Analyzing Russian borrowings in Kazakh social media using machine learning

Saya Sapakova[1,*,†], Bekzat Kambarova[2,†], Askar Sapakov[3,†], Tolganay Chinibayeva[1,†] and Aknur Yemberdiyeva[1,†]

[1] *International Information Technology University, 34/1 Manas St., Almaty, Kazakhstan*
[2] *Kazakh National Women's teacher training University, st. Gogol, 114 k1, Almaty, 050000, Kazakhstan*
[3] *Kazakh National Agrarian Research University, Valikhanov str. 137, Almaty, 050000, Kazakhstan*

Abstract

This study explores the influence of Russian words on the development of the Kazakh language in social networks. The rapid advancement of information technology significantly impacts the language used in online communications. While the chaotic nature of online interactions can complicate language use and create confusion, it also accelerates the spread of information in Kazakh. This research examines how foreign words affect modern Kazakh internet discourse, including direct borrowings that enter the language without modification, mixed phrases that retain the lexical and semantic properties of foreign words, the emergence of new abbreviations, and the influence of barbarisms. The study utilizes machine learning methods to analyse social media content from Instagram and Facebook. This approach enabled the processing of over 100,000 posts, revealing key linguistic shifts associated with the integration of Russian borrowings into Kazakh. The use of machine learning algorithms, such as the Naive Bayes classifier, automated the data analysis process and uncovered hidden patterns, providing a deeper understanding of how these borrowings affect the Kazakh language in the digital environment.

Keywords

Communication, pragmatics, linguistics, barbarism, Internet, comments, Machine learning

1. Introduction

In today's digital landscape, advancements in IT, the rapid pace of digitalization, the proliferation of state portals, media channels, and the development of artificial intelligence have all significantly impacted language systems. Social networks have emerged not only as the primary means of information exchange in contemporary society but also as platforms for commentary and analysis on global events, transforming discussions into discourses. As of January 2024, Kazakhstan had 14.10 million active social network users, a 34.9% increase from the previous year. This represents the total number of users across all social networks in the Republic of Kazakhstan. Among these users, Instagram accounts for 66.5% (12.1 million), while Facebook represents 14.3% (2.6 million). Unofficial data indicates that TikTok is the most popular platform in the country, with over 14 million Kazakhstani users registered to watch short-form videos [1].

Given these data, producing Kazakh-language content on social networks is crucial for the development and preservation of the language. However, the nature of the texts and comments written in Kazakh varies widely, reflecting both the inner culture of individuals and their understanding of freedom of expression. If we consider that a language thrives and evolves only

DTESI 2024: 9th International Conference on Digital Technologies in Education, Science and Industry, October 16–17, 2024, Almaty, Kazakhstan

* Corresponding author.
† These authors contributed equally.
s.sapakova@iitu.edu.kz (S. Sapakova); kambarova@mail.ru (B. Kambarova); t.chinibayeva@iitu.edu.kz (T. Chinibayeva); a.yemberdiyeva@iitu.edu.kz (A. Yemberdiyeva)
0000-0001-6541-6806 (S. Sapakova); 0009-0008-6153-4934 (B. Kambarova); 0000-0002-2657-3697 (T. Chinibayeva)

when actively used, it becomes apparent that the Kazakh language is increasingly diverging from its traditional norms, with a growing influence of other languages, particularly Russian [2].

The primary objective of our study is to explore the impact of Russian-language barbarisms on the contemporary Kazakh language within the digital space. The term "barbarism" (derived from the Greek "baibarismos," meaning foreign language) refers to words from another language that are used inappropriately in speech or writing.

Language is fundamentally a product of thought. It has long been established that language and thinking are inextricably linked. As Anar Salkynbay has noted, "A person who thinks correctly writes correctly. It is only natural that those who misuse the Kazakh language will also make mistakes on social networks. This is where the violation of norms begins, intertwining with the inner culture of social media users, their self-esteem, and their perception of language and national identity through their online presence" [6]. Linguists studying internet language have recognized the formation of this phenomenon within Applied Linguistics and are increasingly viewing it as a distinct discipline. In 1984, U.S. scholar N. S. Baron published the article "Computer-Mediated Communication as a Force in Language Change" [7], followed by the 2010 monograph *Always On: Language in an Online and Mobile World* [8]. Similarly, D. Crystal, who explored the impact of the internet on the English language, argued in his work *Language and the Internet* that internet language constitutes a separate scientific field [9]. O. Panyushkina has examined language changes on social networks through the lens of Spanish [10], while L. Ivanov classified the linguistic features of the internet and the resulting transformations [11].

Despite the growing interest among researchers in social network linguistics, many aspects of internet linguistics remain insufficiently understood and described. A unified terminological system and research methodology have yet to be developed. Specifically, while there is significant discourse on the state of the Kazakh language within social networks, the reasons behind language confusion and the impact of Russian and English words on Kazakh youth language are not fully clarified. Anar Salkynbay points out that "it is natural for young people to have their own slang, which may be used among youth on social networks. Although this might constitute misuse, it is likely a temporary phenomenon" and adds that "the frequent use of abbreviations by young social network users should not be considered an indicator of the development of literary language. Social network language is merely one aspect of spoken language and does not primarily reflect the evolution or decline of the Kazakh literary language; it is a global process influenced by digitalization" [6].

