

# Morphology and Emotion Recognition: How Verb Form Affects Language Model Predictions

## Abstract

Emotion lexicons are an important tool for textual emotion recognition; however, they represent words as lemmas, disregarding morphological variation. This study tests the hypothesis that grammatical form influences emotion classification. For 473 verbs from the Russian Emotion Lexicon, indicative and imperative forms were generated (5,676 verb forms in total). Three encoders (ruBERT-tiny2, ruBERT-base, ruRoBERTa-large), fine-tuned on a corpus compiled from available Russian-language emotion analysis datasets, classified each form. The results demonstrate that past tense feminine forms systematically achieve optimal performance, surpassing the infinitive in F1-score, whilst imperative forms exhibit the poorest performance. Error analysis revealed a consistent pattern across all models: imperative and second-person present tense forms labeled as Sadness were systematically misclassified as Anger. Preliminary experiments with large language models showed similar tendencies with lower morphological sensitivity.

**Keywords:** emotion recognition; BERT; natural language processing; Russian language

## Морфология в распознавании эмоций: как глагольная форма влияет на предсказания языковой модели

### Аннотация

Эмоциональные лексиконы являются важным инструментом распознавания эмоций в тексте, однако они, как правило, представляют лексические единицы в виде лемм, не учитывая морфологическую вариативность. В работе представлено экспериментальное исследование о влиянии грамматической формы на определение эмоционального класса лексической единицы. Для 473 глаголов из Russian Emotion Lexicon были сгенерированы формы изъявительного и повелительного наклонения (всего 5676 глагольных форм). Три энкодера (ruBERT-tiny2, ruBERT-base, ruRoBERTa-large), дообученные на корпусе, составленном на основе доступных русскоязычных датасетов для анализа эмоций, классифицировали каждую форму. Выявлено, что формы прошедшего времени женского рода систематически показывают наилучшие результаты, превосходя инфинитив по F-мере, а формы повелительного наклонения, напротив, имеют наименьшие показатели. Анализ ошибок выявил устойчивую тенденцию, характерную для всех моделей: формы повелительного наклонения и второго лица настоящего времени с истинной меткой «Грусть» систематически классифицировались как «Злость». Предварительные эксперименты с большими языковыми моделями показали схожие тенденции при меньшей чувствительности к морфологии.

**Ключевые слова:** распознавание эмоций; BERT; обработка естественного языка; русский язык

## 1. Introduction

Textual emotion recognition is one of the key tasks in natural language processing (NLP). Emotion lexicons, consisting of lists of words and expressions associated with different emotional categories [1: 34], remain an important tool in this field [2]. However, existing lexicons, like traditional dictionaries, typically disregard morphological variation: words are represented as lemmas. Meanwhile, grammatical form can influence the emotional component of a lexical unit. For example, the verb *мучиться* ('to suffer') is categorised as Sadness in a number of sources [3; 4], but the imperative (*мучься*) is perceived as more aggressive, expressing anger rather than sadness.

The problem of morphological variation is also relevant given the nature of textual data analysed in emotion recognition — often comments from social networks, characterised by brevity and fragmentariness. As noted in [5], short utterances frequently lack sufficient context for determining emotion. This potentially increases the significance of the morphological characteristics of individual lexical units. Accordingly, the following research questions are formulated.

RQ-1: Does the morphological form influence emotion class prediction by language models?

RQ-2: Are there error patterns associated with specific morphological forms?

To address these questions, an experimental study was conducted. Based on the Russian Emotion Lexicon (RusEmoLex) [6], a list of emotion-annotated verbs was compiled. Various morphological forms were generated for each verb. Three encoders (ruBERT-tiny2, ruBERT-base, ruRoBERTa-large), fine-tuned on a combined corpus from four Russian-language datasets (total corpus size: 22,852 annotated examples), classified each form. Performance was evaluated using the F1-score.

## 2. Related Work

The field of textual emotion recognition has undergone significant development with the advent of deep learning methods [7; 8].

Despite the considerable number of datasets available for this task, reviews [5; 9] identify the shortage of annotated data as one of the main challenges in this area.

