



FACTUAL ERRORS IN JOURNALISTIC TEXTS: A CORPUS-BASED ANALYSIS

Dildora Xaitbaevna Turaeva

Lecturer, Uzbekistan University of Journalism and
Mass Communications

<https://doi.org/10.5281/zenodo.19309984>

Abstract: This article analyzes factual errors in journalistic texts using a corpus linguistics approach. Within the framework of the study, a corpus of media texts published in mass media is compiled and systematically examined to identify and classify different types of factual inaccuracies. In particular, numerical errors, contextual distortions, and misinterpretations are explored as key categories of factographic errors. The study investigates the linguistic and journalistic factors contributing to the emergence of these errors. By applying corpus linguistic methods, the frequency, distribution, and functional characteristics of errors within the texts are identified and analyzed. The findings of the research contribute to improving the accuracy and credibility of journalistic content and provide scientific insights for enhancing fact-checking practices in media production.

Keywords: *journalistic text, factual errors, corpus linguistics, fact-checking, numerical errors, contextual distortion, misinterpretation, media analysis.*

Introduction: In modern journalism, the accuracy of information is a fundamental principle that ensures public trust and credibility of media institutions. However, the increasing speed of information production and dissemination, especially in digital environments, has led to a rise in factual errors within journalistic texts. These errors can take various forms, including numerical inaccuracies, contextual distortions, and misinterpretations, all of which negatively affect the quality of information delivered to the audience.

In recent years, corpus linguistics has emerged as an effective methodological approach for analyzing large volumes of textual data in a systematic and empirical manner. By applying corpus-based methods, researchers are able to identify recurring patterns, frequency distributions, and structural characteristics of language use in media texts. This makes corpus linguistics particularly suitable for examining factual errors in journalism, as it allows for both quantitative and qualitative analysis. Despite the growing interest in misinformation and media reliability, there is still a lack of studies that specifically focus on factual errors from a corpus linguistic perspective. Therefore, this study aims to fill this gap by analyzing journalistic texts through a corpus-based approach, identifying the types, frequency, and underlying causes of factual



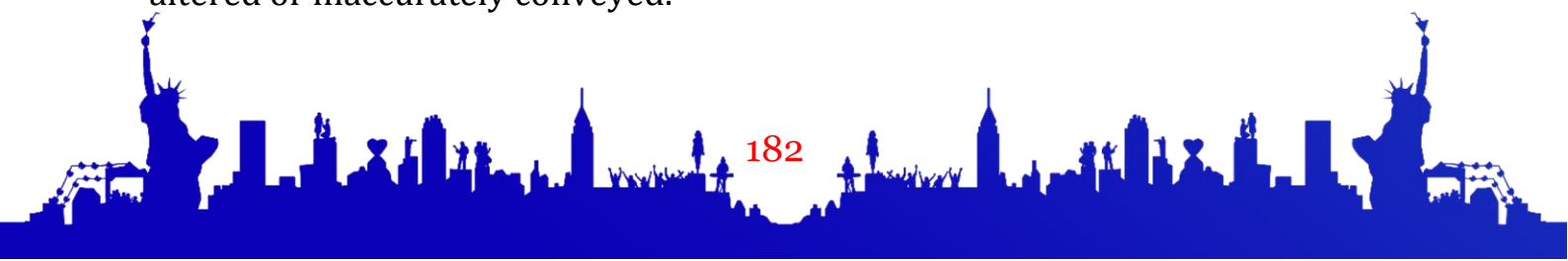


inaccuracies. The significance of this research lies in its potential to contribute to improving journalistic standards, enhancing fact-checking practices, and promoting more reliable information dissemination in mass media.

Methodology: This study adopts a corpus-based research design to analyze factual errors in journalistic texts. The research is based on the compilation and examination of a specialized corpus consisting of media texts collected from various mass media sources, including online news platforms, digital publications, and official news websites. The corpus includes approximately 100–150 journalistic texts selected through purposive sampling. The selection criteria focus on texts that contain identifiable factual inconsistencies, such as numerical inaccuracies, misleading contextual information, and interpretative errors. The texts are manually annotated to classify different types of factual errors according to predefined categories. The methodological framework combines both quantitative and qualitative approaches. Quantitative analysis is used to determine the frequency and distribution of different types of factual errors across the corpus. This includes calculating the occurrence rates of numerical errors, contextual distortions, and misinterpretations. Qualitative analysis, on the other hand, is employed to examine the linguistic and contextual features of these errors, as well as the factors contributing to their occurrence.