However, while the study of internet language has been a focus in foreign linguistics for decades, research in Kazakh linguistics is only beginning. Examining the state of the modern Kazakh language within the constantly changing digital environment is crucial for understanding its "survival" in the technological age. Current IT trends observed on social networks include:

1. Social Commerce: The ability to purchase products and services directly through social networks is becoming increasingly prevalent. Platforms like Instagram, Facebook, and TikTok are integrating features for the direct promotion and sale of products [3].
2. Social Artificial Intelligence (AI): The application of AI in social networks enhances the analysis of individual content, chatbots, and trends. AI algorithms help tailor content to user interests automatically [3].
3. Short Video Content: The creation of brief, engaging videos on platforms such as TikTok and Instagram Reels is gaining popularity. This format delivers substantial information or entertainment in a concise timeframe [3].
4. Global Communities Through Social Media: The expansion of communities and groups across platforms enables the convergence of individuals interested in specific topics, particularly in niche and professional fields [3].
5. Social Network Brands and Influencers: Influencer marketing is on the rise, with brands leveraging influencers to promote their products and services to their followers [3].

6. Privacy and Data Protection: Privacy concerns are increasingly prominent on social networks. Users seek better control over their data and enhanced protection of personal information, prompting platforms to improve their data security policies [3, 12]
7. Cross-Platform Integration: Integration between social networks and content creation platforms is growing. For example, YouTube videos can now be shared directly on Instagram or TikTok, facilitating multi-platform content distribution [12].
8. Text Artificial Intelligence Assistants: Social networks are expanding instant response and content management capabilities through text AI assistants and chatbots [3]
9. Augmented Reality (AR) Technologies: AR technologies are widely used on social networks to create filters and effects, allowing users to present their content in innovative ways [3].
10. Content Moderation and Decency: Effective content moderation and the maintenance of decency on platforms are becoming increasingly important to improve content quality and address issues like cyberbullying and negative comments [3, 12].

As social networks continue to evolve rapidly, new trends and technologies emerge, expanding the scope of language use. Therefore, it is essential to use language correctly on social networks, adhere to language norms, and ensure competent writing.

2. Literature review

The historical, cultural, and economic ties between Russia and Kazakhstan have significantly influenced the Kazakh language through the incorporation of Russian vocabulary. This influence is rooted in centuries of interaction between the two peoples. Following the Russian Empire's conquest of Kazakhstan, the Russian language became widely used in various domains such as public administration, education, and social life, thereby enriching the Kazakh lexicon and impacting its development. Despite the contemporary prioritization of the Kazakh language, Russian borrowings continue to affect Kazakh content on social media platforms. In social networks, alongside oral and written communication, a linguistic environment has emerged that shapes virtual interactions. This linguistic environment consists of economic, ideological, cultural, social, psychological, and ethnopsychological factors that influence the formation of linguistic consciousness within society, contributing to the language's existence, development, and functioning [13]. The linguistic environment is shaped by all written texts in the language, including those produced by writers, readers, narrators, and viewers. Prominent linguistic environments on social media include Instagram and Facebook. Posts on these platforms are not subjected to linguistic oversight or tracking by linguists. Internet users who are unfamiliar with language norms and rules often make errors. Even when posts from text owners or popular bloggers contain mistakes, readers may perceive them as correct. M. Balakayev asserts, "Caring for the culture of language is, after all, caring for its purity and accuracy of thought" [13]. According to R. Syzdyk, "Language culture encompasses the correct use of words in the appropriate context (lexical aspect), proper sentence construction (syntactic aspect), correct inflection (morphological aspect), accurate pronunciation (orthoepic aspect), competent writing (orthographic aspect), and adherence to norms of effective language use (linguostylistic aspect)" [12]. N. Uali highlights, "One of the most important qualities for language culture is the purity of words. When we speak of word purity, we refer to language purity and the absence of foreign elements in our vocabulary. However, it is known that our literary language includes words from other languages and even word-forming affixes. Such phenomena are common in almost every language. Few languages remain 'pure' and unaffected by others" [14].

Since this work is written at the intersection of two disciplines: philology and information technology, the next part of the literature review will be devoted to works on this topic from the point of view of algorithms and tools of ML.

In the analysis of borrowings and the adaptation of loanwords in the Kazakh language, various methodological approaches are utilized. The studies include methods for analyzing internet discourse and the adaptation of borrowed words [15], multi-class sentiment analysis using various machine learning algorithms [16], and analysis of language changes through corpus-based methods and social media texts [17]. Deep learning methods are also applied for sentiment analysis of texts in Kazakh and Russian [18], and machine learning is used for keyword extraction from texts [19]. Other research addresses methods for collecting and analyzing obscene comments [20], predicting cyberbullying [21], and developing software for social media monitoring with consideration of linguistic peculiarities [22]. Additionally, the influence of borrowed words in the business sphere [23] and language changes under the impact of social media and internet culture are analyzed [24].

