Latin American Journal of Education
www.lajoe.org

DIGITAL CORPUS-BASED APPROACHES IN CONTEMPORARY FRENCH LINGUISTICS

https://doi.org/10.5281/zenodo.18920403

Mirzayev Tukhtasin Adhamjonovich
Lecturer at the Department of French Language Theory and Practice
Andijan State Institute of Foreign Languages, Andijan, Uzbekistan
e-mail: toxtasinmirzayev@gmail.com
https://orcid.org/0009-0002-2417-8754

Abstract: *This article examines modern approaches to the study of the French language through corpus linguistics and digital technologies. The research focuses on theoretical foundations, methodological principles, and practical applications of corpus-based linguistic analysis in French linguistics. Special attention is given to lexical, grammatical, and semantic patterns revealed through digital corpora and computational analysis. The study demonstrates that corpus linguistics and artificial intelligence significantly contribute to the understanding of current linguistic changes in French.*

Keywords*: French linguistics, corpus linguistics, digital linguistics, computational linguistics, language change, semantics.*

The rapid development of digital technologies has significantly transformed modern linguistic research. Traditional approaches to language analysis are increasingly complemented by digital methods that allow researchers to analyze large amounts of textual data with high precision. In this context, corpus linguistics has become one of the most influential fields in contemporary linguistic studies. Corpus linguistics refers to the systematic analysis of large collections of authentic texts stored in electronic form. According to Tony McEnery and Andrew Hardie, corpus linguistics enables researchers to investigate language patterns based on empirical evidence rather than intuition (1, p. 7). This methodological shift has profoundly influenced linguistic research in many languages, including French.

In French linguistics, corpus-based research has opened new possibilities for analyzing lexical innovation, syntactic variation, and semantic change. The French language has undergone considerable transformation in the digital age, particularly due to globalization, technological communication, and increasing contact with other languages. Scholars such as François Rastier argue that digital corpora allow linguists to examine linguistic phenomena across various discourse types, including literary

Latin American Journal of Education
www.lajoe.org

texts, media discourse, and digital communication (2, p. 84). As a result, corpus linguistics provides valuable insights into both the structure and the evolution of modern French.

Another important aspect of modern linguistic research is computational linguistics, which integrates linguistic theory with computer science. Computational models enable automatic text processing, syntactic parsing, and semantic analysis. According to Noam Chomsky, although linguistic competence is fundamentally cognitive, computational tools can help reveal structural regularities in language usage (3, p. 45).

The main objective of this article is to analyze the role of corpus linguistics and digital technologies in the study of contemporary French. The research focuses on theoretical frameworks, methodological principles, and practical applications of digital linguistic analysis. Corpus linguistics emerged as a distinct field in the second half of the twentieth century, although the use of text collections in linguistic studies dates back much earlier. Early structural linguists emphasized the importance of authentic linguistic data in describing language systems. One of the pioneers of corpus-based analysis was John Sinclair, who emphasized that linguistic research should be grounded in real language use rather than invented examples (4, p. 101). Sinclair's work on lexical patterns demonstrated that words often occur in predictable combinations known as collocations.

Collocational analysis has become an essential component of corpus linguistics. For instance, in French, certain verbs frequently occur with specific nouns or prepositions. By analyzing large corpora, linguists can identify such patterns and determine their frequency and distribution. In addition to collocations, corpus linguistics also investigates grammatical structures. According to Douglas Biber, corpus-based methods enable researchers to identify systematic variations in grammatical usage across different genres and registers (5, p. 67). This approach has been widely applied in French linguistics to study differences between spoken and written language. Furthermore, corpus linguistics contributes to semantic analysis. Large-scale text data allow linguists to trace semantic shifts and metaphorical extensions of words over time. For example, digital communication has introduced numerous new lexical items and semantic innovations in French.

Digital corpora represent structured collections of texts stored in electronic format and annotated for linguistic analysis. These corpora may include literary works, newspapers, academic texts, social media content, and spoken language transcripts. One of the most widely used resources in French linguistics is the Frantext, which contains thousands of French literary and scientific texts spanning several centuries.

Latin American Journal of Education
www.lajoe.org

Researchers use this corpus to analyze historical language changes and lexical developments.

Another important resource is the Corpus de Référence du Français Contemporain, which provides a comprehensive representation of contemporary French usage.

Digital corpora offer several advantages:

Large data volume Linguists can analyze millions of words, which increases the reliability of research findings.

Authentic language usage Corpus data reflect real communication rather than artificially constructed examples.

Statistical analysis Computational tools allow researchers to identify patterns and trends that would be difficult to detect manually.

As Geoffrey Leech notes, corpus linguistics bridges the gap between theoretical linguistics and empirical research by providing quantitative evidence for linguistic hypotheses (6, p. 112). The integration of artificial intelligence into linguistic research has significantly expanded the possibilities of language analysis. Computational linguistics uses algorithms and machine learning techniques to process and interpret linguistic data. Modern natural language processing systems can automatically identify grammatical categories, analyze sentence structure, and detect semantic relations. These technologies are increasingly applied in French linguistic research.

According to Christopher Manning, machine learning models trained on large corpora can reveal complex linguistic patterns that traditional methods may overlook (7, p. 203).

