Rule-Based Sentiment Analysis of Macedonian

Sofija Kochovska[1], Branko Kavšek[1,2] and Jernej Vičič[1,**]

[1]University of Primorska, UP FAMNIT, Glagoljaška 8, 6000 Koper, Slovenia
[2]Jožef Stefan Institute, JSI, Jamova cesta 39, 1000 Ljubljana, Slovenia

Abstract

This paper addresses the problem of sentiment analysis in Macedonian, a low-resource language with limited annotated datasets, sentiment lexicons, and linguistic tools. We present the construction of a sentiment lexicon containing positive and negative words, which serves as the foundation for a rule-based sentiment classification system. The final lexicon consists of approximately 4,000 entries, making it a substantial linguistic resource for sentiment analysis in the absence of large-scale annotated data. To enhance the accuracy and contextual sensitivity of the system, we incorporate affirmative and non-affirmative (AnA) modifiers such as intensifiers, diminishers, and polarity shifters which were also prepared in the scope of the presented research. These elements allow for more nuanced handling of sentiment polarity, especially in the presence of negation or emphasis. Additionally, stopwords are taken into account to preserve the structural and semantic integrity of analyzed text. The system also accounts for specific linguistic characteristics of Macedonian, such as the absence of grammatical case distinctions, which influence how sentiment is expressed and interpreted. The rule-based methodology is particularly suitable for low-resource environments, as it reduces reliance on large-scale annotated data. The results demonstrate that linguistically informed, rule-based approaches can provide effective sentiment analysis in under-resourced language settings.

Keywords

Sentiment analysis, Macedonian language, Low-resource languages, Rule-based classification, Sentiment lexicon, Linguistic modifiers, Natural language processing

1. Introduction

Macedonian language [1] is an Eastern South Slavic language. It is spoken as a first language by around 1.6 million people and serves as the official language of North Macedonia. Macedonian has a high degree of mutual intelligibility with Bulgarian and varieties of Serbo-Croatian. Section 4 describes the datasets and lexicons used in this study for sentiment analysis of Macedonian text, including both manually annotated and automatically generated resources. We detail the compilation and merging process of these lexicons to create a comprehensive sentiment resource tailored for Macedonian social media data.

Sentiment analysis is a widely studied task in natural language processing (NLP), aiming to determine the emotional tone of textual content. While significant advances have been made for high-resource languages, low-resource languages such as Macedonian still lack fundamental tools, annotated datasets, and reliable sentiment resources.

In this work, we present a rule-based sentiment analysis system for Macedonian, developed in the absence of large-scale annotated data. At the core of the system is a sentiment lexicon, initially built manually and carefully curated to include approximately 4,000 sentiment-bearing words. This initial lexicon served as the foundation for early system development and evaluation.

To expand the coverage and explore the effect of larger lexical resources, we later merged our lexicon with the publicly available sentiment lexicons from the work by Jovanoski et al. [2]. These resources

ITAT'25: Information Technologies –– Applications and Theory, September 26––30, 2025, Telgárt, Slovakia

***Corresponding author.
†These authors contributed equally.
kochovskasofija@gmail.com (S. Kochovska); branko.kavsek@upr.si (B. Kavšek); jernej.vicic@upr.si (J. Vičič)
https://github.com/sof42 (S. Kochovska)
0000-0002-9092-1633 (B. Kavšek); 0000-0002-7876-5009 (J. Vičič)

were obtained from the authors' GitHub repository and combined with our existing entries, resulting in a merged lexicon of approximately 8,000 words.

We tested our rule-based sentiment classification algorithm on both lexicon versions: the smaller, high-precision 4,000-word lexicon, and the broader 8000-word merged lexicon. This allowed us to evaluate the trade-offs between lexical precision and coverage in the context of rule-based classification. For evaluation, we used the same manually annotated Twitter test dataset from the work of Jovanoski et al. [2], enabling a direct comparison of our system's performance with previous approaches applied to Macedonian sentiment analysis.