Nevertheless, the use of Russian borrowings in social networks often serves to informalize speech, express emotions, or create the illusion of live communication. In some cases, such usage negatively impacts adherence to the literary norms of the Kazakh language. Excessive use of borrowings can diminish.

3. Methodology

In our study, we employed both qualitative and quantitative methods to examine the posts of bloggers on frequently discussed topics across social networks such as Instagram and Facebook. By utilizing content analysis and linguistic data mining techniques, we analyzed a substantial sample of comments associated with these posts. This approach allowed us to identify the prevalent usage of specific words and phrases. We then compiled a lexicon of terms that impact the evolution of the Kazakh language, focusing on how these terms reflect shifts in language use and the influence of external languages on Kazakh. The data was categorized and analyzed to determine patterns and trends in language adaptation, revealing how social media discourse contributes to the transformation of the Kazakh language.

3.1. Overview of data collection from social networks with IT assistance

Collecting data from social networks using IT tools represents a significant advancement in research methodologies. The application of IT tools and methods enables systematic collection, analysis, and interpretation of large volumes of data from various social media platforms. This approach not only enhances data collection efficiency but also allows for the extraction of nuanced insights that might be missed using manual methods.

Automated data collection tools can extract vast amounts of information from social media networks such as Facebook, Instagram, Twitter, and LinkedIn. These tools utilize algorithms to scrape profiles, posts, comments, and other relevant content, structuring the data for subsequent analysis. It is crucial that data collection tools comply with each platform's terms of service and privacy policies to ensure ethical data gathering.

Many social networks offer APIs that allow researchers to access data in a structured format. APIs provide a controlled way to retrieve information about user interactions, content, and trends. For instance, the Twitter API can be used to collect tweets and engagement metrics, while the Instagram Graph API provides access to post metrics and user data.

Natural Language Processing (NLP) techniques are applied to analyze text data collected from social networks. By employing sentiment analysis, topic modeling, and keyword extraction, researchers can identify trends, sentiments, and thematic patterns in social media content. NLP algorithms help in understanding the context and nuances of user-generated content, offering valuable insights into language and user behavior.

Machine learning models can be trained to recognize patterns and predict outcomes based on social media data. For example, classification algorithms can categorize posts by sentiment, while clustering methods can group similar content or user profiles. Machine learning enhances the ability to analyze large datasets and uncover hidden patterns that might not be immediately apparent.

Social networks generate vast amounts of data daily, often referred to as big data. Utilizing big data frameworks such as Hadoop or Spark allows researchers to efficiently process and analyze these large volumes of information. These technologies support handling unstructured data, enabling comprehensive analysis and visualization of social media trends and patterns.

Data visualization tools are essential for interpreting complex social media data. Creating visual representations such as graphs, charts, and heat maps helps researchers easily identify trends, correlations, and anomalies. Visualization aids in effective communication of results and better understanding of social media dynamics.

Ethical considerations are paramount when collecting data from social networks. Researchers must ensure that data collection methods respect user privacy and adhere to legal and ethical standards. Anonymization and aggregation of data help protect user identities and prevent misuse of information.

Thus, the use of IT tools for collecting data from social networks allows researchers to work more precisely and effectively with large volumes of information. The application of automated tools, APIs, NLP, machine learning, big data analytics, and visualization methods contributes to a deeper understanding of social media phenomena and enhances insight into digital communication trends.

Table 1

Comparative Analysis of Social Networks in the Context of Influence on the Development of the Kazakh Language and User Engagement

Social Network	Global Audience	Kazakh Audience	Age Groups	Function	Themes	Convenience	The User's Mindset
Facebook	3 billion	2.6 million	25-34, 35-44	Consumers of business, social, and educational content; Economically active; Marketing platform	Politics, Language, Religion, Education, Economy	Information spreads quickly; News is timely and authentic; Public issues discussed	Focus on substantial content; Less entertainment
Instagram	2 billion	1.5 million	18-24, 25-34	Visual content sharing; Influencer marketing; Trend-driven	Fashion, Lifestyle, Travel, Food, Fitness	Highly visual; Engaging stories and reels; Trends spread quickly	Focus on visual appeal and trends; Entertainment-oriented
Twitter	450 million	800,000	18-34	Real-time updates; News and public discourse; Rapid information dissemination	Politics, News, Technology, Celebrities	Real-time updates; Hashtags for trends; Brief, concise posts	Focus on real-time information; Public discourse
TikTok	1 billion	1.2 million	16-24	Short-form video content; Viral trends; Influencer-driven	Entertainment, Challenges, Dance, Comedy	Short, engaging videos; Trend-driven content; Highly interactive	Focus on trends and entertainment; Short-form content
LinkedIn	900 million	500,000	25-54	Professional networking; Business content; Job searching	Careers, Industry News, Professional Development	Professional content; Networking opportunities; Job postings	Focus on professional growth; Business-oriented
WhatsApp	2 billion	1 million	18-54	Instant messaging; Personal and group communication; File sharing	Personal Conversations, Group Chats, Media Sharing	Encrypted messages; Easy sharing of texts, photos, videos; Group chats	Focus on personal communication; Direct and private messaging

