Western European Journal of Linguistics and Education
Volume 2, Issue 11, November 2024
https://westerneuropeanstudies.com/index.php/2
ISSN (E): 2942-190X Open Access| Peer Reviewed

BEHAVIORAL LEXEMES AND THEIR IMPORTANCE IN SENTIMENTAL ANALYSIS

Rahimov Hasanboy Komiljonovich
2nd year basic doctoral student of ASU,
intern teacher of Namangan State
Institute of Foreign Languages
xasanraximov@adu.uz

Annotation
This article is dedicated to exploring the importance of sentimental analysis of behavioral lexemes in the Uzbek language. Behavioral lexemes, as words that express a person’ s ethical and emotional qualities, embody positive, negative, and neutral connotations. The article examines the emotional value of these lexemes and explores the thesaurus-based approaches to their classification. Furthermore, it highlights the linguistic relationships that underlie behavioral lexemes for sentimental analysis, emphasizing the practical significance of building synonym and antonym networks. Methods for processing behavioral lexemes to detect emotions in texts, the connotative meanings of words, and their socio-cultural impact are also analyzed. The article serves as a means to delve deeper into the unique semantic features of the Uzbek language and offers new perspectives on the language through sentimental analysis.
Keywords: Sentimental analysis, neutral lexemes, thesaurus, connotation, Euclidean distance

Introduction

In linguistics, lexemes that express behavior (character traits and personal attributes) hold significant importance. These lexemes describe a person’ s ethical and emotional qualities. Studying these lexemes in relation to sentimental analysis in the Uzbek language enables precise and effective evaluation of emotions within texts. This, in turn, has applications across various fields. For example, sentimental analysis can enhance business practices by evaluating feedback on social media platforms or product advertisements. In the legal domain, systems based on deep learning can detect domestic violence, while in security, extremist and non-extremist posts can be classified. Additionally, identifying emotion-related units and associating them with keywords allows artificial intelligence to improve machine translation, leading to significant advancements. This article discusses the works of researchers who have utilized sentimental analysis for various purposes and explores how behavioral lexemes can be effectively applied in the process of sentimental analysis.

Literature Review

Sentiment analysis has emerged as a vital field of study, particularly in its application to behavioral lexemes. These lexemes, forming the foundation of a thesaurus for sentiment analysis, play a significant role in simplifying the evaluation and classification of emotions. Positive, negative, and neutral lexemes, when systematically categorized, provide deeper insights into social, psychological, and ethical aspects within texts. Despite the significance, research on this subject in Uzbekistan remains scarce, highlighting the need to explore methodologies developed by international researchers. Over the past decade, various sentiment analysis models have been introduced globally, addressing the identification and understanding

Western European Journal of Linguistics and Education
Volume 2, Issue 11, November 2024
https://westerneuropeanstudies.com/index.php/2
ISSN (E): 2942-190X Open Access| Peer Reviewed

of emotions, especially in textual data on social media platforms. A lot of research work has been done in this regard.

For instance, Tao [2] proposed a keyword-based sentiment recognition approach, where sentences are constructed using Emotional Function Words (EFW). These EFWs are categorized into emotional keywords, modifiers, and metaphors. Drawing from Ekman' s six basic emotions, emotional keywords are assigned specific weights, enhancing their effectiveness in sentiment analysis.

In addition, Bruyne and colleagues [3] developed a sentiment classification system for English tweets, utilizing preprocessing techniques such as tokenization, stemming, lowercasing, and part-of-speech tagging. They extracted features through n-grams, lexical properties, and syntactic attributes. This approach employs multiple binary classifiers, each targeting specific emotions, thus improving the granularity of emotion detection.

Deep learning techniques have gained prominence in sentiment analysis, as highlighted by multiple researchers [4– 7]. These methods employ neural networks to learn complex patterns and relationships in data. After initial preprocessing, an embedding layer converts words into numerical representations, which are then processed through neural network layers for classification. Such methods are particularly effective in capturing nuanced emotional expressions within texts.

Rathnayaka and collaborators [8] proposed a multi-label sentiment detection system for microblog posts, using tools like Ekphrasis for normalization, tokenization, and segmentation. Their system classified emotions into 11 categories and offered mechanisms to filter negative sentiments on social media platforms.