To increase the sensitivity of our method to context and emphasis, we incorporated a set of affirmative and non-affirmative (AnA) modifiers, including intensifiers, diminishers, and polarity shifters. Additionally, Macedonian stopwords are taken into account to maintain semantic clarity.

Our rule-based approach is especially suited for low-resource settings, offering transparency and interpretability without the need for large annotated datasets. By leveraging linguistic insights and available lexical resources, we demonstrate that rule-based sentiment analysis can be a practical and effective solution for under-resourced languages like Macedonian.

The paper is structured as follows: Section 2 presents State-of-the-art in the scientific field of sentiment analysis for Macedonian language. Section 3 presents the methodology used in the presented experiment. Next follows the description of the properties and availability of data. Section 6 presents the results with a brief discussion and the paper finishes with conclusions and future work.

2. Related Work

Sentiment analysis for the Macedonian language has received limited attention in the NLP community due to the scarcity of linguistic resources, annotated corpora, and standardized evaluation benchmarks. However, a few key studies have laid the groundwork for sentiment classification in this low-resource language.

One of the most influential contributions is the work by Jovanoski et al. [2], who developed and evaluated several sentiment lexicons for Macedonian using Twitter data. Their study compared three manually created lexicons and introduced a lexicon-based classification algorithm that served as a baseline for Macedonian sentiment analysis. We incorporated both the manually created and manually extended lexicons from this work into our own sentiment lexicon to assess the trade-off between lexical precision and vocabulary coverage. The authors also made their resources publicly available[1], which greatly facilitated reproducibility and further development in this domain.

In a closely related effort, Jahić and Vičič [3] developed an annotated sentiment lexicon for the Bosnian language. While the lexicon was created for Bosnian, the two languages share many similarities due to their common South Slavic origin. To address the lack of sentiment resources in Macedonian, we translated the Bosnian lexicon into Macedonian as a foundation for our initial 4,000-word lexicon. This translation was followed by manual validation and curation to ensure linguistic accuracy and contextual relevance. Their methodological framework, which includes rule-based sentiment classification, the use of linguistic modifiers, and attention to low-resource constraints served as a basis and inspiration for our own approach.

Although Bosnian benefits from slightly more NLP support, the success of Jahić and Vičič's lexicon-based system supports the feasibility of applying similar techniques to Macedonian. Our work extends these ideas by adapting them to the specific linguistic features of Macedonian and incorporating additional resources such as affirmative and non-affirmative (AnA) modifiers and stopword handling.

Recent work on Bulgarian sentiment analysis [4] demonstrates the potential of leveraging typologically similar languages, although their focus on fine-grained emotion datasets differs from our lexicon-based approach. While Bulgarian's grammatical similarities to Macedonian make it a promising candidate for cross-lingual transfer, we prioritized Bosnian due to: (1) the immediate availability of its manually annotated sentiment lexicon [3], and (2) shared South Slavic lexical similarities in core

[1] https://github.com/badc0re/sent-lex/tree/master/sentiment_analysis_macedonian

sentiment vocabulary despite morphological differences. Future work could explore hybrid approaches that combine Bulgarian resources with Macedonian-specific adaptations.

Together, these studies provide valuable foundations for lexicon-based sentiment analysis in under-resourced South Slavic languages, and our contribution demonstrates how these approaches can be extended, translated, and tailored for the Macedonian language.

3. Methodology

The development of our rule-based sentiment analysis system for Macedonian followed a multi-step approach centered on the creation and expansion of a sentiment lexicon. Given the low availability of sentiment resources for Macedonian, we began by adapting resources from a closely related South Slavic language.

3.1. Initial Lexicon Construction via Translation

Our initial lexicon was derived by translating an existing Bosnian sentiment lexicon into Macedonian. This lexicon, introduced by Jahić and Vičič [3], contains manually annotated positive and negative words intended for sentiment analysis in Bosnian. Given the high lexical and morphological similarity between Bosnian and Macedonian, especially in sentiment-bearing vocabulary, this translation provided a suitable foundation. The initial translation of the Bosnian sentiment lexicon into Macedonian was performed manually by the first author. To ensure linguistic and semantic adequacy in the Macedonian context, the translated lexicon was subsequently reviewed by a total of five native speakers, including members of the author's family, fellow colleagues, and academic peers. The resulting lexicon consisted of approximately 4,000 entries.