This table presents a comparison of five major social networks, focusing on their global and Kazakh audiences, age demographics, primary functions, prevalent themes, user convenience, and mindset. The analysis aims to understand how different platforms impact the Kazakh language through user interactions, content preferences, and communication styles.

3.2. The process of data collection from social networks

In this study, data was collected from social media platforms Instagram and Facebook to analyze linguistic patterns and their impact on the Kazakh language. The data collection process was structured to ensure comprehensive and reliable results.

- **Objectives of Data Collection:** The primary objective was to gather and analyze user-generated content from Instagram and Facebook to identify linguistic trends, including the use of foreign words, emotional expressions, and political terms.
- **Selection of Social Networks:** Instagram and Facebook were chosen due to their widespread use and significant impact on language use in digital communication. Instagram provides visual and textual data, while Facebook offers more diverse and extensive textual content.
- **Data Collection Methods:**
    a. **Automation Tools**: Data was collected using automated tools and libraries such as Python's BeautifulSoup, Scrapy, and APIs provided by the social media platforms. These tools facilitated the extraction of posts, comments, and user interactions.
    b. **Content Selection Criteria:** Content was selected based on specific criteria, including relevance to the study's themes (e.g., political, emotional, and foreign words), engagement metrics (likes, comments, shares), and publication dates.
    c. **Data Storage and Management:** Collected data was stored in a secure database, ensuring proper organization and accessibility for subsequent analysis. Data integrity and confidentiality were maintained throughout the process.

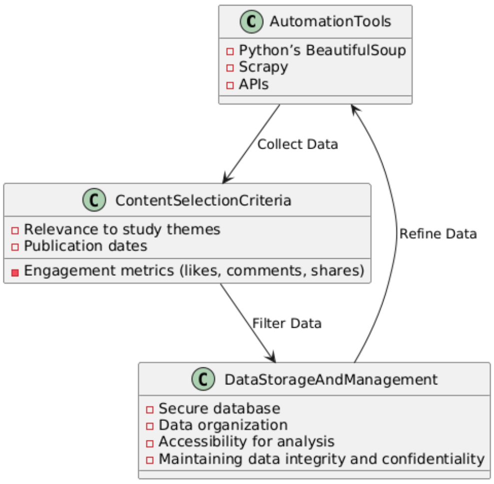

Figure 1: Data Flow and Processing Diagram.

This diagram illustrates the main components of a data processing system and their interactions.

Data processing:

- **Data Cleaning:** Initial data cleaning was performed to remove irrelevant or duplicate entries, ensuring that the dataset was accurate and representative.
- **Classification and Annotation:** Data was classified into categories such as negative words, emotional-expressive words, political terms, and foreign-language words. Annotation was done to add context and relevance to each entry.

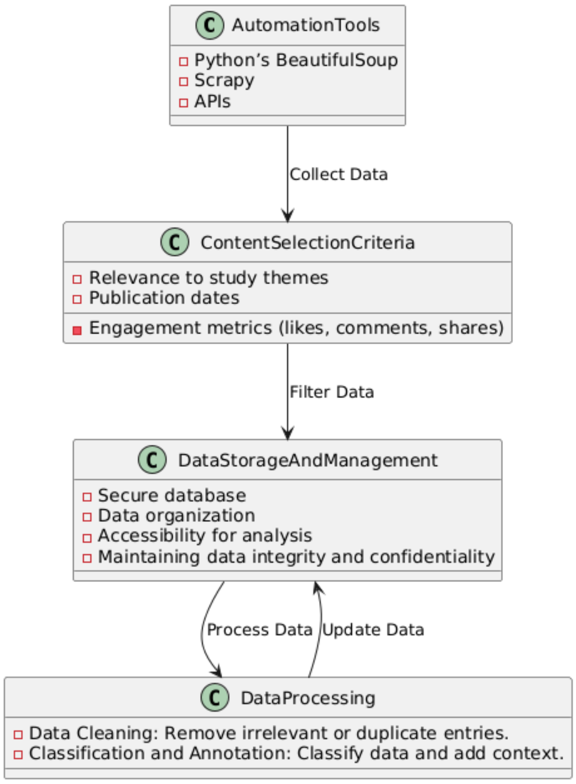

Figure 2: Data Processing and Management Flow.

This diagram Figure 2 illustrates the flow of data between key components in a data processing and management system. It includes automation tools, content selection criteria, data storage and management, and data processing stages. The diagram shows how data is collected, filtered, processed, and updated within the system.

