

GENDER REPRESENTATION AND COLLOCATIONAL PATTERNS: A COMPARATIVE ANALYSIS OF BNC AND COCA

Rixsiboyeva Farangis,

*Toshkent Davlat Tibbiyot Universiteti, Til o'rgatish markazi ,
Toshkent shahri*

Abstract: This study investigates gender stereotypes in English by analyzing their linguistic representation in the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). Using corpus-based methods, the research identifies collocational patterns, adjectives, and verbs commonly associated with male and female subjects. The comparative analysis highlights similarities and differences between British and American English in portraying gender roles. Findings suggest that language both reflects and reinforces social stereotypes, with certain traits and actions more frequently linked to one gender than the other. This study contributes to corpus linguistics and gender studies by providing empirical evidence of how English constructs gendered meanings, offering insights for future research on language, culture, and social perception.

Keywords: Gender stereotypes, Corpus linguistics, BNC (British National Corpus), COCA (Corpus of Contemporary American English), Collocation analysis, Linguistic representation, English language, Cross-cultural comparison

Language is not merely a tool for communication; it also reflects and shapes social perceptions, including notions of gender. Gender stereotypes, preconceived ideas about the attributes, behaviors, and roles of men and women, are frequently embedded in everyday discourse, influencing how individuals are perceived and treated in society. In English, these stereotypes can be observed through the words and expressions used to describe men and women, revealing implicit biases that persist across cultures.

Corpus linguistics provides a powerful method for empirically investigating these patterns. By analyzing large, representative collections of authentic language data, researchers can identify collocational patterns, adjectives, verbs, and other linguistic features that signal gendered meanings. The British National Corpus (BNC) and the Corpus of Contemporary American English (COCA) are two of the most widely used corpora, offering extensive datasets that reflect contemporary British and American English usage. Comparing these corpora allows for cross-cultural insights into how gender stereotypes are constructed and maintained in different varieties of English.

This study aims to examine the linguistic representation of gender in BNC and COCA, focusing on how language reinforces social stereotypes. By identifying patterns of collocations and word usage associated with male and female subjects, the research contributes to a deeper understanding of the interaction between language, culture, and social perception.

1. How are male and female subjects linguistically represented in the BNC and COCA?
2. What collocational patterns, adjectives, and verbs are commonly associated with each gender?
3. Are there notable differences in gender representation between British and American English?
4. How do these linguistic patterns reflect or reinforce societal gender stereotypes?

This study employs a corpus-based approach to examine gender stereotypes in English, using two major corpora: the **British National Corpus (BNC)** and the **Corpus of Contemporary American English (COCA)**. The BNC offers a comprehensive representation of British English in written and spoken forms, while COCA provides a large-scale corpus of contemporary American English across various genres. These corpora were selected due to their size, representativeness, and relevance for cross-cultural comparison.¹²

The focus of the analysis is on gender-specific lexical items, including nouns (e.g., *man*, *woman*), pronouns, adjectives, and verbs frequently associated with male and female subjects. Collocational analysis is conducted to identify recurring word pairings that signal stereotypical representations. Frequency counts and concordance lines are examined to determine the contexts in which these words appear and their semantic associations.¹³

Data extraction and analysis are performed using **AntConc** and **Sketch Engine**, which enable efficient concordance generation, frequency statistics, and collocation identification. A mixed-methods approach is adopted: quantitative analysis provides statistical evidence of collocational trends, while qualitative interpretation explores the pragmatic and semantic implications of these patterns.¹⁴

Finally, a comparative perspective between BNC and COCA is applied to uncover potential cultural variations in gender representation. This methodology ensures a rigorous and comprehensive investigation of how English encodes and perpetuates gender stereotypes.¹⁵

¹² Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

¹³ G. Ergasheva, T. Alimov, and M. Mamurova, "Gender Concepts in the Phraseological World Picture of the English and Uzbek Languages," *Turkish Journal of Computer and Mathematics Education* 12, no. 6 (2021): 345–356.

¹⁴ Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.

¹⁵ McEnery, T., & Hardie, A. (2011). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

The analysis of BNC and COCA revealed consistent patterns in the linguistic representation of gender. Male nouns and pronouns were frequently associated with adjectives such as *strong*, *leader*, *active*, and *intelligent*, and verbs including *command*, *win*, and *decide*, reflecting traditional masculine traits.¹⁶ Female nouns and pronouns co-occurred more often with adjectives such as *beautiful*, *kind*, *emotional*, and verbs like *care*, *support*, and *nurture*, reinforcing socially constructed feminine roles.¹⁷

Quantitative frequency counts indicate that action verbs are more commonly linked to male subjects, while descriptive adjectives predominantly characterize female subjects. Concordance analysis further shows that the context surrounding male terms often emphasizes agency and independence, whereas contexts for female terms highlight appearance and relational attributes.¹⁸

Comparative analysis between BNC and COCA suggests subtle cross-cultural differences. BNC exhibits slightly more formal or neutral descriptions, whereas COCA demonstrates a higher frequency of colloquial expressions and occasionally more stereotypical language. Despite these differences, both corpora consistently reflect underlying gender stereotypes, demonstrating that language both mirrors and perpetuates societal norms regarding gender.¹⁹

These findings underscore the importance of corpus-based studies for uncovering implicit biases in language. By systematically examining collocations and contextual usage, researchers can empirically demonstrate how English encodes gendered social expectations, providing insights relevant to linguistics, sociolinguistics, and gender studies.²⁰

This study examined gender stereotypes in English using corpus-based analysis of the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). The findings indicate that male subjects are predominantly associated with traits related to strength, leadership, and agency, whereas female subjects are linked to attributes emphasizing appearance, kindness, and relational roles.²¹ Collocational patterns

¹⁶ G. I. Ergasheva, "Basic Social Categories: Natural Semantic Metalanguage (NSM) Approach," *Philology Matters*, no. 2 (2019): 57–65.

