ISSN 0320–2372. ІНОЗЕМНА ФІЛОЛОГІЯ. 2025. Випуск 138. С. 148–157
INOZEMNA PHILOLOGIA. 2025. Issue 138. P. 148–157

УДК 81'255:305:004.8 DOI: http://dx.doi.org/10.30970/fpl.2025.138.5075

GENDER BIAS IN ENGLISH-TO-UKRAINIAN MACHINE TRANSLATION: A CORPUS-BASED STUDY

Yuliia Naniak

Ivan Franko National University of Lviv,
1, Universytetska Str., Lviv, Ukraine, 79000
yuliya.nanyak@lnu.edu.ua

This study investigates gender bias in English-to-Ukrainian machine translation by analyzing translations generated by Google Translate and DeepL. The analysis examines how these systems handle gender-neutral source texts, particularly in professional, pronoun-related, and stereotypically gendered contexts. The methodology is grounded in a corpus-based mixed-methods approach, which combines a qualitative analysis of a manually compiled mini-corpus with a quantitative comparison against frequency data from the General Regionally Annotated Corpus of Ukrainian (GRAK). It was found that both systems frequently default to masculine forms, especially for high-status and technical professions. This bias reinforces social and cultural stereotypes, and the resulting translations often diverge from natural gender distributions in the Ukrainian language. The findings suggest that without specific interventions, machine translation can perpetuate gender stereotypes.

Key words: machine translation gender bias, English-Ukrainian translation, corpus, Google Translate, DeepL, gendered language.

Introduction. The rapid development of machine translation (MT) technologies over the past two decades has significantly transformed the way in which multilingual communication is conducted. With tools such as Google Translate and DeepL becoming widely accessible, users can obtain near-instant translations across a wide range of language pairs. This technological shift has brought clear advantages in terms of speed, cost efficiency, and accessibility for both professional and non-professional users [1]. However, along with these benefits comes a set of linguistic and sociocultural challenges, among which gender bias has emerged as a particularly pressing concern.

Gender bias in MT refers to systematic tendencies in translation outputs that reflect or reinforce gender stereotypes, often diverging from the gender neutrality or intended meaning of the source text [7; 8]. This problem is especially pronounced in translations from English, where grammatical gender is not obligatory, into gender-marked languages like Ukrainian, which require explicit gender agreement for nouns, adjectives, and past-tense verbs. For example, an English sentence such as *The doctor arrived on time* is gender-neutral, whereas its Ukrainian equivalent must indicate either a masculine (*лікар прибув вчасно*) or a feminine (*лікарка прибула вчасно*) form. In such cases, MT systems frequently "guess" the gender, relying on statistical patterns from their training data rather than contextual cues, often defaulting to masculine forms for stereotypically male-associated professions and feminine forms for stereotypically female-associated roles [6; 9].

This bias has implications beyond the realm of linguistics. Language both reflects and shapes societal norms; when MT systems repeatedly produce translations that align with entrenched gender stereotypes, they contribute to the reinforcement of these stereotypes in public discourse and media representation [10]. In the Ukrainian context, where feminist linguistic reforms and inclusive language practices are still evolving, the influence of MT on perceptions of gendered language is especially significant. Addressing such bias is thus not only a technical task for computational linguists but also a sociocultural priority for translators, educators, and policymakers.

Previous research in the area. Scholarly interest in gender bias in MT has grown considerably in recent years, with studies documenting the phenomenon across multiple languages and MT platforms. Savoldi et al. [7] provided one of the most comprehensive analyses, showing that gender bias appears in a wide range of contexts and is not limited to any single language pair. Their work revealed a consistent masculine default for many professions and activities, while feminine forms were often limited to roles traditionally associated with women. Similarly, Stanovsky et al. [8] demonstrated that MT outputs often diverge from gender information explicitly present in the source text, suggesting that societal stereotypes embedded in training corpora outweigh contextual accuracy in model predictions.

