

## CORPUS ANALYSIS OF SPEECH UNITS IN SOCIAL NETWORKS

**Urinboeva Nazokat**

Master's degree student  
of the National University of Uzbekistan.  
E-mail: [nnazokat71@gmail.com](mailto:nnazokat71@gmail.com)

**Kuvondikova Gavhar Isomiddinovna**

Teacher of the National University of Uzbekistan.

**Abstract:** This article describes the use of an electronic database to study and verify the analysis of speech units in social media with the help of a corpus, the identification of marked speech changes, their specific characteristics and the reasons for their changes.

**Keywords:** NLP, API, Twitter, Reddit, Facebook, LinkedIn.

**Аннотация:** В данной статье описано использование электронной базы данных для изучения и проверки анализа речевых единиц в социальных сетях с помощью корпуса, выявления выраженных речевых изменений, их специфических особенностей и причин их изменений.

**Ключевые слова:** NLP, API, Twitter, Reddit, Facebook, LinkedIn.

### Introduction

Corpus analysis of speech units in social networks means identifying speech changes in Internet networks, analyzing them, determining the causes of the specified language changes, and creating a set of data about their specific characteristics. Texts on the Internet, posts on social media, blogs, forums, news and other data are considered the main sources, using corpora to study the prevalence of speech changes, their conversation topics, what thoughts or trends to pay attention to, or other important data can be obtained.

### Literature analysis and methods

Many scientists have been studying changes in social processes through corpus analysis for several decades. However, the specific starting point may vary depending on the context and the particular line of research. The 1970s and 1980s saw the founding of corpus linguistics, with major contributions by scholars such as John Sinclair and Randolph Quirk, who focused more on linguistic description and analysis, and later on the study of social processes through corpus analysis. . In the 1990s, corpus linguistics began to expand its scope to include sociolinguistic inquiries. Scholars such as Douglas Bieber and Susan Conrad began using corpus methods to study language change in different social contexts. From the 2000s to the present, the use of corpus analysis in the study of changes in social processes has grown significantly. Advances in technology and the availability of large digital corpora have facilitated more sophisticated analyses.

Dan Thomas [1], David Bamman [4], Jacob Eisenstein [5], Cynthia Van Hee [6], David Juergens [7] have conducted research using social media discourse analysis, including Twitter

data. Their work explored linguistic and sociolinguistic aspects of online communication, including language change, sentiment analysis, and social dynamics.

Also, in a corpus-based analysis of changes in social processes, Timothy Graham [8], Nathan Mathias [9], Bernhard Rieder, Scott Hale, Christoph A. Hafner, and other researchers together examine various social phenomena, including political discourse, community formation, and language use. conducted research using Reddit data to explore topics related to models, online communities, collective action, and the impact of technology on social processes. Scientists David Lazer, Michael S. Bernstein [10] conducted a study using Facebook data to study social phenomena, such as information sharing, communication patterns and topics related to social networks on the platform, political communication, and collective behavior in communities. They used Facebook as a corpus to analyze social processes, online behavior, and changes in communication patterns within the platform [Abdurakhmonova, N. 2019,2021,2022].

In particular, the study of processes has developed over time, with the contribution of sciences and scientists. It remains an active area of research with constant changes in methodology and theory. The process of analyzing speech changes in social networks with the help of a corpus is carried out in certain stages.

First, select a suitable dataset for analysis and this dataset will include data from social networks such as Twitter, Facebook, Reddit, blogs, news portals and other sources. Data can be retrieved using automated or manual downloads, APIs (Application Programming Interfaces) or browser extensions to retrieve the required data from the collection. The received information is placed in a corpus or database, transferred to a suitable format and stored. It is also possible to use automated algorithms or data mining methods to detect language changes from the data in the corpus.

At the same time, a detailed analysis of selected speech changes can be done through language learning, distribution, exchange of ideas, use of words and expressions, comparison of actions on words, and other factors. The results of the analysis are then made available to data subjects, end users, or employees who have access to the data. Results can be presented through graphs, statistics, presentations and other visualizations. A detailed explanation of the results leads to a greater understanding of social media changes, trends, relationships between people, and more [Abdurakhmonova, N. 2019,2021,2022]. These results are fundamental to the knowledge and learning gained from the analysis of language change. Each step of this process requires challenge and reinforcement in data collection, analysis, and presentation of results.