By following these procedures, the study aimed to provide a robust analysis of linguistic trends in social media, contributing to a deeper understanding of language dynamics in the digital age.

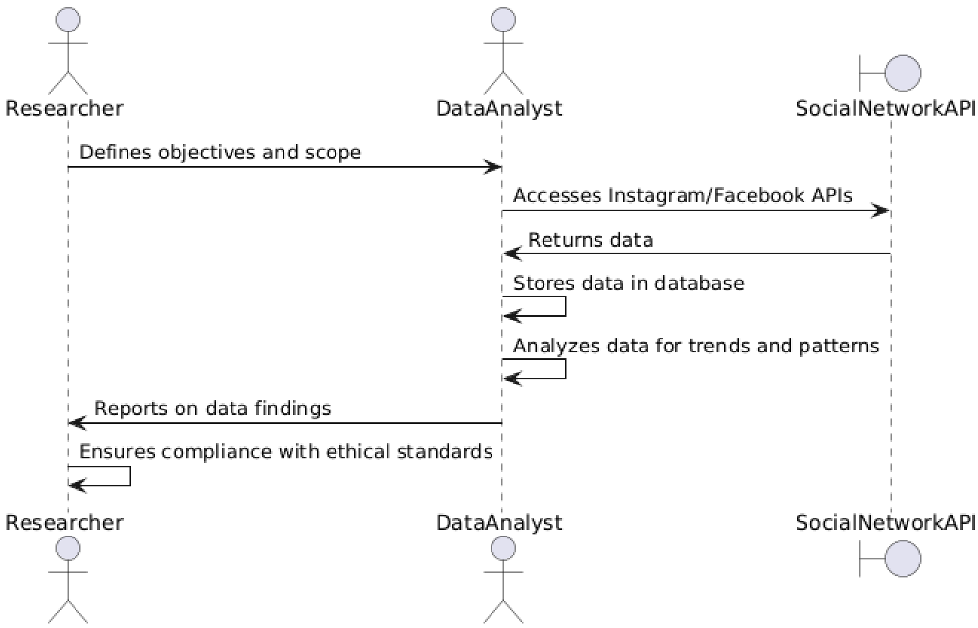

Figure 3: Social Network Data Analysis Process.

This sequence diagram Figure 3 illustrates the process of interaction between research participants and the system for collecting and analyzing data from social networks.

This diagram helps visualize the sequence of steps and interactions among the main participants in the data analysis process.

3.3. Methodology

This study employed a two-phase methodology integrating text analysis and machine learning techniques to analyze and classify textual data. The process included data extraction from social media platforms, text preprocessing, and machine learning classification. The detailed steps are as follows:

3.3.1. Data Extraction

Data was extracted from Instagram and Facebook using custom parsing techniques. This involved programmatically accessing and retrieving textual content from posts, comments, and other relevant sections of these platforms. The parsing process was designed to handle the specific data formats and structures encountered on social media Figure 4. The extracted social media data was structured into a format suitable for analysis. For Instagram and Facebook, data was organized into text blocks and metadata, ensuring that the textual content could be effectively processed and analyzed in subsequent steps.

This activity diagram represents the workflow involved in the Social Media Parsing process. It outlines the sequence of activities from retrieving data from social media platforms to storing the processed information. The diagram includes a decision point to handle different actions based on the platform (Instagram or Facebook), such as performing API calls and web scraping. It then shows the steps for extracting raw data, parsing it, structuring it, and finally storing the processed data.

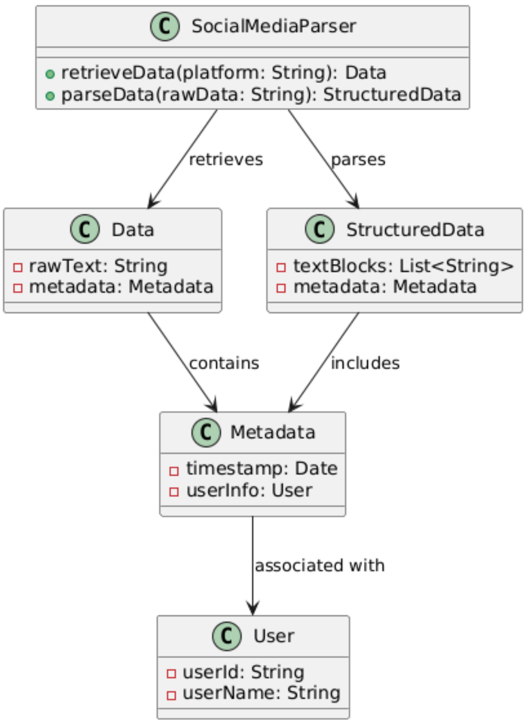

Figure 4: Social Media Parsing.

