



## ETHNICAL CONSIDERATIONS IN CORPUS CONSTRUCTION AND USE

***Khusniddinova Sevinchoy Sirojiddin qizi***

*Faculty of Philology, Jizzakh State Pedagogical University, Jizzakh,  
Uzbekistan*

[\*xusniddinovasevinchoy15@gmail.com\*](mailto:xusniddinovasevinchoy15@gmail.com)

***Abdullajonova Khakima Abduqadirovna***

*Senior lecturer at Jizzakh State Pedagogical University, Jizzakh, Uzbekistan*

[\*hakimabotirova9@gmail.com\*](mailto:hakimabotirova9@gmail.com)

***Abstract:*** *This article analyzes the ethical considerations involved in corpus construction and use in linguistic research. As corpora increasingly include data from real-life communication, ethical issues such as informed consent, privacy and anonymity, copyright, fair representation, and responsible data usage become essential. The study highlights the importance of protecting participants and ensuring transparency in research practices. By following ethical guidelines, corpus linguistics can produce reliable and trustworthy results while respecting individual rights.*

***Key words:*** *Corpus linguistics; ethical considerations; informed consent; privacy and anonymity; copyright; data protection*

***Аннотация:*** *В данной статье рассматриваются этические аспекты создания и использования корпусов в лингвистических исследованиях. Поскольку корпуса все чаще включают данные из реального общения, этические вопросы, такие как информированное согласие, конфиденциальность и анонимность, авторское право, справедливое представление и ответственное использование данных, становятся крайне важными. Исследование подчеркивает важность защиты участников и*



обеспечения прозрачности в исследовательской практике. Следуя этическим принципам, корпусная лингвистика может давать надежные и достоверные результаты, уважая при этом права личности.

**Ключевые слова:** Корпусная лингвистика; этические соображения; информированное согласие; конфиденциальность и анонимность; авторское право; защита данных

**Annotatsiya:** Ushbu maqola korpus qurilishi va tilshunoslik tadqiqotlarida foydalanish bilan bog'liq axloqiy fikrlarni o'rganadi. Korporatsiya real hayotdagi aloqa ma'lumotlarini tobora ko'proq o'z ichiga olganligi sababli, xabardor qilingan rozilik, maxfiylik va anonimlik, mualliflik huquqi, adolatli vakillik va ma'lumotlardan mas'uliyatli foydalanish kabi axloqiy masalalar muhim ahamiyat kasb etadi. Tadqiqot ishtirokchilarni himoya qilish va tadqiqot amaliyotida shaffoflikni ta'minlash muhimligini ta'kidlaydi. Axloqiy me'yorlarga rioya qilgan holda, korpus lingvistikasi individual huquqlarni hurmat qilgan holda ishonchli va ishonchli natijalar berishi mumkin.

**Kalit so'zlar:** Korpus lingvistikasi; axloqiy mulohazalar; xabardor qilingan rozilik; maxfiylik va anonimlik; mualliflik huquqi; ma'lumotlarni himoya qilish.

Corpus linguistics is a field of linguistic study that analyzes language through large collections of authentic spoken or written texts known as corpora. Grammar, discourse, and sociolinguistics studies are supported by these corpora, which offer insightful information about actual language use. However, the creation and use of corpora involve significant ethical issues because they frequently contain data created by actual people. To preserve participants' rights and prevent any harm, researchers must make sure that linguistic data is acquired and evaluated properly. With a focus on informed permission, privacy and anonymity, copyright and ownership, fair and balanced representation, sensitive content, and responsible use



of corpus data, this article aims to explore the major ethical issues in corpus development and use.

A fundamental ethical prerequisite for corpus development is informed consent, especially when data is directly from people. Participants must be made fully aware of the corpus's objectives, the kind of linguistic data being gathered, the duration of the data's storage, and who will have access to it. Consent must be freely given, free from coercion or obligation. Additionally, participants must be made aware of their ability to withdraw their data at any point during the study. Written authorization forms are often needed for spoken corpora and interviews, however ethical approval and public data policies may be applicable for some large-scale digital corpora. Research credibility is damaged when informed permission is not obtained, and there may be ethical and legal repercussions. Another crucial ethical principle in corpus linguistics is privacy preservation. Names, localities, occupations, and personal experiences are examples of personal information found in linguistic data that may inadvertently disclose a participant's identity. Researchers must anonymize corpus data by eliminating or substituting identifiable features with symbols or codes in order to solve this problem. In oral and online corpora, where speakers may discuss sensitive or intimate subjects, anonymity is particularly crucial. Researchers, consumers, and the general public cannot identify participants thanks to ethical corpus design. This approach fosters trust between participants and researchers while shielding people from possible damage. Copyright and ownership issues arise when corpora include written texts, audio recordings, or digital content created by others. Many sources, such as newspapers, books, websites, and social media posts, are protected by copyright law. Researchers must obtain permission from copyright holders or rely on legally permitted uses, such as fair use for academic purposes. Proper citation and acknowledgment of data sources are also required. Respecting copyright is not only a legal obligation but also an ethical responsibility that reflects academic honesty and respect for intellectual



property. Copyright and ownership issues arise when corpora include written texts, audio recordings, or digital content created by others. Many sources, such as newspapers, books, websites, and social media posts, are protected by copyright law. Researchers must obtain permission from copyright holders or rely on legally permitted uses, such as fair use for academic purposes. Proper citation and acknowledgment of data sources are also required. Respecting copyright is not only a legal obligation but also an ethical responsibility that reflects academic honesty and respect for intellectual property. Ethical responsibility does not end with corpus construction. Researchers must analyze and interpret data honestly and transparently. Misrepresenting findings, manipulating data, or ignoring limitations violates ethical research standards. Corpus users should also respect licensing agreements and cite corpora correctly in academic work. Responsible use of corpus data strengthens the reliability of linguistic research and contributes to the development of the field.

In conclusion, the creation and application of corpora in language research depend heavily on ethical issues. Essential values that safeguard participants and uphold academic integrity include informed consent, privacy and anonymity, respect for copyright, fair representation, and appropriate management of sensitive and digital data. Linguistic research is kept reliable, courteous, and socially conscious thanks to ethical corpus procedures. As digital and real-world communication become more prevalent in corpora, researchers must constantly modify ethical standards to fit new situations. Increased ethical consciousness can also help with cross-cultural communication, policy-making, and language instruction. In order to ensure that corpus linguistics develops responsibly and inclusively, future research should investigate ethical criteria for new types of online and multilingual corpora.



### References:

1. Gumperz, J. J. (1982). Discourse strategies. Cambridge University Press.
2. Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics*, 18(7–8), 581–618.  
<https://doi.org/10.1515/ling.1980.18.7-8.581>
3. Myers-Scotton, C. (1993). Social motivations for codeswitching: Evidence from Africa. Oxford University Press.
4. Auer, P. (1998). Code-switching in conversation: Language, interaction and identity. Routledge.
5. Biber, D., Conrad, S., & Reppen, R. (1998). Corpus linguistics: Investigating language structure and use. Cambridge University Press.
6. McEnery, T., & Hardie, A. (2012). Corpus linguistics: Method, theory and practice. Cambridge University Press.
7. McEnery, T., Baker, P., & Hardie, A. (2006). A glossary of corpus linguistics. Edinburgh University Press.
8. Baker, P. (2010). Sociolinguistics and corpus linguistics. Edinburgh University Press.
9. Eckert, P. (2000). Linguistic variation as social practice. Blackwell.
10. Holmes, J., & Wilson, N. (2017). An introduction to sociolinguistics (5th ed.). Routledge.