д. Избранные труды. Т. 2: Интегральное описание языка и системная лексикография. М.: Школа «Языки русской культуры», 1995. [Apresian, Yury D. 1995. Izbrannyye trudy. Integral'noe opisanie yazyka i sistemnaya leksikografiya (Selected Works. Integral Description of a Language and Systematic Lexicography). Vol. 2. Moscow: Shkola Iazyki russkoi kultury Publ., Moscow. (In Russ.)].
- Гладкова А.Н. Русская культурная семантика: Эмоции, ценности, жизненные установки. М.: Языки славянской культуры, 2010. [Gladkova, Anna 2010. Russkaya kul'turnaya semantika: Emotsii, tsennosti, zhiznennyye ustanovki (Russian Cultural Semantics: Emotions, Values, Attitudes). Vol. 2. Moscow: Iazyki slavianskoi kul'tury Publ. (In Russ.)].
- Иоанесян Е.Р. Исследования по семантике эмоциональных предикатов в типологическом аспекте. 2-е изд., расширенное и доп. М.; Ярославль: Издательство Канцлер, 2024. [Ioanesian, Evgenia R. 2024. Issledovaniya po semantike emotsional'nykh predikatov v tipologicheskom aspekte (Research on the Semantics of Emotional Predicates in the Typological Aspect). 2nd edn. Moscow; Yaroslavl: Kantsler Publ. (In Russ.)].
- Ионова С.В., Шаховский В.И. Проспекция лингвокультурологической теории эмоций Анны Вежбицкой // *Russian Journal of Linguistics*. 2018. Т. 22. No 4. С. 966–987. [Ionova, Svetlana & Viktor Shakhovsky. 2018. Anna Wierzbicka's linguocultural theory of emotions in the development dynamics. *Russian Journal of Linguistics* 22 (4). 966–987 (In Russ.)]. <https://doi.org/10.22363/2312-9182-2018-22-4-966-987>.

⁷ Текущая версия RusEmoLex доступна по ссылке: <https://github.com/nl-pi/rusemolex>.

- Ионова С.В., Штеба А.А. Смешанные эмоции: к вопросу о лингвистической репрезентации и метаязыке описания // *Вопросы психолингвистики*. 2019. № 2 (40). С. 63–81. [Ionova, Svetlana V. & Alexey A. Shteba. 2019. Contradictory Emotions: The question of the linguistic representation and the metalanguage description. *Journal of Psycholinguistics* 2 (40). 63–81. (In Russ.)].
- Ионова С.В. Эмоциональная доминанта текста: некоторые лингвистические аспекты исследования // *Вестник Волгоградского государственного университета*. Сер. 2: Языкознание. 2023. Т. 22. № 1. С. 13–27. [Ionova, Svetlana V. 2023. Emotional dominant of the text: Some linguistic aspects of research. *Science Journal of Volgograd State University. Linguistics* 22 (1). 13–27. (In Russ.)].
- Котельников Е.В., Разова Е.В., Котельникова А.В., Вычегжанин С.В. Современные словари оценочной лексики для анализа мнений на русском и английском языках (аналитический обзор) // *Научно-техническая информация. Серия 2: Информационные процессы и системы*. 2020. № 12. С. 16–33. [Kotelnikov, Evgeny V., Elena V. Razova, Anastasia V. Kotelnikova, Sergey V. Vychezhzhanin. 2020. Modern sentiment lexicons for opinion mining in English and Russian (analytical survey). *Nauchno-tehnicheskaya informaciya. Seria 2. Informacionnye processy i sistem* 12. 16–33. (In Russ.)].
- Ларина Т.В. Прагматика эмоций в межкультурном контексте // *Russian Journal of Linguistics*. 2015. № 1. С. 144–163. [Larina, Tatiana V. 2015. Pragmatics of emotions in intercultural context. *Russian Journal of Linguistics* 1. 144–163. (In Russ.)].
- Ларина Т.В. Эмотивная экологичность и эмотивная вежливость в жанре английской и русской анонимной рецензии // *Вопросы психолингвистики*. 2019. Т. 1. № 39. С. 38–57. [Larina, Tatiana V. 2015. Emotive ecology and emotive politeness in English and Russian: Blind peer-review. *Journal of Psycholinguistics* 1 (39). 38–57. (In Russ.)]. <https://doi.org/10.30982/2077-5911-2019-39-1-38-57>.
- Ларина Т.В. Эмотивная вежливость в лицеугрожающих речевых актах: кросс-культурный аспект // *Слово.ру: балтийский акцент*. 2025. Т. 16. №2. С. 118–135. [Larina, Tatiana V. Emotive politeness in face-threatening speech acts: Cross-cultural perspectives. *Slovo.ru: Baltic Accent* 16 (2). 118–135. (In Russ.)] <https://doi.org/10.5922/2225-5346-2025-2-7>.
- Озюменко В.И., Ларина Т.В. Искусственный интеллект в переводе: сильные и слабые стороны // *Вестник Волгоградского государственного университета*. 2025. Серия 2. Языкознание. 24 (1). С. 122–135. [Ozyumenko, Vladimir I. & Tatiana V. Larina. 2025. Artificial intelligence in translation: Advantages and limitations. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie / Science Journal of Volgograd State University. Linguistics* 24 (1). 122–135. (In Russ.)]. <https://doi.org/10.15688/jvolsu2.2025.1.10>
- Падучева Е.В. Динамические модели в семантике лексики. М.: Языки славянской культуры, 2004. [Paducheva, Elena V. 2004. Dynamic Models in Lexical Semantics. Moscow: Iazyki slavianskoi kul'tury Publ. (In Russ.)].
- Савчук С.О., Архангельский Т.А., Бонч-Осмоловская А.А., Донина О.В., Кузнецова Ю.Н., Ляшевская О.Н., Орехов Б.В., Подрядчикова М.В. Национальный корпус русского языка 2.0: новые возможности и перспективы развития // *Вопросы языкознания*. 2024. № 2. С. 7–34. [Savchuk, Svetlana O., Timofey Arkhangelskiy, Anastasiya A. Bonch-Osmolovskaya, Ol'ga V. Donina, Yuliya N. Kuznetsova, Ol'ga N. Lyashevskaya, Boris V. Orekhov & Mariya V. Podryadchikova. 2024. Russian National Corpus 2.0: New opportunities and development prospects. *Voprosy yazykoznanija* 2. 7–34. (In Russ.)].