This can be attributed to two factors. Firstly, the dominance of English-language resources: the majority of available datasets are in English (see, for example, the description of various resources in the review [5]). Consequently, the lack of data in languages other than English is particularly acute [10]. Secondly, there exists a considerable number of theoretical approaches to describing emotions, and dataset annotation approaches differ depending on the underlying theoretical framework [5]. According to the statistics presented in [11], the approaches of Paul Ekman [12] and Robert Plutchik [13] are most frequently used in contemporary NLP research on emotion analysis. The diversity of theoretical approaches also results in the problem of imbalanced emotion classes — some emotions are overrepresented, others underrepresented.

For Russian, alongside machine learning datasets (described in Section 4), there exist specialised lexicons and results from native speaker surveys.

The RuSentiLex lexicon [3] contains over 12,000 evaluative units, of which 1,760 belong to emotional vocabulary grouped under the general tag ‘feeling’. The NRC Emotion Lexicon [14] is an English-language crowdsourced resource automatically translated into 108 languages, including Russian. It comprises over 14,000 units annotated for eight basic emotions from Plutchik’s model [13]: anger, fear, anticipation, trust, sadness, joy, disgust, and surprise — on a binary 0/1 scale. Both lexicons include various parts of speech.

The ENRuN database (Emotional Norms for Russian Nouns) [15] contains emotional ratings for 1,800 nouns across five classes (joy, sadness, anger, fear, disgust) on a scale from 0 to 5.

The integrative resource the Russian Emotion Lexicon (RusEmoLex) [6] combines data from lexicographic sources (dictionaries and thesauri), the semantically annotated part of the Russian National Corpus [16], psychological surveys on word–emotion associations, and datasets used in emotion recognition tasks. It includes 1,024 lexical units with part-of-speech information and annotation for five emotions (joy, sadness, anger, surprise, fear).

However, all the aforementioned resources share a common limitation: lexical units are represented as lemmas, which does not allow for assessment of the influence of grammatical form on emotional content.

## 3. Data

The general idea of the experiment is to determine the extent to which encoder language model’s predictions depend on verb form. The list of verbs annotated for emotions, required for conducting the experiment, was taken from the RusEmoLex, where lexical units are provided with part-of-speech tagging.

RusEmoLex contains 473 verbs annotated across five emotion classes, with each verb assigned only one label: sadness (181 verbs), joy (107), anger (81), fear (78), and surprise (26). For each verb, forms were generated in the indicative mood across past, present, and future tenses in all possible persons and numbers (where applicable, masculine and feminine forms were also generated, for example, for the past tense singular), as well as second-person imperative forms in both singular and plural. The total number of verb forms amounts to 5,676 (12 forms per verb). The verb forms were input sequentially into the language model as separate tokens.

## 4. Models

### 4.1. Datasets

To fine-tune the encoders, a combined corpus was constructed from available emotion-annotated datasets in Russian. A brief description of the resources used is provided below, with the main characteristics of each dataset presented in Table 1.

- CEDR [17] is a dataset in which each sentence is assigned one or more emotion class labels; sentences without class labels are also present, i.e. those deemed by annotators not to contain an emotional component. The text sources comprised comments from social networks, blogs, and news resources<sup>1</sup>.
- AIST [18] — the publicly available part of this dataset was used for fine-tuning the models. The dataset sources comprised comments from social networks (VK, Telegram)<sup>2</sup>.
- BRIGHTER [19] is a collection of human-annotated emotion recognition datasets for 28 languages, used in the SemEval-2025 competition on multilingual emotion recognition. The Russian-language portion of the dataset consists of comments from social networks<sup>3</sup>.
- GoEmotions [20] is a dataset consisting of emotion-annotated comments from Reddit. For fine-tuning the encoders within the experiment, an automatically translated Russian version of the dataset was used<sup>4</sup>.

| Dataset      | Labels                                       | Number of labelled texts      |
|--------------|----------------------------------------------|-------------------------------|
| CEDR         | joy, sadness, surprise, fear, anger          | 9,410                         |
| AIST         | joy, sadness, anger, uncertainty, neutrality | 50,750                        |
| BRIGHTER     | anger, disgust, fear, joy, sadness, surprise | 3,443 (train set for Russian) |
| ruGoEmotions | 27 emotions + neutrality                     | 58,000                        |

Table 1: Datasets for Emotion Classification in Russian

Examples labelled as sadness, joy, anger, fear, and surprise were selected from each dataset in accordance with the categories represented in RusEmoLex. Table 2 presents the distribution of examples by emotion in the combined dataset, with multi-class instances taken into account.