In addition, concordance analysis and keyword analysis are applied to identify patterns of language use associated with factual inaccuracies. Comparative methods are also used to analyze differences between various types of media sources in terms of error frequency and characteristics. To ensure the reliability and validity of the research, the study follows systematic data selection and classification procedures, and the findings are interpreted within the framework of established theories in corpus linguistics and journalism studies.

Results and discussion: The analysis of the compiled corpus reveals that factual errors in journalistic texts occur with noticeable frequency and follow identifiable patterns. The quantitative results indicate that numerical errors account for approximately 35% of all identified inaccuracies, making them the most common type. These include incorrect statistics, misreported figures, and inconsistencies in numerical data. Contextual distortions constitute around 33% of the errors, often resulting from incomplete information, omission of relevant details, or misleading framing of events. Misinterpretations account for approximately 32%, reflecting cases where the original meaning of information is altered or inaccurately conveyed.





The corpus-based analysis further demonstrates that factual errors are not randomly distributed but tend to occur more frequently in fast-paced digital media environments, where the pressure for rapid publication often compromises accuracy. Concordance and keyword analysis reveal that certain lexical patterns, such as the use of vague expressions, exaggerated terms, and ambiguous references, are strongly associated with the occurrence of errors. From a qualitative perspective, the findings suggest that factual errors are closely linked to both linguistic and journalistic factors. Linguistically, errors arise from improper word choice, ambiguity, and lack of precision. Journalistically, they are often the result of insufficient fact-checking, reliance on unverified sources, and time constraints in news production. Moreover, comparative analysis shows that online media platforms exhibit a higher frequency of factual errors compared to traditional media sources, highlighting the impact of digitalization on journalistic practices. These findings confirm the effectiveness of corpus linguistics as a methodological tool for identifying and analyzing patterns of factual inaccuracies in media texts.

Conclusion: This study has explored factual errors in journalistic texts through a corpus linguistics approach, providing both quantitative and qualitative insights into their nature and distribution. The results demonstrate that factual inaccuracies are a significant issue in modern journalism, with numerical errors, contextual distortions, and misinterpretations being the most prevalent types. The findings highlight that factual errors are influenced by a combination of linguistic and journalistic factors, including imprecise language use, inadequate verification processes, and the increasing speed of information production in digital media. The study also confirms that corpus-based methods are highly effective in identifying patterns and frequencies of errors, offering a systematic framework for analyzing large-scale textual data.

In conclusion, this research contributes to the improvement of journalistic standards by emphasizing the importance of accuracy, precision, and rigorous fact-checking practices. It also underscores the need for integrating corpus linguistics into media studies as a valuable tool for enhancing the quality and reliability of journalistic content. Future research may focus on expanding the corpus size, applying automated error detection techniques, and exploring cross-linguistic comparisons of factual errors in media texts.

References:

1. McEnery, Tony, Hardie, Andrew Corpus Linguistics: Method, Theory and Practice. – Cambridge: Cambridge University Press, 2012. – 294 p.





2. Biber, Douglas, Conrad, Susan, Reppen, Randi Corpus Linguistics: Investigating Language Structure and Use. – Cambridge: Cambridge University Press, 1998. – 300 p.
3. Bell, Allan The Language of News Media. – Oxford: Blackwell, 1991. – 277 p.
4. Allan, Stuart News Culture. – 3rd ed. – Maidenhead: Open University Press, 2010. – 238 p.
5. Abdujabborova, K. (2025). Methodology of English Language Teaching to Journalism Students: Experience of Foreign Countries. Ethiopian International Journal of Multidisciplinary Research, 12(02), 42-46.
6. Kamola Xusniddinqizi Abdujabarova. (2025). IMPORTANCE AND DEVELOPMENT OF ENGLISH LANGUAGE COMPETENCES IN MODERN JOURNALISM. International Multidisciplinary Journal for Research & Development, 12(10), 310–315. Retrieved from <https://www.ijmrd.in/index.php/imjrd/article/view/3864>
7. Abduganiyevna, A. D. (2025). PROBLEMS OF TRANSLATING METAPHORS AND FIGURATIVE LANGUAGE: A CASE OF UZBEK LITERATURE. Shokh Articles Library, 1(1).
8. Kasimova, S. (2025). TRANSLATION OF EXAMPLES OF EASTERN LITERATURE INTO UZBEKISTAN: ISSUES OF STYLE, CONTENT AND AESTHETIC CONSISTENCY. International Journal of Artificial Intelligence, 1(3), 116-119.