- Сысоева Т.А., Люсин Д.В. Разработка расширенной базы данных с эмоциональными оценками существительных ENRuN-2: успехи, проблемы и перспективы // *Психология познания: материалы Всероссийской научной конференции. ЯрГУ, 6–8 декабря 2024 г.* / под ред. И.Ю. Владимирова, С.Ю. Коровкина. Ярославль: Филигрань, 2024. С. 316–320. [Sysoeva, Tatiana A. & Dmitrii V. Lyusin. 2024. Development of an extended database with emotional ratings of nouns ENRuN-2: Successes, problems and prospects. In Ilya Yu. Vladimirov, Sergey Yu. Korovkin (eds.), *Psychology of cognition: Proceedings of the All-Russian Scientific Conference, YARUSU, December 6-8*, 316–320. Yaroslavl : YARSU. (In Russ.)].
- Шаховский В.И. Лингвистическая теория эмоций. М.: Гнозис, 2008. [Shakhovsky, Viktor I. 2008. *Lingvisticheskaya teoriya emotsii* (Linguistic theory of emotions). Moscow: Gnosis Publ. (In Russ.)].
- Шаховский В.И. Язык и эмоции в аспекте лингвокультурологии. Волгоград: Перемена, 2009. [Shakhovsky, Viktor I. 2008. *Yazyk i emotsii v aspekte lingvokul'turologii* (Language and emotions in the aspect of linguoculturology). Volgograd: Peremena Publ. (In Russ.)].
- Alba-Juez, Laura & Michael Haugh (eds.). 2025. *The Sociopragmatics of Emotion*. Cambridge: Cambridge University Press.
- Alba-Juez, Laura & Tatiana Larina. 2018. Language and emotion: Discourse-pragmatic perspectives. *Russian Journal of Linguistics* 22 (1). 9–37.
- Buechel, Sven, Susanna Rücker & Udo Hahn. 2020. Learning and evaluating emotion lexicons for 91 languages. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 1202–1217. <http://doi.org/10.18653/v1/2020.acl-main.112>
- Cavicchio, Federica. 2025. *Emotion Detection in Natural Language Processing*. Cham: Springer.
- Cortal, Gustave, Alain Finkel, Patrick Paroubek & Lina Ye. 2023. Emotion recognition based on psychological components in guided narratives for emotion regulation. *Proceedings of the 7th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. 72–81. <http://doi.org/10.18653/v1/2023.latechclfl-1.8>.
- Dalal, Dhairya, Gaurav Negi & Davide Picca. 2025. LLMs and Emotional Intelligence: Evaluating emotional understanding through psychometric tools. *Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization (UMAP '25)*. 323–328. <https://doi.org/10.1145/3699682.3728315>.
- Demszky, Dorottya, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade & Sujith Ravi. 2020. GoEmotions: A dataset of fine-grained emotions. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 4040–4054. <http://doi.org/10.18653/v1/2020.acl-main.372>.
- Ekman, Paul. 1992. Are there basic emotions? *Psychological Review* 99 (3). 550–553.
- Gladkova, Anna. 2010. ‘Sympathy’, ‘compassion’, and ‘empathy’ in English and Russian: A linguistic and cultural analysis. *Culture & Psychology* 16 (2). 267–285.
- Gladkova, Anna. 2022. Emotions and attitudes in present day Russian through the prism of new words: Cultural semantics of zhest’ and related concepts. *Russian Journal of Linguistics* 26 (4). 970–994. <https://doi.org/10.22363/2687-0088-32167>.
- Izard, Carroll. E. 1991. *The Psychology of Emotions*. New York: Plenum.
- Larina, Tatiana & Douglas M. Ponton. 2022. I wanted to honour your journal, and you spat in my face: Emotive (im)politeness and face in the English and Russian blind peer review. *Journal of Politeness Research* 18 (1). 201–226. <https://doi.org/10.1515/pr-2019-0035>.
- Loukachevitch, Natalia & Anatolii Levchik. 2016. Creating a general Russian sentiment lexicon. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. 1171–1176.