3.2. Lexicon Expansion through Merging

To improve coverage and capture a broader range of sentiment expressions, we merged the translated Bosnian lexicon with two publicly available Macedonian sentiment lexicons developed by Jovanoski et al. [2]. These resources include:

- A manually annotated lexicon of sentiment-bearing words, created for sentiment analysis of Macedonian Twitter data by the aforementioned authors. This includes both the original manual lexicon and an extended manual lexicon developed by Jovanoski et al. [2] to enrich and expand the set of sentiment words. Both were used together in the merging process to enhance the final lexicon.
- A second lexicon automatically generated by the aforementioned authors using the Pointwise Mutual Information (PMI) method, which quantifies the association between a word and a sentiment class (positive or negative) based on their co-occurrence in a large unlabelled corpus. Words with higher PMI scores relative to sentiment classes are included as indicative of that polarity.

The merging process involved deduplication, harmonization of polarity labels, and manual review to resolve potential conflicts and inconsistencies across sources. The final merged lexicon comprises approximately 8,000 unique entries labelled as positive or negative.

Table 1 presents a quantitative overview of the overlap and unique contributions across the various sentiment lexicons used in this study, highlighting the extent to which each source contributed to the final merged resource.

Notably, the manual lexicon, the manual extended lexicon, and the PMI lexicon are all resources from Jovanoski et al. [2]. These lexicons represent both manually annotated and automatically generated sentiment vocabularies developed specifically for Macedonian Twitter data. The table details the number of entries unique to each lexicon as well as the intersections between pairs of lexicons. For instance, the "Translated lexicon only" row shows words exclusive to the translated lexicon, while "Manual extended lexicon only" and "PMI lexicon only" indicate unique entries from those Jovanoski et al. [2] lexicons. The "Manual lexicon only" row has zero entries, indicating that all original manual lexicon words overlap with other resources. The intersections, such as "Merged ∩ Manual extended" and "Merged ∩ PMI," reveal how many words the final merged lexicon shares with these sources, reflecting their substantial contributions. Overall, this table illustrates how the merging process effectively combined multiple complementary lexicons to create a comprehensive sentiment resource.

Table 1
Overlap and uniqueness across sentiment lexicons

Comparison	Entries
Translated lexicon only	2,702
Merged lexicon only	3,310
Manual lexicon only	0
Manual extended lexicon only	1,688
PMI lexicon only	1,827
Translated ∩ Manual	65
Translated ∩ PMI	62
Manual ∩ PMI	381
Merged ∩ Manual extended	2,993
Merged ∩ PMI	2,260
Total in final merged lexicon	**8,257**

3.3. Use of Affirmative and Non-Affirmative Modifiers

In addition to sentiment words, the system incorporates a curated list of affirmative and non-affirmative (AnA) modifiers. These include intensifiers, diminishers, and polarity shifters. These elements were manually identified and assigned contextual polarity transformation rules, allowing the system to handle negation and emphasis more accurately.

3.4. Inclusion of Stopwords and Linguistic Features

To ensure robustness and avoid spurious sentiment cues, a Macedonian stopword list was applied during preprocessing. These include common function words that do not carry sentiment information. Additionally, the system was tailored to account for Macedonian-specific linguistic features, such as the lack of grammatical case inflections, which simplifies certain syntactic parsing tasks compared to other Slavic languages, but also increases the need for a bigger lexicon size to achieve comparable results as lemmatized word forms do not cover the same amount of text.

3.5. Rule-Based Classification

To perform sentiment classification on Macedonian text, we developed a rule-based sentiment analyzer leveraging lexicons and modifier rules. First, we translated the Bosnian sentiment lexicon into Macedonian and merged it with two manually annotated lexicons from Jovanoski et al. [2], as well as a lexicon generated via Pointwise Mutual Information (PMI) for research purposes. The combined lexicon contains sets of positive and negative words.