The RAN model [9] focuses on sentence-level sentiment analysis, emphasizing interrelations among matrices and elements to perform an in-depth analysis. Similarly, the Matrix-Vector Recursive Neural Network (MVRNN) [10] is an innovative approach combining CNN and RNN for short-context sentiment analysis. Long Short-Term Memory (LSTM) models utilize sentiment lexicons as context, improving the analysis process's efficiency. CNN-LSTM models integrate localized CNN components and LSTM layers for enhanced sentiment prediction.

Souri and colleagues [11] analyzed Facebook user activities to develop a personality classification model, leveraging machine learning techniques to identify personal traits. Risch and Krestel [12] focused on social media content collection, employing deep learning for aggression identification and mitigation. Their studies underline the role of sentiment analysis in reducing harmful impacts of social media interactions.

Other researchers have explored unique applications of sentiment analysis. For instance, a deep learning-based system [13] was designed to identify domestic violence by classifying posts into categories such as empathy, awareness, and fundraising. Ahmad et al. [14] developed a model to classify Twitter content into extremist and non-extremist posts. Budiharto and Meiliana [15] applied sentiment analysis to predict election results in Indonesia, demonstrating its utility in political forecasting.

Al Shehhi and colleagues [16] measured happiness levels in the United Arab Emirates by analyzing tweets in English and Arabic. Similarly, other researchers [17– 18] examined student feedback on educators and football fans' emotions in response to game events, respectively. Ibrahim et al. [19] proposed a toxicity detection model based on CNN, GRU, and bi-directional LSTM, achieving high accuracy rates using Wikipedia datasets.

These approaches underscore the versatility of sentiment analysis, enabling its application in diverse contexts such as identifying toxicity, understanding social behaviors, and analyzing

Western European Journal of Linguistics and Education
Volume 2, Issue 11, November 2024
https://westerneuropeanstudies.com/index.php/2
ISSN (E): 2942-190X Open Access| Peer Reviewed

public opinion. By integrating these methodologies, sentiment analysis of behavioral lexemes in the Uzbek language can achieve greater accuracy and relevance in evaluating emotional expressions across textual data.

Methodology

This study utilizes a thesaurus-based approach to classify and analyze behavioral lexemes in the Uzbek language for sentimental analysis. The methodology involves several steps, which are outlined below:

1. **Linguistic Data Collection**

A comprehensive database of Uzbek texts, including literary works, social media posts, and spoken language samples, was compiled. These texts serve as the source for extracting behavioral lexemes.

2. **Lexeme Structuring**

Behavioral lexemes were categorized into three semantic layers:

Denotative semantics (naming sema): Refers to the primary meaning of lexemes, such as "arslon" (lion) or "quyosh" (sun).

Connotative semantics (expressive sema): Captures the emotional and evaluative meanings of lexemes, such as "adolatli" (just) or "sabrsiz" (impatient).

Functional semantics (task-oriented sema): Explores the contextual application of lexemes in communication.

3. **Thesaurus Construction**

The behavioral lexemes were mapped with their synonyms, antonyms, and related terms to construct a thesaurus. The thesaurus was designed to highlight semantic relationships such as:

Synonymy (e.g., "chaqqon" – "epchil"),

Antonymy (e.g., "sabrli" – "sabrsiz"),

Hypernymy and hyponymy (e.g., "jasur" – "qahramon").

4. **Sentiment Classification**

Lexemes were classified into **positive**, **negative**, and **neutral** sentiments based on their emotional implications:

Positive examples: "jasur" (brave), "mehribon" (kind).

Negative examples: "zolim" (tyrant), "dangasa" (lazy).

Neutral examples: "loqayd" (indifferent), "mahmadona" (boastful).