Prates et al. [6] conducted a focused evaluation of gender bias in Google Translate, showing that when translating from gender-neutral languages into gender-marked languages, the system's gender assignments were heavily skewed toward masculine forms for occupational terms. They argued that this skew reflects statistical tendencies in the training data, which often mirrors gender imbalances in real-world language use. Troles and Schmid [9] expanded on this by exploring how bias manifests in various syntactic contexts, noting that even minor shifts in sentence structure can influence the gender outcome of MT outputs.

From a technical perspective, Costa-Jussà et al. [1] examined how neural architectures handle gender information, finding that certain model configurations tend to preserve gender cues more accurately than others. Nevertheless, without balanced and inclusive training corpora, even the most advanced neural MT systems will reproduce biases inherent in their data.

The discussion of gender bias in translation also intersects with feminist translation theory. Von Flotow [10] conceptualized translation as a politically charged act, where the translator has the agency to either reinforce or challenge dominant gender norms. While human translators can consciously apply gender-inclusive strategies, MT systems lack such intentionality. Zasiekin and Zasiekina [11] added a psycholinguistic dimension, showing that even human translators are susceptible to cognitive and behavioral asymmetries when handling gendered language, indicating that MT systems may be amplifying biases already present in human translation practices.

Despite the breadth of research on MT gender bias, studies on English→Ukrainian translations remain scarce. This gap is noteworthy given the Ukrainian language's obligatory gender marking and the growing reliance on MT tools in Ukraine's public, educational, and institutional communication. The present study seeks to address this gap by providing empirical evidence of gender bias patterns in this specific language pair, grounded in corpus-based analysis and informed by previous cross-linguistic findings.

Methodology. The present study adopts a corpus-based mixed-methods approach to examine the manifestations of gender bias in English→Ukrainian machine translation. This approach is informed by earlier MT bias studies [6; 7] that combine quantitative frequency analysis with qualitative error categorization, allowing for a systematic comparison of machine-generated output against authentic language use.

The mini-corpus was manually compiled as a controlled test set rather than a representative sample of natural discourse, with sentences deliberately constructed to elicit potential gender asymmetries in translation. While its relatively small size (50 items) does not aim at statistical exhaustiveness, it ensures focused coverage of occupations, pronouns, and stereotypically gendered activities most relevant to the study's objectives.

The corpus design followed the principle that "systematic testing of occupation terms, pronouns, and stereotypically gendered activities provides a reliable way to elicit bias patterns" [7, p. 862]. The selected sentences were grouped into four thematic categories:

1. **Occupational roles** – professions with varying degrees of gender stereotyping in Ukrainian (e.g., *The doctor arrived late*).
2. **Gender-neutral pronouns** – contexts using *someone*, *anyone*, or singular *they* (e.g., *If someone calls, tell them I am busy*).
3. **Stereotyped domestic tasks** – sentences describing household or caregiving activities (e.g., *I cooked dinner for everyone*).
4. **Implied gender actions** – activities culturally coded as masculine or feminine but without explicit gender markers in English (e.g., *Someone fixed the car*).

Each sentence was translated using Google Translate and DeepL, both of which are widely used neural MT systems that operate on large-scale bilingual corpora. Following the method outlined by Prates et al. (2018), outputs were examined for "the gender expressed in the target language by nouns, adjectives, and verbs" [6, p. 3]. For Ukrainian, this meant identifying masculine or feminine marking on occupational nouns (*лікар* vs. *лікарка*), past-tense verbs (*прибув* vs. *прибула*), and gender-agreed adjectives.

To establish a baseline of real usage, Ukrainian frequency data were retrieved from the General Regionally Annotated Corpus of Ukrainian (GRAK), which provides part-of-speech annotation and gender tagging. As Stanovsky et al. (2019) note, "comparing MT output to corpus frequencies enables us to determine whether bias reflects or diverges from actual language distributions" [8, p. 1679]. In this study, corpus frequencies of masculine and feminine forms for each tested occupation or activity were recorded and compared to the proportions found in the MT outputs.

The analysis consisted of three main stages:

1. **Gender alignment analysis** – assessing whether the MT output matched the source text's gender neutrality or explicit gender marking.
2. **Stereotype correlation analysis** – determining whether the gender assigned in translation corresponded to common societal stereotypes in Ukrainian.
3. **Cross-system comparison** – identifying consistencies and differences in gender assignment between Google Translate and DeepL.