¹⁷ Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

¹⁸ McEnery, T., & Hardie, A. (2011). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press, 102-105.

¹⁹ G. Ergasheva, "Conceptual Gender Analysis of Gender-Marked Phraseological Units and Proverbs in the Uzbek Language," *Philology Matters* (2021), p. 48.

²⁰ Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell, 77-80.

²¹ [Gender concepts in the phraseological world picture of the English and Uzbek languages](#). G. Ergasheva, TA Alimov, MJ Mamurova. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12 (6), 345-356

and concordance lines consistently reflect these gendered representations across both corpora, demonstrating the pervasive nature of social stereotypes in language.

Comparative analysis revealed minor differences between British and American English. BNC tended to use slightly more formal and neutral expressions, while COCA included more colloquial and occasionally overtly stereotypical language. Despite these variations, the overall patterns confirm that English, as used in everyday communication, encodes gendered assumptions in a systematic way.²²

These results contribute to corpus linguistics, sociolinguistics, and gender studies by providing empirical evidence of how language perpetuates social norms. They highlight the importance of critical awareness in language use and the potential for corpus-based research to inform educational, media, and policy initiatives aimed at reducing gender bias. Future research could extend this analysis to other English varieties, specific media genres, or longitudinal studies to track changes in gender representation over time.

The findings of this study have several important implications for both linguistics research and practical applications. Firstly, understanding how gender stereotypes are encoded in English can inform educational practices, helping teachers and curriculum developers foster awareness of implicit biases in language. Secondly, media professionals and content creators can benefit from these insights to avoid reinforcing harmful gender stereotypes in texts, advertisements, and digital content.

From a research perspective, this study demonstrates the value of corpus-based methods in identifying subtle patterns of social bias. It highlights the necessity of combining quantitative collocation analysis with qualitative contextual interpretation to gain a comprehensive view of language use. Future research could expand to include specialized corpora such as social media, news outlets, or literary texts, or perform longitudinal studies to monitor changes in gender representation over time.

In sum, by systematically mapping the linguistic construction of gender, this research not only contributes to corpus linguistics and gender studies but also provides actionable recommendations for education, media, and policy, aimed at promoting a more equitable representation of genders in language.

References:

1. S. Hall, "Cultural Identity and Diaspora," in *Identity: Community, Culture, Difference*, ed. J. Rutherford (London: Lawrence & Wishart, 1996).
2. G. Ergasheva, T. Alimov, and M. Mamurova, "Gender Concepts in the Phraseological World Picture of the English and Uzbek Languages," *Turkish Journal of Computer and Mathematics Education* 12, no. 6 (2021): 345–356.

²² McEnery, T., & Hardie, A. (2011). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press, 120-122.

3. T. McEnery and A. Hardie, *Corpus Linguistics: Method, Theory and Practice* (Cambridge: Cambridge University Press, 2012).
4. G.I.Ergasheva, "Basic Social Categories: Natural Semantic Metalanguage (NSM) Approach," *Philology Matters*, no. 2 (2019): 57–65.
5. T. McEnery and A. Hardie, *Corpus Linguistics: Method, Theory and Practice* (2012).
6. G. Ergasheva, "Conceptual Gender Analysis of Gender-Marked Phraseological Units and Proverbs in the Uzbek Language," *Philology Matters* (2021), p. 48.
7. G. I. Ergasheva, "Basic Social Categories: Natural Semantic Metalanguage (NSM) Approach," *Philology Matters* (2019), p. 60.
8. Gender concepts in the phraseological world picture of the English and Uzbek languages. G Ergasheva, TA Alimov, MJ Mamurova. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12 (6), 345-356
9. A.Moyer, "Do Language Attitudes Determine Accent?" *Journal of Multilingual and Multicultural Development* (2007).
10. G. I. Ergasheva, "Language Potential Through Cognitive Semantics Lens: Frames and ICMs of Gender Terms," *Experimental Phonetics: Problems and Solutions*, Vol. 1, No. 1 (2024), p. 15.
11. G. Ergasheva. *Experimental phonetics: problems and solutions*. Vol. 1 No. 1 (2024)
12. J. Holmes, *An Introduction to Sociolinguistics*, 4th ed. (London: Routledge, 2013).
13. Ergasheva, G. I. (2019). Basic social categories: Natural semantic metalanguage (NSM) approach. *Philology Matters*, 2, 57–65.
14. Stefanowitsch, A. (2006). Corpus-based approaches to metaphor and metonymy. In A. Stefanowitsch & S. T. Gries (Eds.), *Corpus-based approaches to metaphor and metonymy* (pp. 1–16). Berlin: Mouton de Gruyter.
15. McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.
16. Davies, M. (2008). The corpus of contemporary American English (COCA): 560 million words, 1990–present. *International Journal of Corpus Linguistics*, 14(2), 159–190.
17. Aston, G., & Burnard, L. (1998). *The BNC handbook: Exploring the British National Corpus*. Edinburgh: Edinburgh University Press.
18. Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
19. Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. *Proceedings of the Workshop on Comparing Corpora*, 1–6.