5. **Semantic Relationship Mapping**

Advanced tools such as **Euclidean distance formulas** were employed to identify the strength of relationships between lexemes based on their semantic proximity. The formula is given as:

E.D(x,y)=∑i=1d(xi−yi)2+∑j=1d(xj−yj)2E.D(x, y) = \sqrt{\sum_{i=1}^{d}(x_i - y_i)^2 + \sum_{j=1}^{d}(x_j - y_j)^2}E.D(x,y)=i=1∑d(xi−yi)2+j=1∑d(xj−yj)2

Results

The thesaurus of behavioral lexemes developed in this study facilitated the sentiment analysis process, enabling the precise identification of positive, negative, and neutral emotional meanings of words. The results of the study are outlined as follows:

1. **Classification of Behavioral Lexemes**
   a. **Positive lexemes**, such as *"jasur"* (brave) and *"mehribon"* (kind), effectively conveyed positive emotions like happiness and compassion.
   b. **Negative lexemes**, such as *"zolim"* (tyrant) and *"dangasa"* (lazy), clearly expressed dissatisfaction or negative sentiment.

ISSN (E): 2942-190X Open Access| Peer Reviewed

c. **Neutral lexemes**, such as *"loqayd"* (indifferent) or *"sinchkov"* (meticulous), demonstrated variability in emotional connotations depending on the context.

2. **Thematic Grouping**
Behavioral lexemes were categorized into thematic groups, for instance:
   a. Lexemes related to speech (*"gapdon"*—eloquent).
   b. Physical activity (*"chaqqon"*—agile).
   c. Appearance or clothing habits.
3. **Linguistic Relationships**
Semantic relationships, such as **synonyms** (*"chaqqon"*—*"epchil"*, agile), **antonyms** (*"do' stona"*—*"dushmanona"*, friendly—hostile), and **hierarchical structures** (*"jasur"*—*"qahramon"*, brave—hero), were effectively categorized. This approach enhanced understanding of contextual emotional impact.
4. **Utilization of Euclidean Distance**
The study employed Euclidean distance calculations to rank lexemes by their semantic proximity, demonstrating the method' s efficacy in analyzing emotions within large text corpora.

Discussion

This study emphasized the significance of integrating linguistic tools, particularly thesauruses, into sentiment analysis workflows. The findings provide substantial insights into the role of behavioral lexemes in sentiment detection:

1. **Role of Thesauruses in Sentiment Analysis**
The categorization of lexemes into thematic and semantic groups improved the precision of sentiment detection. It also laid the groundwork for artificial intelligence systems to interpret emotions in a structured manner.
2. **Cultural and Contextual Sensitivity**
Emotional connotations of words were shown to be heavily influenced by cultural and social contexts. For instance, the lexeme *"loqayd"* (indifferent) could represent either positive or negative sentiment depending on the situational context.
3. **Practical Applications Across Disciplines**
The thesaurus developed in this study has wide-ranging applications, including:
   - **Social Media Analysis**: Detecting extremist content in platforms like Twitter.
   - **Business and Marketing**: Analyzing customer feedback and product reviews.
   - **Artificial Intelligence and Machine Translation**: Identifying contextual emotions for better translation accuracy.
4. **Alignment with Global Research**
The results align with global sentiment analysis studies, such as J. Tao' s keyword-based sentiment detection methodology and Rathnayaka et al.' s multi-label classification models using neural networks. Incorporating CNN and LSTM models with the proposed thesaurus can provide even deeper insights into sentiment analysis.
5. **Challenges and Limitations**
   - **Neutral Lexemes**: Ambiguity in emotional connotations posed significant challenges.
   - **Large-Scale Analysis**: Technical difficulties in processing extensive text corpora highlighted the need for more advanced machine learning models.

Western European Journal of Linguistics and Education
Volume 2, Issue 11, November 2024
https://westerneuropeanstudies.com/index.php/2
ISSN (E): 2942-190X Open Access| Peer Reviewed

This work contributes significantly to enhancing sentiment analysis in the Uzbek language by focusing on behavioral lexemes. It paves the way for deeper exploration of linguistic features and demonstrates its practical applications in diverse fields, including social sciences, computational linguistics, and artificial intelligence.

Conclusion

By identifying and adapting behavioral lexemes in the Uzbek language for sentimental analysis, it becomes possible to determine the positive and negative meanings of words and express emotions based on them. This approach helps in thoroughly analyzing the internal structures of language and is useful for assessing information, drawing ethical conclusions, and understanding social and cultural relationships. Moreover, sentimental analysis can contribute positively to various fields, such as analyzing social media data on product reviews, facilitating trade and marketing, predicting diseases, improving agriculture, and enhancing psychological studies and legal frameworks.