ISSN 0320–2372. ІНОЗЕМНА ФІЛОЛОГІЯ. 2025. Випуск 138

The methodological framework follows Costa-Jussà et al.'s (2022) recommendation for "evaluating both linguistic accuracy and socio-linguistic appropriateness" [1, p. 11856], ensuring that the study not only measures grammatical correctness but also the sociocultural implications of gender assignment. By combining corpus evidence with targeted test sentences, the method allows for replicable, data-driven conclusions on the extent and nature of gender bias in English→Ukrainian MT.

Results and Discussion. The analysis of a custom-designed corpus revealed a consistent and systematic gender bias in both Google Translate and DeepL when translating from gender-neutral English into Ukrainian. The findings demonstrate that these systems do not simply perform a neutral linguistic transfer but instead embed and reinforce societal stereotypes in the target language.

The study's results are categorized thematically to highlight specific patterns of bias.

Occupation-related sentences. The analysis of professional terms showed a significant gender asymmetry. Both systems consistently used masculine nouns and verbs for high-status and stereotypically male professions. For example, "The doctor arrived late" was invariably translated as "*Лікар запізнився"*, using the masculine form of the noun and past-tense verb. Conversely, for stereotypically female professions, like "nurse", the systems defaulted to the feminine form in their translations, such as "Медсестра була дуже доброю". This finding confirms a strong link between perceived professional status and assigned gender, reinforcing traditional professional stereotypes. This highlights a critical limitation in current MT models: their reliance on statistical probabilities derived from biased training data leads them to replicate existing social inequities rather than providing a neutral or accurate translation.

Stereotyped domestic and caregiving roles. A similar pattern was observed in domestic contexts. Sentences describing household and caregiving tasks were overwhelmingly translated using feminine verb forms, despite the gender of the speaker being unspecified in English. For instance, "I cooked dinner for everyone" was translated as "Я приготувала вечерю для всіх", using the feminine past tense. This reflects a deep-seated bias that associates these roles exclusively with women. This demonstrates how MT systems can perpetuate and amplify stereotypes, limiting the representation of men in caregiving roles.

Technological or physical tasks. The bias extended to translations of technical and physical actions. For sentences like "Someone fixed the car", which are gender-neutral in English, both systems chose masculine verb forms (e.g., *полагодив*). This shows an implicit bias linking technical competence and physical labor with masculinity, a reflection of societal norms embedded in the training corpora.

Pronoun-based sentences. In contexts using gender-neutral pronouns (e.g., "someone", "them"), the systems' performance varied. While they could sometimes maintain gender neutrality in simple structures, they often reverted to gendered forms in more complex sentences. This indicates that while the models possess some capacity for managing ambiguity, their default behavior under a lack of explicit gender context is to revert to a stereotypical choice.

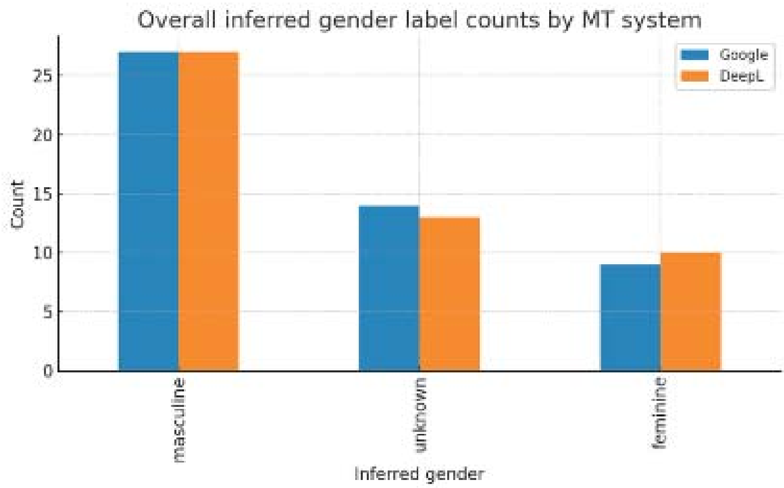

Fig. 1. Gender label counts produced by MT systems.