The classification pipeline begins with text preprocessing, which includes cleaning input sentences by removing URLs, usernames, and punctuation, followed by normalization of character repetitions (e.g., reducing repeated letters). We utilize the `classla` library [5] for tokenization and lemmatization tailored to the Macedonian language. Tokens identified as stopwords are filtered out.

Negation handling is applied by marking tokens that appear within the scope of polarity-shifting negation words with a special prefix to indicate a negated context.

Sentiment scoring is computed by iterating over the lemmatized tokens and summing polarity values assigned by the lexicon: $+\gamma$ for positive words, $-\gamma$ for negative words, and 0 otherwise. Modifier tokens preceding sentiment words adjust this score by intensifying (doubling) or diminishing (halving) the weight. Additionally, if the sentiment word appears within a negation context, its polarity is inverted. In our implementation, γ was set to 1.

The final sentiment score for a given text is obtained by summing the weighted polarity values of all *sentiment-bearing tokens*, namely, tokens that appear in either the positive or negative sentiment lexicons. This aggregate score is then normalized by the total number of sentiment-bearing tokens, ensuring that the resulting value reflects the average polarity strength regardless of the length of the input text. Based on the normalized score and a predefined threshold parameter ϵ, the text is classified as *positive*, *negative*, or *neutral*. The parameter ϵ defines a margin around zero within which a sentiment is considered neutral. To determine the most effective decision boundary, multiple values of ϵ were empirically evaluated on a manually annotated test set.

4. Data

Initially, the sentiment lexicon was obtained by translating the Bosnian lexicon into Macedonian. The sizes of the lexicons before merging are as follows:

- **Positive words:** 1204 entries
- **Negative words:** 2341 entries
- **Intensifiers:** 69 entries
- **Diminishers:** 37 entries
- **Polarity shifters:** 21 entries
- **Stopwords:** 311 entries

The data are available in Zenodo repository in two separate datasets:

- Sentiment polarity lexicon of Macedonian language, DOI: 10.5281/zenodo.15790587, https://doi.org/10.5281/zenodo.15790587,
- Macedonian stopwords list and Sentiment polarity lexicon of Macedonian language, DOI: 10.5281/zenodo.15794762, https://doi.org/10.5281/zenodo.15794762,

After merging with the manually annotated and PMI-generated lexicons by Jovanoski et al. [2], only the positive and negative word lists were expanded, while the other categories remained unchanged:

- **Positive words:** 1997 entries
- **Negative words:** 6260 entries

This merged lexicon forms the core resource for our sentiment analysis model, ensuring coverage and accuracy for the Macedonian language domain.

5. Experimental Setup

5.1. Dataset

For evaluation, we utilized the annotated Twitter dataset introduced by Jovanoski et al. [2], comprising 1,139 Macedonian tweets. Each tweet is labelled with one of three sentiment classes: positive (1), neutral (0), or negative (-1). These labels served as the ground truth in our experiments. The data was extracted into a test set file (`test_data.mo`), where each line contains a tweet followed by its sentiment label, separated by a comma.

5.2. Lexicons

As mentioned above, the lexicons used in our lexicon-based sentiment analysis were preprocessed and loaded as Python sets to enable efficient lookup during the sentiment scoring process.

5.3. Preprocessing

Each tweet was processed through the following pipeline:

1. **Noise Removal**: URLs, user mentions, and punctuation were removed.
2. **Normalization**: Text was lowercased and repeated characters normalized.
3. **Stopword Removal**: Tokens found in the stopword list were removed.
4. **Negation Marking**: Tokens appearing within a three-token window after a polarity shifter were marked to indicate negation scope.
5. **Lemmatization**: Tokens were lemmatized using the Classla NLP toolkit's Macedonian language model [5].