References

1. Maksumova, Saida. " O' zbek tilida konnotativ ma' noning hosil bo' lishi." O' DJTU, Uzbek Language and Literature Department, Associate Professor, Ph.D. in Philology, Uzbekistan, Tashkent. Available at: https://uzresearchers.com/index.php/Conferences/article/view/3157/3094.
2. Boivin, M., Hymel, S., & Bukowski, W. M. (1995). The roles of social withdrawal, peer rejection, and victimization by peers in predicting loneliness and depressed mood in childhood. *Development and Psychopathology, 7*(04), 765-785.
3. Chou, K.L., Liang, K., & Sareen, J. (2011). The association between social isolation and DSM-IV mood, anxiety, and substance use disorders: Wave 2 of the National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Clinical Psychiatry, 72*(11), 1468-1476.
4. Davila, J., Hershenberg, R., Feinstein, B. A., Gorman, K., Bhatia, V., & Starr, L. R. (2012). Frequency and quality of social networking among young adults: Associations with depressive symptoms, rumination, and co-rumination. *Psychology of Popular Media Culture, 1*(2), 72-86.
5. Deters, F., & Mehl, M. R. (2012). Does posting Facebook status updates increase or decrease loneliness? An online social networking experiment. *Social Psychological and Personality Science*, 1948550612469233.
6. Dick, J. (2013). Why Do Social Networks Increase Stress? *The Huffington Post*, 11/09/2013. Retrieved on October 9, 2016, from http://www.huffingtonpost.in/entry/social-networks-and-stress_b_3534170.
7. Drouin, M., Kaiser, D. H., & Miller, D. A. (2012). Phantom vibrations in undergraduates: Prevalence and associated psychological characteristics. *Computers in Human Behavior, 28*, 1490-1496.
8. Souri, A., Hosseinpour, S., & Rahmani, A. M. (2018). Personality classification based on profiles of social network users and the five-factor model of personality. *Human-Centric Computing and Information Sciences, 8*, 24. [CrossRef]
9. Morente-Molinera, J. A., Kou, G., Samuylov, K., Urena, R., & Herrera-Viedma, E. (2019). Carrying out consensual group decision-making processes under social

ISSN (E): 2942-190X Open Access| Peer Reviewed

networks using sentiment analysis over comparative expressions. *Knowledge-Based Systems*, *165*, 335-345. [CrossRef]

10. Risch, J., & Krestel, R. (2018). Aggression identification using deep learning and data augmentation. *Proceedings of the First Workshop on Trolling, Aggression, and Cyberbullying (TRAC-2018)*, Santa Fe, NM, USA, pp. 150-158.
11. Abualigah, L. M. Q. (2019). Feature Selection and Enhanced Krill Herd Algorithm for Text Document Clustering. Springer: Berlin/Heidelberg, Germany.
12. Budiharto, W., & Meiliana, M. (2018). Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis. *Journal of Big Data*, *5*, 51. [CrossRef]
13. Cury, R. M. (2019). Oscillation of tweet sentiments in the election of João Doria Jr. for Mayor. *Journal of Big Data*, *6*, 42. [CrossRef]
14. Al Shehhi, A., Thomas, J., Welsch, R., Grey, I., & Aung, Z. (2019). Arabia Felix 2.0: A cross-linguistic Twitter analysis of happiness patterns in the United Arab Emirates. *Journal of Big Data*, *6*, 33. [CrossRef]
15. Subramani, S., Michalska, S., Wang, H., Du, J., Zhang, Y., & Shakeel, H. (2019). Deep learning for multi-class identification from domestic violence online posts. *IEEE Access*, *7*, 46210-46224. [CrossRef]
16. Pong-inwong, C., & Songpan, W. (2018). Sentiment analysis in teaching evaluations using sentiment phrase pattern matching (SPPM) based on association mining. *International Journal of Machine Learning and Cybernetics*, *10*, 2177-2186. [CrossRef]
17. Aloufi, S., & El Saddik, A. (2018). Sentiment identification in football-specific tweets. *IEEE Access*, *6*, 78609-78621. [CrossRef]
18. Arias, M., Arratia, A., & Xuriguera, R. (2013). Forecasting with Twitter data. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *5*, 8. [CrossRef]
19. Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, *60*, 2169-2188. [CrossRef]