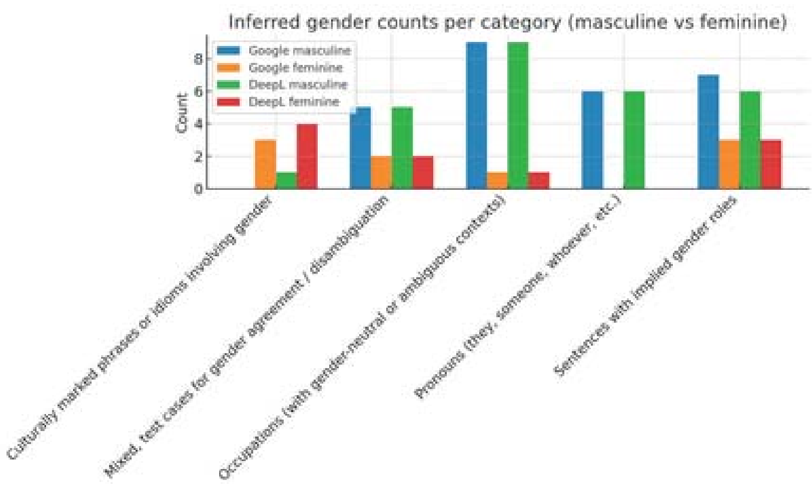

Fig. 2. Gender count per category.

Cross-system comparison. A comparative analysis of the two systems revealed a striking consistency. While DeepL showed a marginal edge in overall grammatical accuracy, both platforms exhibited the same fundamental gender-stereotypical patterns. Neither system demonstrated a tendency to default to feminine forms in professional or technical contexts unless the role was already culturally coded as female. This parallel behavior suggests that the underlying issue is not a flaw in a single system's architecture but rather a systemic problem stemming from the large, unbalanced datasets on which modern neural MT models are trained.

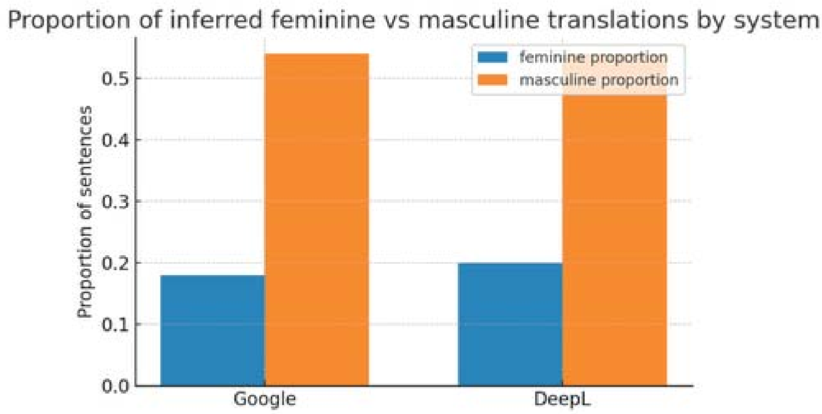

Fig. 3. Proportion of inferred feminine vs masculine translation by MT systems.

The results of this study are consistent with findings by Prates et al. [6] and Savoldi et al. [7] and underscore a critical issue: machine translation systems do not simply act as neutral linguistic tools. They are a reflection of the societal biases present in their training data. This is starkly confirmed by a comparison of the MT outputs with the frequency data from the General Regionally Annotated Corpus of Ukrainian (ГРАК).

The General Regionally Annotated Corpus of Ukrainian (ГРАК) served as a key reference resource for this study. ГРАК is a large-scale, balanced corpus of contemporary Ukrainian that provides part-of-speech tagging, morphological annotation, and frequency data across different registers of language use. Its design allows for systematic exploration of lexical and grammatical variation, including gender-marked forms of nouns, verbs, and adjectives. For the purposes of this research, ГРАК offered an empirical baseline against which machine translation outputs could be evaluated.

