



A COMPARATIVE ANALYSIS OF GENDER REPRESENTATION IN THE LEXICON OF UZBEK AND GERMAN

Gulkhumor Yusupova

Durdona Matsabirova

PhD candidate, Urgench State University
named after Abu Rayhon Beruni,
Student, Urgench State University named after
Abu Rayhon Beruni
Urgench, Uzbekistan,
agulkhumor.yusupova@mail.ru,
b.durdonamatsabirova@gmail.com
<https://doi.org/10.5281/zenodo.15673573>

Abstract:

This study compares gender representation in the lexicons of Uzbek and German, focusing on grammatical structure, lexical usage, and socio-cultural influences. Despite Uzbek's lack of grammatical gender and German's explicit gender marking, both languages exhibit gender bias in professional, social, and familial terms. The analysis also examines how Natural Language Processing (NLP) systems reflect and reinforce these biases. The findings highlight that gender bias is not solely a grammatical phenomenon but is shaped by broader ideological and cultural factors. The study advocates for inclusive language reforms and bias-aware AI development.

Keywords: Gender representation, lexical bias, Uzbek, German, grammatical gender, sociolinguistics, NLP, inclusive language, linguistic ideology.

Introduction. The lexicon of a language provides crucial insights into its speakers' cultural norms, historical development, and social structures. Gender representation in the lexicon reflects deep-rooted ideologies and societal expectations regarding gender roles. This study aims to compare how gender is represented in the Uzbek and German languages, considering their distinct linguistic systems, historical backgrounds, and cultural contexts. The significance of this analysis lies in understanding how grammatical and lexical structures shape perceptions of gender and contribute to gender biases in language use and technology.

Literature Review. Uzbek, a Turkic language, and German, an Indo-European language, differ significantly in their treatment of grammatical gender. German nouns are classified into masculine, feminine, or neuter genders, influencing pronoun usage and adjective agreement (Kramer, 2013; Piazzolla et al., 2023). This system often lacks a direct correlation with biological sex but profoundly impacts cognition and communication (Samuel et al., 2019). In contrast, Uzbek lacks grammatical gender and relies on semantic markers or context to denote gender (Ruzieva, 2020). Scholars such as Bobojonova (2021) and Ronneberger-Sibold (2007) have highlighted the typological distinctions and historical influences that shape these languages.

Gender representation in professional titles, social roles, and kinship terms also reveals societal attitudes. While German has traditionally used gendered forms for professions, recent shifts toward gender-neutral expressions reflect evolving attitudes (Djavadghazaryans, 2020). Uzbek faces the challenge of avoiding gender bias in vocabulary despite its non-gendered grammatical structure (Rajapova, 2021).

The literature also emphasizes the role of language technologies in perpetuating gender biases. Corpora and machine learning models often reflect the dominant Anglo-American worldview, embedding gender stereotypes (Bansal, 2022; Ferrer et al., 2021). This is evident in AI systems and chatbots that frequently reproduce gender norms, such as assigning feminine traits to service roles (Caldarini et al., 2022).

Methods. This study employs a comparative linguistic analysis supported by corpus-based methods and a review of secondary data.

Corpus analysis was conducted using German data extracted from the Deutsches Referenzkorpus, with a focus on gendered vocabulary across various genres and contexts. For Uzbek, lexical data were gathered from official government documents, educational materials, and widely accessible online texts.

Textual comparison centered on selected lexical fields such as professional terms, familial roles, and institutional titles, to examine how each language portrays gender.

Additionally, the study incorporated a technological dimension through a review of existing literature on Natural Language Processing (NLP) tools and machine learning models. This part of the analysis evaluated how gender bias is manifested in automated systems and computational language processing in both Uzbek and German.

Results and Analysis

The comparative analysis of gender representation in the lexicons of Uzbek and German has revealed several important linguistic and cultural insights. These findings highlight the complexity of how gender is encoded, perceived, and potentially reinforced through both language structures and broader societal conventions. Below is a detailed exploration of the key results, accompanied by illustrative examples and interpretation.

1. Grammatical Gender in German and Its Societal Implications

German maintains a grammatically rigid gender system, where every noun is classified as masculine (der), feminine (die), or neuter (das). This grammatical gender is not always aligned with the biological sex of the referent. However, in the domain of professional and social roles, the grammatical gender often reflects and reinforces traditional gender roles.

For example: der Lehrer (male teacher) vs. die Lehrerin (female teacher)
der Arzt (male doctor) vs. die Ärztin (female doctor)

Such pairs indicate that masculine forms are often seen as default, while feminine forms are marked derivatives. In mixed-gender groups, masculine plural is usually used (e.g., die Lehrer can refer to a group of male teachers or a mixed group), which reinforces a male-as-norm ideology.

This system not only structures language but also influences perception. Research has shown that German speakers are more likely to visualize a male figure when a masculine noun is used generically, which may unconsciously reinforce gender stereotypes in professional and academic contexts.