---

<sup>1</sup> [https://huggingface.co/datasets/sagteam/cedr\\_v1](https://huggingface.co/datasets/sagteam/cedr_v1)

<sup>2</sup> <https://github.com/asbabi/AIST>

<sup>3</sup> <https://brighter-dataset.github.io/>

<sup>4</sup> <https://github.com/searayeah/ru-goemotions>

| Emotions            | Number of labelled texts |
|---------------------|--------------------------|
| Joy                 | 9,884                    |
| Sadness             | 4,993                    |
| Anger               | 3,583                    |
| Surprise            | 2,257                    |
| Fear                | 1,703                    |
| Surprise, Joy       | 107                      |
| Anger, Sadness      | 79                       |
| Fear, Sadness       | 68                       |
| Anger, Surprise     | 48                       |
| Sadness, Joy        | 37                       |
| Surprise, Sadness   | 31                       |
| Fear, Joy           | 19                       |
| Fear, Surprise      | 19                       |
| Anger, Fear         | 13                       |
| Anger, Joy          | 10                       |
| Anger, Sadness, Joy | 1                        |
| <b>Total</b>        | <b>22,852</b>            |

Table 2: The Combined Dataset Used for Fine-Tuning Encoders

#### 4.2. Fine-Tuning the Encoders

Three models were used for the experiments: ruBERT-tiny2<sup>5</sup>, ruBERT-base<sup>6</sup> and ruRoBERTa-large<sup>7</sup>. All layers of the models were fine-tuned for the emotion classification task. The dataset was split 80/20 into training and validation sets. Texts were tokenized with a maximum length of 100 tokens (with truncation and padding applied). The following configuration was employed: AdamW optimizer, batch size of 64, linear scheduler with updates after each batch, and BCEWithLogitsLoss loss function. Learning rates were set at  $1 \times 10^{-5}$  (ruBERT-tiny2) and  $5 \times 10^{-6}$  (ruBERT-base, ruRoBERTa-large). Training was conducted until optimal validation performance was achieved: 19 epochs (ruBERT-tiny2), 7 epochs (ruBERT-base), and 3 epochs (ruRoBERTa-large). On the validation set, the models achieved the following F1-scores: ruBERT-tiny2 — 0.78, ruBERT-base — 0.79, ruRoBERTa-large — 0.9.

Following the fine-tuning of the encoders, metrics were obtained for the classification of verbs in infinitive form from RusEmoLex (Table 3). The infinitive form is treated as the baseline in this study.

<sup>5</sup> <https://huggingface.co/cointegrated/rubert-tiny2>

<sup>6</sup> <https://huggingface.co/DeepPavlov/rubert-base-cased>

<sup>7</sup> <https://huggingface.co/ai-forever/ruRoberta-large>

| Model           | Precision | Recall | F1   |
|-----------------|-----------|--------|------|
| ruRoBERTa-large | 0.77      | 0.76   | 0.77 |
| rubert-base     | 0.53      | 0.51   | 0.52 |
| ruBERT-tiny2    | 0.56      | 0.54   | 0.55 |

Table 3: Baseline for Emotion Classification of Verbs from RusEmoLex

The results obtained demonstrate that determining the emotion class of isolated verbs poses a challenge for encoder models trained on contextual representations. The ruRoBERTa-large model, with  $F1 = 0.77$ , shows acceptable performance, which may indicate the ability of large encoders to capture the emotional component at the level of individual lexical units. The significant drop in performance for ruBERT-base ( $F1 = 0.52$ ) and ruBERT-tiny2 ( $F1 = 0.55$ ) underscores the critical dependence of smaller models on contextual information for determining emotion class.