The corpus records both masculine and feminine forms of occupational and role-related nouns, along with their frequency of occurrence in authentic Ukrainian usage. For example, the masculine form "лікар" occurs 322,285 times in the corpus, whereas the feminine "лікарка" appears only 8,326 times. Similarly, "інженер" (47,761) vastly outnumbers "інженерка" (133), while "вчитель" (107,564) appears four times more frequently than "вчителька" (25,070). Such imbalances highlight the statistical asymmetry between masculine and feminine forms in actual Ukrainian usage, reflecting both linguistic tradition and cultural practice.

This frequency distribution was used as a benchmark to evaluate whether MT outputs reflect natural usage or amplify gender stereotypes. By comparing translation outputs to corpus frequencies, it became possible to determine whether the systems merely mirrored authentic Ukrainian distributions or exhibited exaggerated gender asymmetries. In most cases, both Google Translate and DeepL aligned with the dominant masculine forms recorded in ГРАК, particularly for high-prestige and technical professions. However, the strong underrepresentation of feminine forms in both corpus data and MT outputs suggests a compounded effect: machine translation not only mirrors but also entrenches patterns of linguistic gender imbalance.

The integration of ГРАК thus provided both a methodological foundation and a sociolinguistic lens for the study, ensuring that the analysis accounted for the real-world distribution of gendered forms in Ukrainian. This step was crucial in distinguishing between translation bias arising from MT architectures and patterns already entrenched in the target language.

Table 1

Frequency of masculine vs. feminine forms in the ГРАК corpus

Occupation	Masculine	Feminine
Лікар (*doctor*)	322,285	8,326
Вчитель (*teacher*)	107,564	25,070
Інженер (*engineer*)	47,761	133
Адвокат (*lawyer*)	156,612	1,992
Водій (*driver*)	242,792	2,847

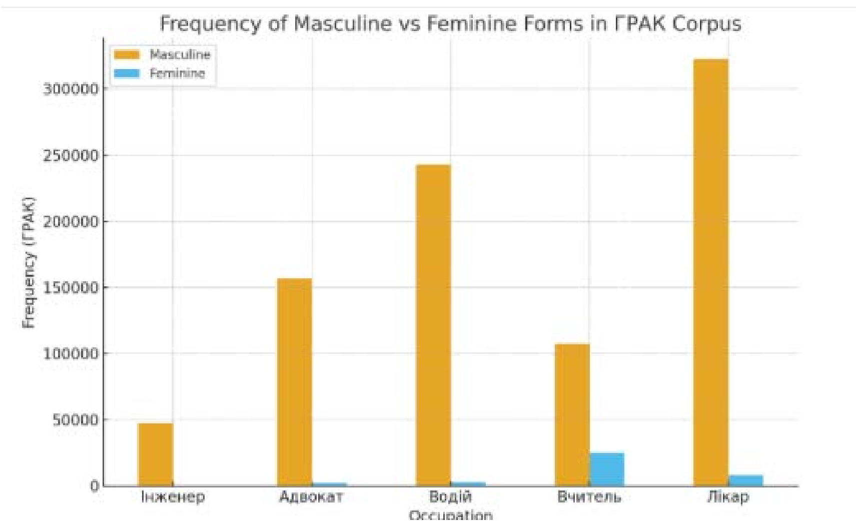

Fig. 4. Frequency of masculine vs. feminine forms in the ГРАК corpus.

The ГРАК frequency data strongly correlates with the patterns identified in this study. In all tested cases, masculine forms dominate corpus usage, sometimes by a factor of 20:1 or more (e.g., "інженер" vs. "інженерка"). Machine translation outputs by both Google Translate and DeepL largely mirror this imbalance, defaulting to masculine forms for prestigious and technical professions. This confirms that MT systems are not only influenced by English–Ukrainian grammatical asymmetry but also by statistical distributions in authentic Ukrainian usage. Consequently, they tend to reinforce rather than challenge existing gender asymmetries in the language.