In French linguistics, computational tools are widely used in:

automatic translation systems

speech recognition technologies

sentiment analysis in digital communication

syntactic parsing and discourse analysis

These technologies also support lexicographic research by identifying new words and expressions appearing in digital discourse. The digital era has introduced significant changes in the structure and usage of the French language. One of the most noticeable developments is lexical innovation. Globalization and technological communication have led to the adoption of numerous loanwords from English. According to Jean Pruvost, lexical borrowing is a natural process that reflects cultural and technological exchange (8, p. 59).

Another important phenomenon is the emergence of internet slang and abbreviations. Digital communication platforms encourage shorter and more informal forms of expression, which gradually influence standard language usage.

Latin American Journal of Education
www.lajoe.org

Corpus-based research has shown that many new lexical forms originate in online communication before entering mainstream media and everyday speech. Furthermore, syntactic simplification has been observed in informal digital discourse. Sentences tend to be shorter, and punctuation rules are often relaxed.

These changes illustrate how language evolves in response to social and technological factors. The present research employs several methodological approaches:

corpus-based linguistic analysis
comparative analysis
statistical frequency analysis
computational text processing

The study analyzes a selection of French texts representing various discourse types, including literary works, journalistic texts, and online communication. Frequency analysis is used to identify common lexical and grammatical patterns. Computational tools assist in processing large datasets and visualizing linguistic trends. The results of corpus-based analysis demonstrate that digital methods provide valuable insights into contemporary French linguistic processes. One of the main advantages of corpus linguistics is its empirical nature. Unlike traditional linguistic research based primarily on theoretical assumptions, corpus studies rely on real data. This approach enhances the reliability and objectivity of linguistic analysis.

Moreover, digital corpora enable researchers to investigate language change over time. Historical corpora make it possible to compare linguistic patterns across different periods. Another significant advantage is the interdisciplinary nature of digital linguistics. Collaboration between linguists, computer scientists, and data analysts has led to the development of sophisticated analytical tools. However, some limitations should also be acknowledged. Corpus data may not always represent all varieties of language equally. Certain social or regional dialects may be underrepresented in digital corpora. Despite these challenges, corpus linguistics remains one of the most promising approaches in modern language research.

In conclusion, corpus linguistics and digital technologies have become essential tools in contemporary French linguistic research. The analysis of large digital corpora allows scholars to investigate lexical, grammatical, and semantic phenomena with unprecedented accuracy. Computational linguistics and artificial intelligence further expand the possibilities of linguistic analysis by enabling automatic text processing and pattern recognition. The integration of these methods contributes to a deeper understanding of language change and variation in modern French. Future research should continue developing digital corpora and analytical tools in

Latin American Journal of Education
www.lajoe.org

order to explore new aspects of linguistic structure and communication.

REFERENCES:

1. McEnery, T., Hardie, A. *Corpus Linguistics: Method, Theory and Practice.* Cambridge University Press, 2012, p. 7.
2. Rastier, F. *Sémantique interprétative.* Paris: Presses Universitaires de France, 2009, p. 84.
3. Chomsky, N. *Aspects of the Theory of Syntax.* MIT Press, 1965, p. 45.
4. Sinclair, J. *Corpus, Concordance, Collocation.* Oxford University Press, 1991, p. 101.
5. Biber, D. *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge University Press, 1998, p. 67.
6. Leech, G. *The State of the Art in Corpus Linguistics.* London: Longman, 2001, p. 112.
7. Oripovna, A. I. (2025). Nominativ birliklarning kommunikativ-pragmatik tamoyillari (fransuz va o ‘zbek tillari misolida). American journal of education and learning, 3(4), 265-268.
8. Oripovna, A. I. (2026). FRANSUZ VA O ‘ZBEK TILLARIDA SODDA GAPLAR TIPOLOGIYASI. Latin American journal of education, 6(2), 164-168.
9. Irodaxon Anorboyeva: 14. Анарбоева, И. О. (2025). Гендерные принципы номинативных единиц (на примере французского и узбекского языков). FARS International Journal of Education, Social Science & Humanities, 13(2), 136-139.
10. Normatov Azamatbek Abduhalilovich, Khasanbayeva Nafisahon Olimjonovna, Sotvoldiyeva Istorahon Lutfullo kizi, Samatova Zulhumor Qudtaillla kizi, Shamsutdinova Nazokat Alisherovna (2024). The Efficacy Of Teacher-Centered And Student-Centered Approaches In Foreign Language Teaching: A Quantitative Analysis. Library Progress International, 44(3), 13513-13530. DOI: https://doi.org/10.48165/bapas.2024.44.2.1
11. Oripovna, A. I. (2025). Nominativ birliklarning kommunikativ-pragmatik tamoyillari (fransuz va o ‘zbek tillari misolida). American journal of education and learning, 3(4), 265-268.
12. Oripovna, A. I. (2026). NUTQIY MULOQOTDA BILVOSITA NOM TANLASH MUAMMOSI: SEMIOTIK-MADANIYATSHUNOSLIK, SOTSIOLINGVISTIK VA GENDER JIHATLARI. AMERICAN JOURNAL OF EDUCATION AND LEARNING, 4(2), 514-517.