## 5. Results

Table 4 shows the highest and lowest emotion classification results for each model and the associated verb forms.

| Model           | Baseline<br>infinitive | F1<br>best        | F1<br>worst            |
|-----------------|------------------------|-------------------|------------------------|
| ruRoBERTa-large | 0.76                   | 0.83<br>past_femn | 0.64<br>impr_sing      |
| ruBERT-base     | 0.52                   | 0.61<br>past_femn | 0.49<br>impr_sing      |
| ruBERT-tiny2    | 0.55                   | 0.64<br>past_femn | 0.48<br>pres_sing_2per |

Table 4: Impact of Verb Form on Emotion Classification Performance by Encoders

The results demonstrate a systematic influence of verb form on the determination of the emotion class of lexical units. For all three models, the past tense feminine form yields the highest F1-measure values (0.83, 0.61, and 0.64, respectively), surpassing the baseline infinitive by 0.07–0.09 points. The relative gain amounts to 9% for ruRoBERTa-large, 17% for ruBERT-base, and 16% for ruBERT-tiny2. The lowest F1-measure values are produced by the singular imperative form (ruRoBERTa-large, ruBERT-base) and the present tense second-person form (ruBERT-tiny2).

Thus, substantial fluctuations in metrics are observed depending on verb form. Morphological sensitivity, calculated as the relative performance decrease between the best and worst forms, amounts to 23%, 20%, and 25% for ruRoBERTa-large, ruBERT-base, and ruBERT-tiny2, respectively.

Using the model with the highest F1-score as an example, we examine the main error patterns. Figure 1 presents a comparison of confusion matrices for ruRoBERTa-large for the best (past tense feminine form) and worst (singular imperative form) forms. Since the classes are imbalanced (for example, verbs labelled ‘Surprise’ are significantly fewer than verbs labelled ‘Sadness’), the matrices present percentages rather than raw counts for better readability: each value indicates what percentage of verbs of a given true class was classified as the corresponding predicted class.