5.4. Sentiment Scoring

Sentiment scores for each tweet were calculated using a lexicon-based heuristic approach as follows:

- Each positive word contributes $+\gamma$, each negative word contributes $-\gamma$.
- **Baseline Setup:** Scores of words preceded by intensifiers are multiplied by 2, while those preceded by diminishers are multiplied by 0.5.
- If a word is marked as negated (within the negation context), its score is inverted (multiplied by -1).
- The total sentiment score for a tweet is normalized by dividing by the number of sentiment-bearing words.

The tweet was then classified based on the normalized score and a threshold ε:

- **Positive**: normalized score $> \varepsilon$
- **Negative**: normalized score $< -\varepsilon$
- **Neutral**: otherwise

To evaluate the classifier's sensitivity to sentiment intensity, we experimented with multiple threshold values, $\varepsilon \in \{0.00, 0.10, 0.20, 0.30, 0.40, 0.50\}$. The threshold ε defines a neutral zone around zero, namely, sentiment scores within the interval $[-\varepsilon, +\varepsilon]$ are classified as neutral, while scores above or below this range are labelled as positive or negative, respectively. In our implementation, the base polarity contribution for each sentiment-bearing word was set to $\gamma = 1$ or $\gamma = -1$, making the score normalization and thresholding consistent with unit-weight sentiment scoring.

5.5. Evaluation Metrics

The evaluation was performed on a labelled test dataset by Jovanoski et al. [2], with standard classification metrics such as precision, recall, and F1-score [6]. Performance was measured using:

- **Average F1-score** computed as the arithmetic mean of the F1-scores for the positive and negative classes, with the neutral class excluded. This metric treats both sentiment polarities equally by averaging their individual F1-scores. Using this average F1-score aligns with the evaluation approach of Jovanoski et al. [2], enabling a direct comparison of polarity classification performance.
- **Classification Report**: Including Precision, Recall, and F1-score per class (negative and positive).

Neutral tweets were excluded from F1 score calculations to focus evaluation on polarity classification accuracy. This approach was adopted to enable direct comparison with the work of Jovanoski et al. [2], who stated that *"the systems were evaluated in terms of an F-score that is the average of the F1-score for the positive, and the F1-score for the negative class"*.

6. Results and Discussion

In this study, we conducted sentiment analysis experiments using two different lexicon approaches: one based solely on translated lexicons, and another that combined the translated lexicons with additional manually curated or merged lexicons. The goal was to evaluate whether merging lexicons leads to measurable improvements in sentiment classification performance for Macedonian text [7].

For both approaches, we tested a range of threshold values (ϵ) to determine the optimal decision boundary between positive, negative, and neutral sentiment classes. Standard classification metrics such as precision, recall, and F1-score [6], are reported separately for positive and negative classes.

6.1. Translated Lexicon Performance

We begin by discussing the results obtained using the smaller lexicons (non-merged ones). Figure 1 illustrates how varying the threshold parameter ϵ influences the classification performance on the test dataset.

At $\epsilon = 0.00$, the classifier achieves an average F1-score of 0.591, performing notably better on the positive class (F1 = 0.681) than the negative class (F1 = 0.501). As ϵ increases up to 0.30, performance remains relatively stable, with minimal changes in F1-scores. However, beyond $\epsilon = 0.30$, performance degrades, primarily due to a sharp drop in recall for the negative class. At $\epsilon = 0.50$, the average F1-score falls to 0.549. These results suggest that optimal performance with smaller lexicons is achieved around $\epsilon = 0.00–0.30$, with consistently stronger detection of positive sentiment.