- Kazyulina, Marina, Aleksandr Babii & Alexey Malafeev. 2021. Emotion classification in Russian: Feature engineering and analysis. *Analysis of Images, Social Networks and Texts, AIST 2020, Lecture Notes in Computer Science*, vol. 12602. 135–148.
- Kondratenko, Vladimir, Artem Sokolov, Nikolay Karpov, Oleg Kutuzov, Nikita Savushkin, & Fyodor Minkin. 2022. *Large Raw Emotional Dataset with Aggregation Mechanism*. <https://arxiv.org/abs/2212.12266> (accessed 25 May 2025).
- Mackenzie, John Lachlan & Laura Alba-Juez (eds.). 2019. *Emotion in Discourse* [Pragmatics and Beyond New Series 302]. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Mohammad, Saif M. 2023. Best practices in the creation and use of emotion lexicons. *Findings of the Association for Computational Linguistics: EACL 2023*. 1825–1836. <http://doi.org/10.18653/v1/2023.findings-eacl.136>
- Mohammad, Saif M. & Peter D. Turney. 2013. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence* 29 (3). 436–465. <http://doi.org/10.1111/j.1467-8640.2012.00460.x>
- Muhammad, Shamsuddeen Hassan, Nedjma Ousidhoum, Idris Abdulmumin et al. 2025. BRIGHTER: BRIdging the Gap in Human-annotated Textual Emotion Recognition. Datasets for 28 languages. <https://arxiv.org/abs/2502.11926>. (accessed 25 May 2025).
- Park Seo-Hui, Byung-Chull Bae & Yun-Gyung Cheong. 2020. Emotion recognition from text stories using an emotion embedding model. *IEEE International Conference on Big Data and Smart Computing. Busan, Korea (South), 2020*. 579–583. <http://doi.org/10.1109/BigComp48618.2020.00014>.
- Plaza-del-Arco, Flor Miriam, Alba A. Cercas Curry, Amanda Cercas Curry & Dirk Hovy. 2024. Emotion analysis in NLP: Trends, gaps and roadmap for future directions. *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*. 5696–5710.
- Plutchik, Robert. 1980. A general psycho-evolutionary theory of emotion. In Robert Plutchik & Henry Kellerman (eds.), *Theories of emotion*, 3–33. Cambridge: Academic Press.
- Sboev, Alexander, Aleksandr Naumov & Roman Rybka. 2021. Data-driven model for emotion detection in Russian texts. *Procedia Computer Science* 190. 637–642.
- Wang, Shuo, Aishan Maolinyazi, Xinle Wu & Xiaofeng Meng. 2020. Emo2Vec: Learning emotional embeddings via multi-emotion category. *ACM Transactions on Internet Technology* 20 (2). 1–17. <https://doi.org/10.1145/3372152>.
- Wierzbicka, Anna. 1992. Talking about emotions: Semantics, culture, and cognition. *Cognition and Emotion* 6 (3–4). 285–319.
- Wierzbicka, Anna. 1999. *Emotions across Languages and Cultures. Diversity and Universals*. Cambridge University Press.
- Wierzbicka, Anna. 2010. The semantics of emotions: Fear and it relatives in English. *Australian Journal of Linguistics* 2. 359–375.
- Wierzbicka, Anna. 2018. Emotions of Jesus. *Russian Journal of Linguistics* 22 (1). 38–53. doi.org/10.22363/2312-9182-2018-22-1-38-53.
- Zappettini, Franco, Douglas M. Ponton & Tatiana V. Larina. 2021. Emotionalisation of contemporary media discourse: A research agenda. *Russian Journal of Linguistics* 25 (3). 586–610. <https://doi.org/10.22363/2687-0088-2021-25-3-586-610>.