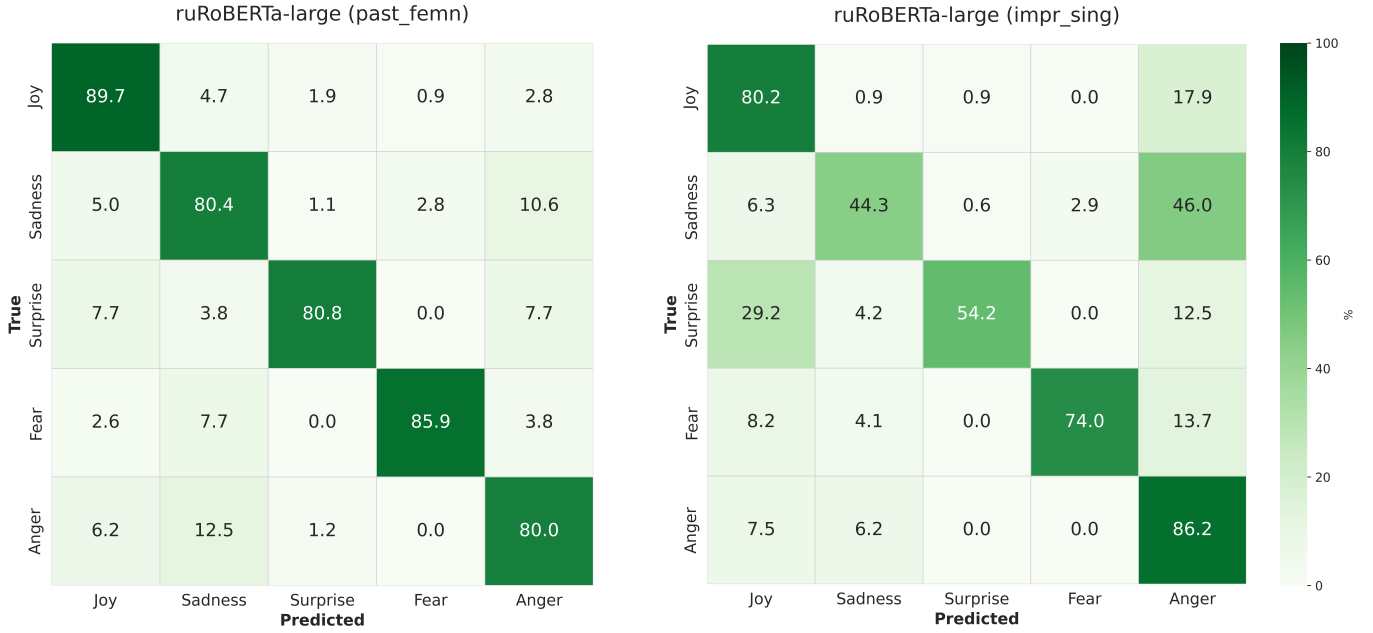


Figure 1: Confusion matrices of the ruRoBERTa-large model for the best-performing (past\_femn, left) and poorest-performing (impr\_sing, right) verb forms

Whilst the model demonstrates relatively high prediction accuracy for the past tense feminine form, accuracy drops sharply for the imperative form across two classes: Sadness and Surprise. The Sadness-Anger confusion is most evident: 46% of Sadness cases are misclassified as Anger. The Surprise class exhibits a bias towards Joy (29.2%), though to a lesser extent. Table 5 presents the most frequent errors for each model in terms of emotion classes and verb form.

| Model           | Form           | True    | Predicted | Count |
|-----------------|----------------|---------|-----------|-------|
| ruRoBERTa-large | impr_sing      | Sadness | Anger     | 80    |
|                 | impr_plur      | Sadness | Anger     | 76    |
|                 | pres_plur_2per | Sadness | Anger     | 61    |
| ruBERT-base     | pres_sing_2per | Sadness | Anger     | 81    |
|                 | impr_plur      | Sadness | Anger     | 79    |
|                 | impr_sing      | Sadness | Anger     | 77    |
| ruBERT-tiny2    | pres_sing_2per | Sadness | Anger     | 113   |
|                 | impr_plur      | Sadness | Anger     | 102   |
|                 | infinitive     | Sadness | Anger     | 100   |

Table 5: Top-3 Most Frequent Classification Errors Across Morphological Forms

For all models, the most frequent pattern is the Sadness–Anger error. This tendency is particularly characteristic of imperative forms, as well as second-person present tense. If we examine the lexical units of the Sadness class that are systematically classified by all three models as Anger, the following observations can be made: typically, these are verbs with high emotional intensity (*убиваться* — ‘to

grieve deeply’, *затравить* — ‘to persecute’, *изнурять* — ‘to exhaust’). A considerable number are verbs with semantics of suppression or violence (*сломить* — ‘to break’, *сокрушить* — ‘to crush’, *раздавить* — ‘to squash’, *подавить* — ‘to suppress’, *давить* — ‘to press’).

Thus, the models’ most frequent errors appear to be influenced by two factors: verb forms implying orientation towards another person (imperative, second-person present tense) and the verb’s semantics.

## 6. Large Language Models: Preliminary Experiments

In recent years, large language models (LLMs) have been increasingly employed in various NLP tasks, and the field of emotion analysis has been no exception (see, for example, [21; 22]). Given the promising potential of LLMs as a tool for emotion recognition in text, preliminary experiments were conducted with verbs from RusEmoLex (Table 6).

During the experiment, verb forms were input sequentially into the model with a plain prompt containing no specific role instructions:

Your task is to identify the emotion associated with the given word. Emotions: Joy, Fear, Surprise, Sadness, Anger. Your answer should consist of a single word indicating the relevant emotion class. No additional comments are required.

| Model                         | Baseline<br>(infinitive) | F1<br>best        | F1<br>worst       |
|-------------------------------|--------------------------|-------------------|-------------------|
| DeepSeek-V3                   | 0.79                     | 0.85<br>past_femn | 0.72<br>impr_sing |
| Qwen3-235B-A22B-Instruct-2507 | 0.85                     | 0.89<br>past_femn | 0.79<br>impr_plur |
| llama-3                       | 0.78                     | 0.83<br>past_femn | 0.69<br>impr_plur |
| YandexGPT 5 Lite              | 0.73                     | 0.77<br>past_femn | 0.66<br>impr_plur |

Table 6: Impact of Verb Form on Emotion Classification Performance by LLMs

The LLM classification results demonstrate systematic superiority over encoders both in absolute performance and in robustness with regard to the influence of morphological form. The average F1-measure for LLMs on the baseline form (infinitive) is 0.79, which exceeds the encoder performance (0.61).