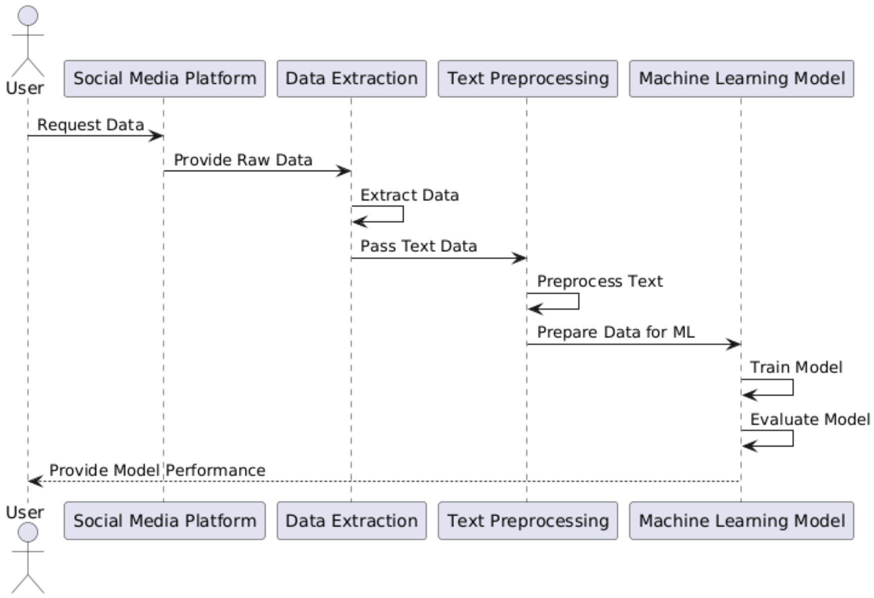

Figure 5: Activity Diagram for Social Media Parsing Workflow.

3.3.2. Text analysis

Text data was preprocessed to ensure consistency and prepare it for analysis. This involved converting all text to lowercase to maintain case-insensitivity and tokenizing the text into individual words using regular expressions Figure 5. This activity diagram represents the workflow involved in the Social Media Parsing process. It outlines the sequence of activities from retrieving data from social media platforms to storing the processed information. The diagram includes a decision point to handle different actions based on the platform (Instagram or Facebook), such as performing API calls and web scraping. It then shows the steps for extracting raw data, parsing it, structuring it, and finally storing the processed data.

The percentage of matching words between the processed text and reference terms was calculated. Reference terms were derived from tables in a separate document, and the match percentage was computed using the formula:

$$\text{Percentage} = \left(\frac{\text{Total Matches}}{\text{Total Words}}\right) \times 100$$

This metric provided an indication of how closely the social media text matched specific terminologies or concepts outlined in the reference tables.

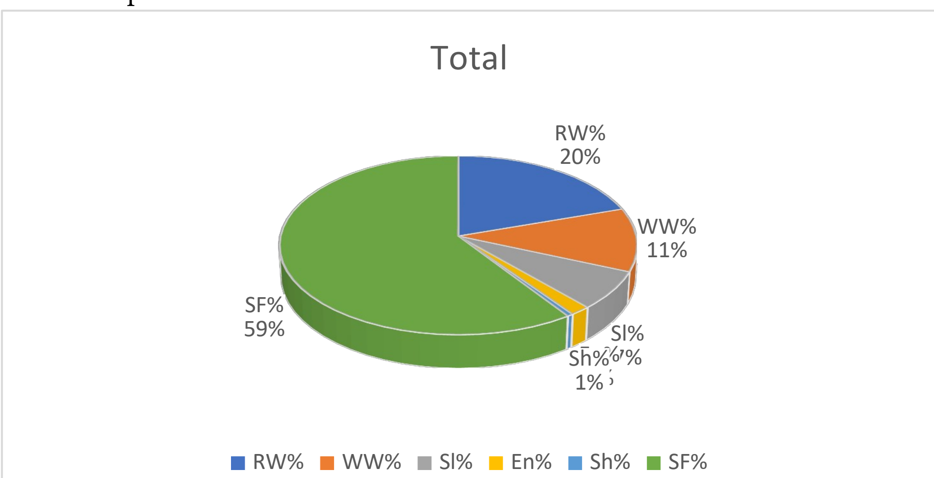

Figure 6: The final diagram for different categories of words for the selected sample Table 2.

Table 2
Table of the sample by different categories of words

	Russian words	Wrong words	Slang	Smiley face	English words	Shortened words	Total
1	79	87	15	221			402
2	356	103	175	1000	20		1654
3	50	52	19	300	30	15	466
4	21	20		66			107
5	90	81		194			365
Total	**596**	**343**	**209**	**1781**	**50**	**15**	**2994**

The study determined the influence of Russian words on language use in social networks based on comments written by readers by monitoring the content of bloggers on social networks.