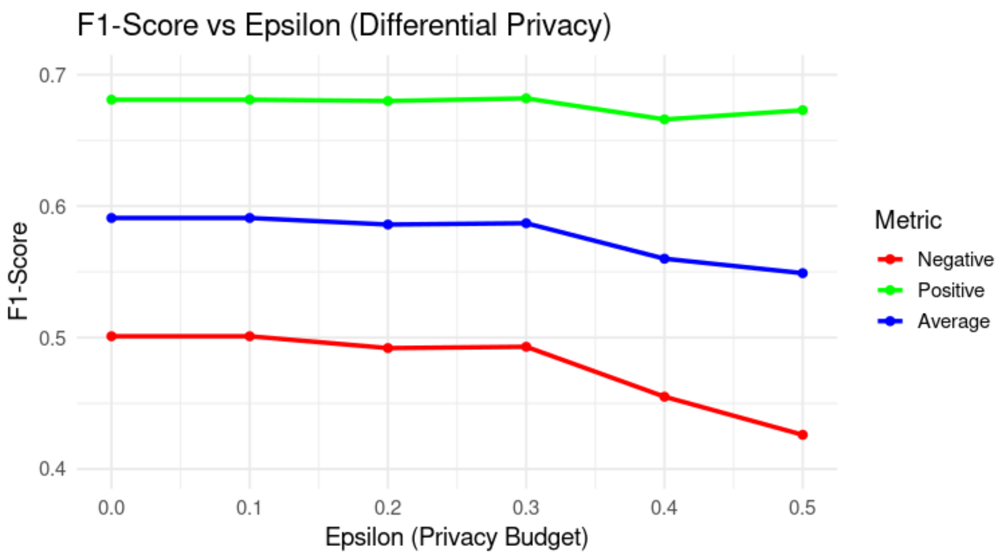

Figure 1: Average F1-score for positive and negative sentiment classes across different threshold values ϵ, using the smaller (non-merged) lexicons.

6.2. Merged Lexicons Performance

We now turn our attention to the results obtained using the merged lexicons, to assess whether combining translated and manually curated resources leads to improved sentiment classification performance across different threshold values.

With merged lexicons, the classifier shows improved and more balanced performance. At $\epsilon = 0.00$, the average F1-score is 0.713, with strong results for both classes ($F1_{neg} = 0.690$, $F1_{pos} = 0.736$). Performance remains stable up to $\epsilon = 0.30$, where the average F1-score peaks at 0.714. Beyond this, results begin to decline, with the negative class F1-score dropping to 0.617 at $\epsilon = 0.50$. Overall, merged lexicons enhance sentiment classification, especially in the low-to-mid ϵ range.

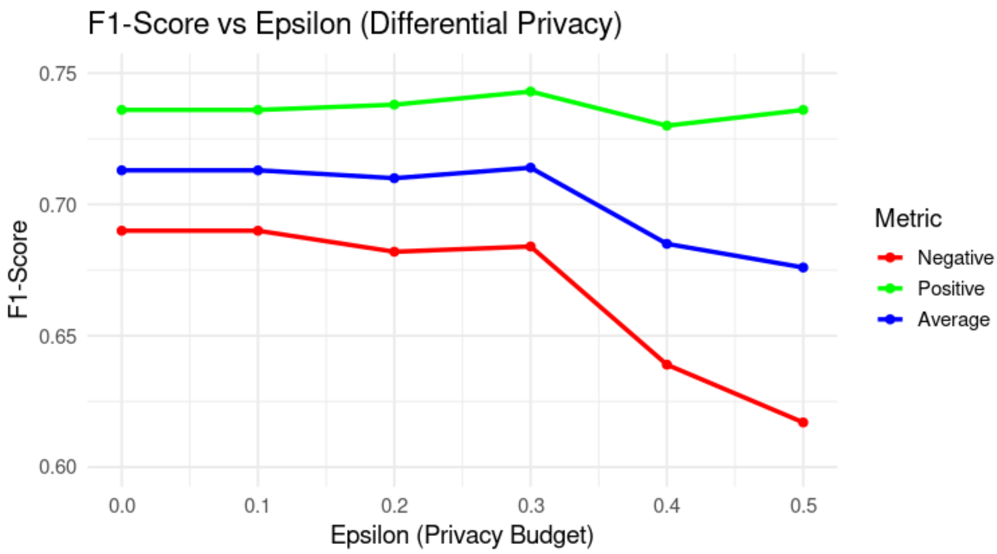

Figure 2: F1-Score vs ϵ using merged lexicons. Positive and average F1-scores remain high at lower ϵ values, while the negative class declines more noticeably with increasing noise.