The patterns in best and worst forms for LLMs are similar to those observed for encoders. All models demonstrate complete consistency, with the past tense feminine form proving optimal for emotion classification. The least effective forms are the plural or singular imperative (the sole exception being rubert-tiny2, for which the worst form is the second person present tense).

Morphological sensitivity, measured as the relative performance decrease between the best and worst forms, averages approximately 14% for LLMs and around 23% for encoders, indicating greater robustness of generative models to morphological variation. The relative performance gain when moving from the infinitive to the optimal form averages 6% for LLMs (substantially lower than for encoders). Further experiments with LLMs are planned.

## 7. Conclusion

This study examined whether morphological form influences the emotion class assigned to a word. Russian verbs from RusEmoLex served as experimental material, with multiple morphological forms generated for each verb.

The results demonstrate morphological influence across all tested models. The past tense feminine form consistently achieved optimal performance, whilst imperative forms produced the poorest results. Error pattern analysis revealed the most pronounced trend: systematic ‘Sadness–Anger’ confusion in imperative and second-person present tense forms. The experiments showed that morphological sensitivity is more pronounced in encoder models than in LLMs, indicating greater robustness of generative models. Importantly, the patterns identifying optimal and suboptimal forms remained consistent across both model types.

These findings suggest that emotion lexicons could benefit from incorporating morphological variants rather than lemmas alone. This approach may prove particularly valuable for analysing texts with limited context, such as social media comments or dialogue utterances.