When analyzing speech units in social networks using a corpus, it is necessary to pay attention to the following important aspects:

- topics of conversation in the language, such as new fashion, political events, technology or sports;
- in the process of analysis, the most used words and expressions in the network are people's opinions, important discussions in the network;
- how long will the exchange of ideas between users on the network change, what important negotiations will be conducted and what issues will be contested;
- the methods of language in social networks, writing, speaking, framing, memoirs and other features, changes in relations between people;
- the phonetics of the language, peculiarities, new words, unchanged words, changes in the style of speaking, specific analytical features of the language and directions of its development;

- discussions, opinions popular in social networks, words of people in social networks, reactions, changes and how they react to social processes in the world;

-in the analysis process, the data is analyzed using statistical analysis, related to many indications, such as how much speech units expand or decrease, which words or topics are used a lot, how the relationship between users' changes.

There are many factors to consider in the analysis process and can be customized for each project or request. The key here is to learn how to analyze language changes and to challenge the results in an advanced manner.

### Practical significance

Corpus analysis of changes in social processes refers to the study of patterns, dynamics, and phenomena that develop within societies by analyzing large collections of texts or corpora [Abdurakhmonova, N. 2019,2021,2022]. This approach refers to the study of how language, behavior, attitudes, and interactions change over time under the influence of various social, cultural, economic, and technological factors.

There are some key aspects of changes in corpus analysis of social processes. Corpus analysis allows researchers to track linguistic changes in a given language or across languages over time. By examining linguistic features such as vocabulary, syntax, discourse structures, and stylistic conventions, researchers can identify changes in language use and how they reflect broader social trends [Abdurakhmonova, N. 2019,2021,2022].

Corpus analysis makes it possible to study social dynamics and laws of interaction within societies, groups and societies. Researchers can study how people communicate, share information, form relationships, express opinions, and negotiate in a variety of social contexts, including online platforms, formal institutions, and everyday interactions.

Also, through corpus analysis, scholars can investigate cultural evolution and innovation by tracking the spread, adoption, and adaptation of cultural artifacts, practices, and norms in societies. This includes the study of popular culture trends, media representations, consumer behavior, and cultural expressions over time.

### Conclusion

In conclusion, the corpus analysis of changes in social processes creates a rich basis for studying the complex interaction between language, culture, and society. By systematically analyzing large data sets of textual materials, researchers can gain insight into the dynamics of human behavior, communication, and social organization in diverse temporal, spatial, and cultural contexts.

### References:

1. Computational Linguistics: An Introduction (Studies in Natural Language Processing). Cambridge University Press.
2. Taweh Beysolow. Applied Natural Language Processing with Python. California, 2018.
3. Michele Zappavigna. Discourse of Twitter and Social Media: How We Use Language to Create Affiliation on the Web. A&C Black, 2012.
4. David Bamman, Noah A. Smith. Contextualized Sarcasm Detection on Twitter. International Conference on Web and Social Media. Computer Science, Linguistics, 2015.

5. Abdurakhmonova, N., Tuliyeu, U., Ismailov, A., & Abduvahobo, G. (2022). Uzbek electronic corpus as a tool for linguistic analysis. In *Компьютерная обработка тюркских языков. TURKLANG 2022* (pp. 231-240).
6. Abduraxmonova, N. Z. Q., & Urazaliyeva, M. Y. (2022). O 'zbek tili elektron korpusida (<http://uzbekcorpus.uz/>) og 'zaki matnlar korpusini yaratishning nazariy va amaliy masalalari. *Academic research in educational sciences*, 3(3), 644-650.
7. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). Development of intellectual web system for morph analyzing of uzbek words. *Applied Sciences*, 11(19), 9117.
8. Jacob Eisenstein. Introduction to Natural Language Processing Adaptive Computation and Machine Learning series. MIT Press, 2019.
9. Cynthia Van Hee, Els Lefever, Veronique Hoste. Exploring the Realization of Irony in Twitter Data. International Conference on Web and Social Media, 2016.
10. D. Nguyen, A. Seza Doğruöz, F. D. Jong. Computational Sociolinguistics: A Survey. Computational Linguistics, 2016.
11. L. Jenny, Davis, T. Graham. Emotional consequences and attention rewards: the social effects of ratings on Reddit. *Information Communication and Society* 24(4):1-18, 2021.
12. J. Nathan Matias. Going Dark: Social Factors in Collective Action Against Platform Operators in the Reddit Blackout. Microsoft Research, 2015.
13. S. Michael, E. Bakshy, B. Karrer. Quantifying the invisible audience in social networks. International Conference on Human Factors in Computing Systems, 2013.
14. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In *of the International Conference on Language Technologies for All (LT4All)*.
15. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In *Proceedings of the 11th Global Wordnet conference* (pp. 8-19).