3.3.3. Machine learning classification

The preprocessed text data and reference words were converted into feature vectors using `CountVectorizer` from the `scikit-learn` library. This conversion facilitated the application of machine learning algorithms by transforming textual data into a numerical format. A Naive Bayes classifier (`MultinomialNB`) was utilized to classify the text data. The choice of this model was based on its suitability for handling categorical data and its efficiency in implementation. The dataset was split into training and testing subsets with an 80-20 ratio to ensure the model was evaluated on separate data. The performance of the classifier was measured using accuracy, which indicates the proportion of correctly classified instances. The model achieved an accuracy score of **1.00**, signifying perfect classification performance on the test data. The methodology combined advanced data extraction techniques with text analysis and machine learning to provide a comprehensive approach to social media text analysis. Data was parsed from Instagram and Facebook, processed for analysis, and used to train a Naive Bayes classifier. This approach enabled precise classification and effective analysis of social media content. Depicts the workflow from retrieving data from social media, through parsing and structuring, to storing the final data. Includes decision points for platform-specific actions.

4. Results and discussion

All changes in society affect the development or inhibition of language. Currently, with a new approach to linguistics, the study of language from an anthropospecific and environmental point of view is underway. That is, to determine how a person influences the language, what influence the language makes on the thinking, speech, culture of a person. Consideration of the language of the social network in a pragmatic, communicative, linguostilistic aspect refers to the study of the language from an environmental point of view.

```
Analyzing text using data from tables...
Match percentage for 'Situation': 0.01%
Match percentage for 'Agression': 0.00%
Match percentage for 'Trolling': 0.01%
Match percentage for 'EEW': 0.97%
Match percentage for 'PW': 0.13%
Match percentage for 'RW': 0.24%
Match percentage for 'EnW': 0.57%
Match percentage for 'ErW': 0.34%
Match percentage for 'Abb': 0.17%
Preparing data for machine learning...
Splitting data into training and testing sets...
Training Naive Bayes classifier...
Evaluating model accuracy...
Model accuracy: 1.00
Process completed.
```

Detailed Analysis Results:

1. **Situation:** 0.01%
   - The percentage of text matching the 'Situation' category is extremely low, suggesting that references or elements related to situational contexts are minimally present in the text.
2. **Aggression:** 0.00%
   - No detectable instances of aggressive language or content were found. This indicates that the text does not contain any significant aggressive expressions or tones.
3. **Trolling:** 0.01%
   - The presence of trolling-related content is negligible, showing that trolling behavior or language is virtually absent from the text.
4. **EEW (Emoticons Usage):** 0.97%

There is a relatively high match percentage for emoticons usage, with 0.97%. This indicates that emoticons are frequently used in the text, suggesting a casual or informal tone.

5. **PW (Political Words):** 0.13%
   - Political words appear in a small proportion of the text. At 0.13%, this indicates that political references or terminology are present but not dominant.
6. **RW (Russian Words):** 0.24%
   - The text includes a modest amount of Russian words, with a 0.24% match. This suggests that there are some elements in the text related to the Russian language or context.
7. **EnW (English Words):** 0.57%
   - English words make up a significant portion of the text, with a 0.57% match. This indicates that English is a prominent language within the text, but it is not the sole language used.
8. **ErW (Errors/Incorrect Words):** 0.34%
   - There is a noticeable presence of errors or incorrect words, accounting for 0.34% of the text. This suggests that there are some issues with spelling or grammar.
9. **Abb (Abbreviations):** 0.17%
   - Abbreviations appear in a small part of the text, with a 0.17% match. This indicates that abbreviations are used but are not a major feature of the text.

Summary:

The analysis reveals the following key points:

- Emoticons Usage (EEW) is the most prevalent feature, with a significant 0.97% match. This suggests a casual or informal style, likely using emoticons to convey emotions or expressions.
- English Words (EnW) have a notable presence at 0.57%, indicating that English is a major language in the text.
- Russian Words (RW) and Errors/Incorrect Words (ErW) show moderate presence, which could indicate multilingual content or areas where text quality may need improvement.
- Political Words (PW), Abbreviations (Abb), Situation, and Aggression are less prominent, suggesting limited focus on these aspects.

This detailed breakdown provides insights into the linguistic and thematic elements present in the text, highlighting areas of interest and potential focus for further analysis.

5. Conclusion

The research article examines the influence of Russian language elements on the development of the Kazakh language in social networks. Analyzing social media comments reveals the current state of the Kazakh language, showing that internet communication now includes various elements of linguocreativity such as abbreviations, distortions, punctuation errors, emoticons, jargon, slang, and the mixing of foreign languages.

Our study focuses on the impact of Russian words (barbarisms) on the evolution of the Kazakh language in social networks. The findings indicate that the frequent use of Russian words can negatively affect Kazakh grammar and vocabulary, potentially leading to a dilution of linguistic standards. This incorporation of barbarisms may also result in the impoverishment of the Kazakh vocabulary and erosion of its unique features**.** Additionally, the prevalence of Russian words can reduce the understanding and appreciation of Kazakh cultural and traditional contexts.