## References

1. Cavicchio Federica. Emotion Detection in Natural Language Processing. — Cham : Springer, 2025.
2. Mohammad Saif. 2023. Best Practices in the Creation and Use of Emotion Lexicons // Findings of the Association for Computational Linguistics: EACL 2023. — Dubrovnik, Croatia : Association for Computational Linguistics, 2023. — P. 1825–1836.
3. Loukachevitch Natalia, Levchik Anatolii. 2016. Creating a General Russian Sentiment Lexicon // Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). — Portorož, Slovenia : European Language Resources Association (ELRA), 2016. — P. 1171–1176.
4. Babenko Lyudmila G. Alphabet of Emotions: Thesaurus of Emotive Lexis. — Yekaterinburg; Moscow : Armchair Scholar, 2022.
5. Deng Jiachen, Ren Fuji. 2023. A Survey of Textual Emotion Recognition and Its Challenges // IEEE Transactions on Affective Computing. — 2023. — Vol. 14, No. 1. — P. 49–67. — <https://doi.org/10.1109/TAFFC.2021.3053275>
6. Iaroshenko Polina V., Loukachevitch Natalia V. 2025. RusEmoLex: Russian Emotion Lexicon // Russian Journal of Linguistics. — 2025. — Vol. 29, No. 3. — P. 659–687. — <https://doi.org/10.22363/2687-0088-44439>.
7. Acheampong Francisca Adoma, Chen Wenyu, Nunziato Alex, Sorensen John. 2021. AI Based Emotion Detection for Textual Big Data: Techniques and Contribution // Big Data and Cognitive Computing. — 2021. — Vol. 5, No. 43. — <https://doi.org/10.3390/bdcc5030043>.
8. Kumar R. Praveen, Ramya Pannala, Teja G. Sai, Krishna V. Rama. 2025. BERTing Emotions: Exploring Textual Emotion Recognition using NLP // 2025 International Conference on Intelligent Computing and Control Systems (ICICCS). — Erode, India, 2025. — P. 888–895. — <https://doi.org/10.1109/ICICCS65191.2025.10984670>.
9. Kusal Sudhanshu, Patil Snehal, Kotecha Ketan, Aluvalu Rajasekhar, Varadarajan Vijayakumar. 2021. AI Based Emotion Detection for Textual Big Data: Techniques and Contribution // Big Data and Cognitive Computing. — 2021. — Vol. 5, No. 3. — Article 43. — <https://doi.org/10.3390/bdcc5030043>
10. De Bruyne Luna. 2023. The paradox of multilingual emotion detection // Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis. — Toronto, Canada : Association for Computational Linguistics, 2023. — P. 458–466.
11. Plaza-del-Arco Flor Miriam, Cercas Curry Alba A., Cercas Curry Amanda, Hovy Dirk. 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. — 2024. — P. 5696–5710.
12. Ekman Paul. 1992. Are there basic emotions? // Psychological Review. — 1992. — Vol. 99, No. 3. — P. 550–553.
13. Plutchik Robert. 1980. A general psycho-evolutionary theory of emotion // Theories of emotion / Robert Plutchik, Henry Kellerman (eds.). — Cambridge : Academic Press, 1980. — P. 3–33.
14. Mohammad Saif, Turney Peter. 2013. Crowdsourcing a word-emotion association lexicon // Computational Intelligence. — 2013. — Vol. 29, No. 3. — P. 436–465. — <http://doi.org/10.1111/j.1467-8640.2012.00460.x>
15. Sysoeva Tatiana A., Lyusin Dmitrii V. 2024. Development of an extended database with emotional ratings of nouns ENRuN-2: Successes, problems and prospects // Psychology of cognition: Proceedings of the All-Russian Scientific Conference, YARSU, December 6-8, 2024 / I. Yu. Vladimirov, S. Yu. Korovkin (eds.). — Yaroslavl : YARSU, 2024. — P. 316–320.
16. Savchuk Svetlana O., Arkhangelskiy Timofey, Bonch-Osmolovskaya Anastasia A., Donina Olga V., Kuznetsova Yulia N., Lyashevskaya Olga N., Orekhov Boris V., Podryadchikova Maria V. 2024. Russian National Corpus 2.0: New opportunities and development prospects // Voprosy Jazykoznanija. — 2024. — No. 2. — P. 7–34.

17. Sboev Alexander, Naumov Aleksandr, Rybka Roman. 2021. Data-driven model for emotion detection in Russian texts // *Procedia Computer Science*. — 2021. — Vol. 190. — P. 637–642.
18. Kazyulina Marina, Babii Alexander, Malafeev Alexander. 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. — Cham : Springer, 2021. — [https://doi.org/10.1007/978-3-030-72610-2\\_10](https://doi.org/10.1007/978-3-030-72610-2_10)
19. Muhammad Shamsuddeen Hassan, Ousidhoum Nedjma, Abdulmumin Idris et al. 2025. BRIGHTER: BRIdging the Gap in Human-Annotated Textual Emotion Recognition Datasets for 28 Languages // *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. — Vienna, Austria : Association for Computational Linguistics, 2025. — P. 8895–8916. — <https://doi.org/10.18653/v1/2025.acl-long.436>
20. Demszky Dorottya, Movshovitz-Attias Dana, Ko Jeongwoo, Cowen Alan, Nemade Gaurav, Ravi Sujith. 2020. GoEmotions: A Dataset of Fine-Grained Emotions // *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. — Online : Association for Computational Linguistics, 2020. — P. 4040–4054.
21. Castro Emmanuel, Calvo Hiram, Kolesnikova Olga. 2025. Emotion and Intention Detection in a Large Language Model // *Mathematics*. — 2025. — Vol. 13, No. 23. — Article 3768. — <https://doi.org/10.3390/math13233768>
22. Mitrofanova Olga, Bakaev Maksim, Yuryevtseva Polina. 2025. What Do LLMs know about human emotions? The Russian case study // *27th International Conference «Speech and Computer», SPECOM 2025. Szeged, Hungary. October 13–15, 2025. Proceedings. Part I. Lecture Notes in Computer Science*. — Vol. 16187. — Cham : Springer Nature, 2025. — P. 129–144.