6.3. Hyperparameter Optimization and Cross-Validation

6.3.1. Baseline Epsilon Sensitivity via 5-Fold Cross-Validation

Before conducting a full grid search, we evaluated the impact of the epsilon (ϵ) threshold on classification performance using 5-fold cross-validation with a fixed set of parameters using the merged lexicon:

- **Intensifier** = 2.0
- **Diminisher** = 0.5
- **Negation window** = 3

This setup was chosen to establish a baseline and isolate the influence of ϵ alone. The results, shown in table 2, represent the average F1-score across the five folds (averaged over the positive and negative classes):

Table 2
Epsilon sensitivity analysis using 5-fold cross-validation

Epsilon	Average F1-score
0.00	0.713
0.10	0.713
0.20	0.709
0.30	0.712
0.40	0.684
0.50	0.677

These results indicate that lower epsilon values (0.00–0.10) consistently yielded better performance, while higher thresholds beyond 0.30 resulted in a notable drop in F1-score. This analysis confirmed the model's sensitivity to the epsilon threshold and motivated the narrowed epsilon range (0.00–0.30) used in the full hyperparameter grid search described in the next subsection.

6.4. Grid Search for Hyperparameter Optimization

To address concerns regarding parameter tuning and overfitting, we performed an extensive grid search to optimize key hyperparameters of our rule-based sentiment classification model. This was done independently of the final test set, using 5-fold cross-validation on the 1,139 annotated Macedonian tweets from Jovanoski et al. [2].

6.4.1. Grid Search Configuration

We systematically explored 108 configurations by varying the following parameters:

- **Intensifier coefficient:** {1.5, 2.0, 2.5}
- **Diminisher coefficient:** {0.3, 0.5, 0.7}
- **Negation window size (in tokens):** {2, 3, 4}
- **Epsilon threshold** (ϵ): {0.00, 0.10, 0.20, 0.30}

This selection of epsilon values was informed by our earlier experiments in 6.3.1, where a fixed configuration (Intensifier = 2.0, Diminisher = 0.5, NegWin = 3) showed that both $\epsilon = 0.00$ and $\epsilon = 0.10$ yielded the highest average F1-score. Therefore, we focused the epsilon range more narrowly (0.00 to 0.30) in the grid search to identify optimal combinations under broader parameter variations.

Each combination was evaluated via 5-fold cross-validation on the manually labelled Macedonian tweet dataset. For each fold, we computed the F1-scores for the positive and negative classes and averaged them to produce a robust estimate of classification quality. The neutral class was excluded from F1, following the evaluation protocol in [2].

6.4.2. Top 5 Performing Configurations

Table 3
Top 5 configurations from the hyperparameter grid search

Avg F1	Intensifier	Diminisher	NegWin	Epsilon
0.736	1.5	0.7	2	0.30
0.736	1.5	0.5	2	0.30
0.735	1.5	0.3	2	0.30
0.732	1.5	0.7	2	0.00
0.732	1.5	0.5	2	0.00

The results, shown in table 3 demonstrate that:

- The optimal intensifier coefficient (1.5) was lower than the initial 2.0, suggesting that Macedonian sentiment expressions benefit from moderate emphasis. For example, phrases like 'very good' required less amplification than in English to avoid over-scoring.
- The 2-token negation window outperformed larger windows, aligning with Macedonian's tendency for negation particles to modify immediately following adjectives/verbs. Longer windows risked capturing irrelevant terms due to flexible syntax. Longer windows introduced noise by capturing unrelated words outside the negation's semantic scope.
- Similarly to intensifiers, diminisher values (0.5–0.7) provided balanced attenuation without suppressing sentiment cues too strongly
- An epsilon threshold of 0.30 slightly outperformed 0.00 in most cases, despite earlier results favoring 0.00 in isolation.