Словари / Dictionaries

- Бабенко Л.Г. Алфавит эмоций: словарь-тезаурус эмотивной лексики. СПб.: Кабинетный ученый, 2022. [Babenko Ludmila G. 2022. Alphabet of Emotions: Thesaurus of Emotive Lexis. Yekaterinburg; Moscow: Armchair Scholar Publ. (In Russ.)].

Русский семантический словарь: толковый словарь, систематизированный по классам слов и значений / Рос. акад. наук, Ин-т рус. яз. им. В.В. Виноградова; под общ. ред. Н.Ю. Шведовой. М.: Азбуковник, 1998. [Russkij semanticheskij slovar. Tolkovyj slovar, sistematizirovannyj po klassam slov i znachenij (The Russian Semantic Dictionary. Explanatory dictionary, systematized by classes of words and meanings). Shvedova, Natalia Yu. (eds.). Azbukovnik, Moscow. 1998. (In Russ.)].

Article history:

Received: 02 June 2025

Accepted: 25 July 2025

Bionotes:

Polina V. IAROSHENKO, PhD, is a Research Fellow in the Laboratory for Information Resources Analysis at the Research Computing Center, Lomonosov Moscow State University; Junior Research Fellow at Ivannikov Institute for System Programming of the Russian Academy of Sciences. Research interests: computational semantics, natural language processing, corpus linguistics.

e-mail: polina.iaroshenko@yandex.ru

<https://orcid.org/0000-0002-9425-5842>

Natalia V. LOUKACHEVITCH PhD (Advanced Doctorate), is a Leading Research Fellow in the Laboratory for Information Resources Analysis at Research Computing Center, Lomonosov Moscow State University; Principal Research Fellow at Ivannikov Institute for System Programming of the Russian Academy of Sciences. Research interests: computational semantics, natural language processing, artificial intelligence.

e-mail: louk_nat@mail.ru

<https://orcid.org/0000-0002-1883-4121>

Сведения об авторах:

Полина Владимировна ЯРОШЕНКО – кандидат филологических наук, научный сотрудник лаборатории анализа информационных ресурсов Научно-исследовательского вычислительного центра Московского государственного университета имени М.В. Ломоносова; младший научный сотрудник Института системного программирования им. В.П. Иванникова Российской академии наук. Сфера научных интересов: компьютерная семантика, обработка естественного языка, корпусная лингвистика.

e-mail: polina.iaroshenko@yandex.ru

<https://orcid.org/0000-0002-9425-5842>

Наталья Валентиновна ЛУКАШЕВИЧ – доктор технических наук, ведущий научный сотрудник лаборатории анализа информационных ресурсов Научно-исследовательского вычислительного центра Московского государственного университета имени М.В. Ломоносова; главный научный сотрудник Института системного программирования им. В.П. Иванникова Российской академии наук. Сфера научных интересов: компьютерная семантика, обработка естественного языка, искусственный интеллект.

e-mail: louk_nat@mail.ru

<https://orcid.org/0000-0002-1883-4121>