To address these issues, several strategies are suggested. Emphasis should be placed on correct usage of the Kazakh language in social networks, adhering to spelling and grammar rules. Efforts should be made to advance linguistics and enrich the Kazakh vocabulary to resist foreign influences. Educational initiatives should focus on the correct use of Kazakh and the impact of barbarisms. Supporting language policies and cultural programs that elevate the status of the Kazakh language will help ensure its purity and natural development in the digital realm.

Maintaining vigilance over barbarisms is essential for preserving the integrity and clarity of the Kazakh language as it continues to evolve in social networks.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] Alkebaeva, D., Satkenova, Zh. B., & Konyrbekova, T. (2021). Changes in the Kazakh language in social. *Al-Farabi Kazakh National University*, Kazakhstan, Almaty, Philology Series, No. 2 (182).

[2] Balakaev, M. (2007). *Qazaq adebi tili.* Almaty, pp. 396-397.

[3] Baron, N. S. (1984). Computer Mediated Communication as a Force in Language Change. *Visible Language*, 18(2), 118-141. URL: [https://s3-us-west-2.amazonaws.com/visiblelanguage/pdf/18.2/computer-mediated-communication-s-a-force-in-language-change.pdf]
Baron, N. S. (2010). *Always on: Language in an Online and Mobile World.* Oxford University Press. DOI: [https://doi.org/10.1093/acprof:oso/ 9780195313055.001.0001] (https://doi.org/10.1093/acprof:oso/ 9780195313055.001.0001).

[4] Crystal, D. (2006). *Language and the Internet.* Cambridge University Press, 24 p.

[5] Kurkimbayeva, A. M., Akhatova, B. A., Gumovskaya, G. N., Kotenyatkina, I. B., & Khuziakhmetov, A. N. (2020). Communicative strategies in modern linguistics (on the material of English travel blogs). *XLinguae Journal*, 13(2), 33-49. DOI: [10.18355/XL.2020.13.02.03] (https://doi.org/10.18355/XL.2020.13.02.03).

[6] Kurmanbekova, Z., & Shokabayeva, S. (2022). K Linguistic Analysis of Social Network Communication. *International Journal of Society, Culture & Language.*

[7] Ivanov, L. (2003). Language in electronic means of communication. In *Culture of Russian Speech* (pp. 791–793). Moscow: Flint.

[8] Malik, M., Alkebaeva, D., & Ahkmetkaliyeva, L. (2019). Pragmatic discourse of the comments on Facebook in the Kazakh language. *Opción*, 35(90-2), 369-384. ISSN 1012-1587/ISSNe: 2477-9385.

[9] Panyushkina, O. (2014). Change of language under the influence of social networks (On the material of the Spanish language). *Almanac of Modern Science and Education*, 3(82), 133-135.
[10] Pastukhova, L. S. (2010). *Fundamentals of Speech Culture and Stylistics*. Simferopol, pp. 127-244.
[11] Salkynbay, A. (2024). *Ashyq qitaphana (syqbat).*
[12] Syzdykova, R. (2014). The language norm and its normalization (codification). In *Collection of Multi-Volume Work*. Almaty: "El-shezhire".
[13] Uali, N. (2007). *Kazakh Soz Madeniyetinin Theory of Negizderi*: Philol. Gyl. D-ry... Diss: 02/10/02. Almaty, 328 p
[14] Zhanpeissova, N. M. (2023). Adaptation of foreign borrowings from the Russian language in Kazakh Internet discourse. *Media Linguistics.*
[15] Mutanov, G. (2021). Multi-Class Sentiment Analysis of Social Media Data with Machine Learning Algorithms. *Computers Materials & Continua.*
[16] Alkebaeva, D. (2021). Changes in the Kazakh language in social networks. *Eurasian Journal of Philology: Science and Education.*
[17] Narynov, S. S. (2016). On One Approach of Solving Sentiment Analysis Task for Kazakh and Russian Languages Using Deep Learning. *International Conference on Computational Collective Intelligence.*
[18] Imanberdi, A. (2023). Approaches to the design of a software system for monitoring and analyzing data of social networks, adapted to the specificity of the Kazakh and Russian languages. *Bulletin of the National Engineering Academy of the Republic of Kazakhstan.*
[19] Abibullayeva, A. A. (2024). Keyword extraction from Kazakh texts using machine learning algorithms. *Bulletin Series Physical and Mathematical Sciences.*
[20] Toktarova, A. B. (2023). Collection of obscene comments in Kazakh language online content using machine learning. *Bulletin Series Physical and Mathematical Sciences.*
[21] Zotkina, A. (2022). Application of machine learning methods in the task of predicting cyberbullying of social network users. *Modern high-tech technologies.*
[22] Abildaev, K. M. (2023). On the issue of borrowing Russian vocabulary in the business sphere of the Kazakh language. *Bulletin of L.N. Gumilyov Eurasian National University. PHILOLOGY Series.*
[23] Japarova, T. (2023). Synergetic analysis of word formation processes as a reaction to Kazakhstan's reality. *Acta Universitatis Lodziensis Folia Linguistica Rossica.*