7. Conclusions and Future Work

Our experiments demonstrate that merging translated lexicons with manually curated lexicons improves sentiment classification performance for Macedonian tweets, achieving an average F1-score of 73.6% at the optimal threshold alongside optimal hyperparameter configurations. This combined lexicon approach yields more balanced results between positive and negative classes compared to using translated lexicons alone.

However, the overall F1-scores remain moderate, potentially due to the prevalence of slang and informal language in tweets, which likely reduces lexicon coverage and impacts classification accuracy.

Compared to the approach by Jovanoski et al. [2], our method achieves competitive results with a much simpler pipeline that relies primarily on lexicon merging and threshold tuning, without requiring complex machine learning models or extensive feature engineering.

The future work will incorporate creation of slang lexicons, context-aware models, or leveraging deeper linguistic features that will further enhance performance on Macedonian social media text. Future research can explore several avenues to further enhance the sentiment analysis framework:

- **Lexicon Expansion:** Increasing the coverage of sentiment lexicons by incorporating more domain-specific terms, slang, and evolving language use could improve detection accuracy. Crowdsourcing or semi-automated lexicon creation methods may aid this process.
- **Machine Learning Integration:** Combining the lexicon-based approach with supervised or unsupervised machine learning models could leverage both linguistic knowledge and data-driven patterns, potentially boosting performance.
- **Broader Dataset Testing:** Evaluating the approach on larger, more diverse datasets, including different genres or social media platforms, will test robustness and generalizability.
- **Multilingual and Cross-domain Adaptation:** Extending the methodology to other languages or adapting lexicons for related domains could broaden the applicability of the system.
- **Contextual and Phrase-level Sentiment:** Incorporating context-aware sentiment analysis, including handling of idiomatic expressions, negation scope, and multi-word phrases, will further refine polarity detection.
- **Parameter Generalization:** Investigating language-specific tuning of coefficients (e.g., training intensifier weights on annotated data) could further optimize rule-based systems for low-resource languages.

These directions offer promising opportunities to improve sentiment analysis for Macedonian and other low-resource languages.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] C. E. Kramer, L. Mitkovska, Macedonian: A course for beginning and intermediate students, University of Wisconsin Press, 2011.

[2] D. Jovanoski, V. Pachovski, P. Nakov, Sentiment analysis in Twitter for Macedonian, in: R. Mitkov, G. Angelova, K. Bontcheva (Eds.), Proceedings of the International Conference Recent Advances in Natural Language Processing, INCOMA Ltd. Shoumen, BULGARIA, Hissar, Bulgaria, 2015, pp. 249–257. URL: https://aclanthology.org/R15-1034/.

[3] S. Jahic, J. Vičič, Annotated lexicon for sentiment analysis in the Bosnian language, Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave 11 (2023) 59–83. doi:10.4312/slo2.0.2023.2.59-83.

[4] I. Temnikova, I. Marinova, S. Gargova, R. Margova, A. Komarov, T. Stefanova, V. Kireva, D. Vyatrova, N. Grigorova, Y. Mandevski, S. Minkov, SM-FEEL-BG - the first Bulgarian datasets and classifiers for detecting feelings, emotions, and sentiments of Bulgarian social media text, in: N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, N. Xue (Eds.), Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), ELRA and ICCL, Torino, Italia, 2024, pp. 14954–14966. URL: https://aclanthology.org/2024.lrec-main.1301/.

[5] N. Ljubešić, L. Terčon, K. Dobrovoljc, CLASSLA-Stanza: The Next Step for Linguistic Processing of South Slavic Languages, in: Conference on Language Technologies and Digital Humanities (JT-DH-2024), Institute of Contemporary History, Ljubljana, Slovenia, 2024. URL: https://doi.org/10.5281/zenodo.13936406. doi:10.5281/zenodo.13936406.
[6] H. Schütze, C. D. Manning, P. Raghavan, Introduction to information retrieval, volume 39, Cambridge University Press, Cambridge, 2008.
[7] B. Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Cambridge University Press, 2015.