2. Ideological Gender Encoding in Uzbek

Unlike German, the Uzbek language does not possess grammatical gender. Nouns, pronouns, and adjectives are not morphologically marked for gender. However, gender representation exists ideologically and is conveyed through context, specific lexical choices, and culturally ingrained expressions.

For instance: O'qituvchi (teacher) is gender-neutral grammatically, but culturally may evoke a female image when referring to primary school teachers, and a male image for university professors.

Rahbar (leader) is commonly associated with male figures, reflecting societal expectations and traditional roles.

In family terminology, gender-specific terms exist, such as ona (mother) and ota (father), but broader social roles are often gender-assigned through usage, not form. Expressions like **Ayol bosh bo'lsa, uy buziladi** (If a woman is the head, the household will collapse) indicate the presence of deep-rooted gender ideologies reflected in phraseology and proverbs.

Thus, even in the absence of grammatical gender, lexical gender bias persists through semantic and cultural association.

3. Sociolinguistic Reforms and Inclusive Language

Both German and Uzbek societies have witnessed movements toward more inclusive and gender-neutral language. In German, this has led to the adoption of gender-inclusive forms such as: **Lehrkraft** (teaching staff, gender-neutral), the use of the gender asterisk: Lehrer*in, or slash forms like Student/-in.

Despite institutional support and activism, the use of such forms is not yet standardized. There are still debates on readability, practicality, and acceptance across different demographics.

In Uzbek, efforts are more nascent. Since the language does not enforce grammatical gender, the focus has been on discouraging stereotypical expressions and promoting neutral or inclusive alternatives. However, public discourse, media, and educational content continue to reflect traditional gender norms, making widespread reform challenging.

4. Gender Bias in NLP Systems

The study also found that Natural Language Processing (NLP) tools and AI-based systems demonstrate embedded gender biases that reflect existing linguistic norms.

In German:

Machine translation tools often default to masculine forms for professions when translating gender-neutral sentences from English. For example, translating **"They are a doctor"** into German might yield **Sie sind ein Arzt** instead of the more neutral **Ärztin** or **Arzt** (a non-standard attempt at neutralization). Voice assistants and chatbots often exhibit gender-coded behaviors, such as using female voices for service roles and male voices for authoritative roles.

In Uzbek:

Although the language lacks overt gender markers, NLP tools trained on biased corpora may still reproduce gender stereotypes. For example, a model may associate **rahbar** (leader) more frequently with male pronouns in usage prediction or sentence generation. The limited availability of high-quality, balanced Uzbek language corpora further amplifies this issue.

These findings underscore the fact that technology not only reflects societal biases but may also amplify them if not addressed intentionally.

Discussion The structural nature of German reinforces gender differentiation, influencing cognition and possibly perpetuating stereotypes through daily language use. Uzbek, despite its non-gendered grammar, reflects traditional gender roles through selective lexical usage. These findings suggest that grammatical structure alone does not determine the

presence or absence of gender bias—socio-cultural and historical factors play an equally critical role.

The presence of gender bias in AI tools and corpora demonstrates the urgent need for inclusive language practices in computational linguistics. Addressing these biases requires interdisciplinary collaboration and conscious design of language resources. The tendency for chatbots and virtual assistants to adopt gendered behaviors illustrates how technology can reinforce stereotypes if not actively countered.

Conclusion This comparative analysis highlights the different linguistic paths Uzbek and German take in encoding gender, revealing how language both reflects and shapes societal norms. While German must confront the rigidity of grammatical gender, Uzbek must navigate implicit lexical biases. Future research should focus on creating balanced corpora, promoting inclusive language reforms, and developing AI tools that challenge rather than replicate existing gender norms

References:

1. Kramer, R. (2013). Gender in Amharic: a morphosyntactic approach to natural and grammatical gender. *Language Sciences*, Volume No. 43. 102-115 pp. <https://doi.org/10.1016/j.langsci.2013.10.004>
2. Piazzolla, S., Savoldi, B. (2023). Good, but not always Fair: An Evaluation of Gender Bias for three commercial Machine Translation Systems. *Computation and Language*, 9 Jun 2023. 209-225 p. <https://doi.org/10.48550/arXiv.2306.05882>
3. Ruzieva, S. (2020). The terms related to the concept family in the english and uzbek language cultures. *International Journal of Advanced Research*, Volume No. 8. 912-915 p. <https://doi.org/10.21474/ijar01/10381>
4. Bobojonova, Kh. (2021). Reviews on the usage of specific borrowed units in the lexicon of the Uzbek language. *Psychology and Education Journal*, Volume No. 48. 499-500 p. <https://doi.org/10.17762/pae.v58i1.1721>
5. Djavadghazaryans, A. (2020) "Please Don't Gender Me!" Strategies for Inclusive Language Instruction in a Gender-Diverse Campus Community. *Springer Nature*. pp 269–287. https://doi.org/10.1007/978-3-030-34342-2_15