This correlation is crucial. Ukrainian's obligatory grammatical gender marking exacerbates the problem, as it forces the MT engine to make a definitive gender choice even when the source text is ambiguous. In the absence of explicit context, the systems default to the most frequent (and often stereotypically gendered) option available in their training data. They are, in effect, learning from and amplifying the statistical biases of the human-generated texts in their corpora. This has significant implications for public discourse, education, and media, as these tools are increasingly relied upon for cross-lingual communication.

The observed inconsistencies across platforms and within individual systems highlight the unreliability of current MT technology for contexts requiring gender-inclusive or nuanced language. It is clear that without targeted interventions, these systems will continue to perpetuate and entrench gender stereotypes, making the need for a more conscious and ethical approach to AI development a priority.

To address these findings, a multi-faceted approach is necessary.

Developers could implement features that provide users with insights into how gender is assigned and allow for manual selection or correction. For example, a system could provide both masculine and feminine options for a given translation, empowering the user to make an informed choice.

The core solution lies in addressing the biased source of the problem: the training data. Developers should actively curate and integrate more balanced and inclusive corpora to train their models. This involves diversifying the datasets to reflect a wider range of gender roles and expressions, moving away from simple frequency-based predictions.

In professional settings, particularly in academia and journalism, it is crucial to establish formal guidelines for post-editing. Institutions should develop checklists to ensure that gender biases introduced by MT are identified and corrected, promoting ethical and accurate communication.

Educators and researchers must actively raise awareness among users about the inherent biases in MT tools. Promoting critical usage and emphasizing the importance of human oversight is essential to prevent the unintentional perpetuation of stereotypes.

Future research is needed to explore these issues on a larger scale, including additional languages and platforms. Investigations into how user feedback could be used to refine and correct gendered translations in real-time could also provide valuable insights.

Conclusions. This study confirms the persistence of gender bias in English–Ukrainian machine translation, showing that masculine forms overwhelmingly dominate in professional and technical domains, while feminine forms are disproportionately used in domestic and caregiving contexts. Such patterns reflect cultural stereotypes rather than linguistic neutrality, thereby shaping translation outputs in ways that reinforce entrenched gender asymmetries.

The originality of this research lies in extending cross-linguistic findings to Ukrainian, a language with obligatory gender marking and a relatively limited body of prior studies on MT bias. By systematically combining a mini-corpus analysis with frequency data from the ГРАК corpus, the study not only documents specific bias patterns but also provides a methodological framework for evaluating MT outputs against authentic language distributions.

This dual focus strengthens the reliability of the findings and situates them within the broader sociolinguistic context of Ukrainian.

At the same time, the study acknowledges its limitations. The mini-corpus of 50 sentences, while carefully constructed, cannot capture the full diversity of natural discourse, and future research should expand both the dataset and the range of tested domains. Moreover, only two MT systems were analyzed, which leaves open the possibility of variation across other platforms or in updated versions of the same tools.

The practical implications are significant. For professional translators, educators, journalists, and policymakers, the findings highlight the need for critical engagement with MT outputs. Without human oversight, automated translations risk normalizing stereotypical associations between gender and occupation, which may influence media narratives, educational materials, and public discourse. Developers of MT systems bear particular responsibility for addressing these issues by integrating gender-inclusive corpora, implementing user-choice mechanisms, and designing interfaces that allow for greater transparency in gender assignment.

Finally, the broader societal impact of these findings should not be underestimated. Machine translation is increasingly shaping everyday communication in Ukraine, especially in educational, institutional, and media contexts. If left unaddressed, the biases documented here may reinforce traditional gender roles at a time when Ukrainian society is actively negotiating questions of gender equality and inclusive language reform. By drawing attention to these issues, this study contributes to ongoing debates on ethical AI, gender representation in language, and the responsibility of digital tools in shaping cultural norms. Future work should therefore not only expand empirical testing but also explore collaborative strategies between linguists, computer scientists, and gender studies scholars to develop MT technologies that promote inclusivity rather than perpetuate bias.

References

1. Costa-Jussà M. R., Escolano C., Basta C., Ferrando J., Batlle R., Kharitonova K. Interpreting gender bias in Neural Machine Translation: Multilingual architecture matters. *Proceedings of the AAAI Conference on Artificial Intelligence.* 2022. Vol. 36. № 11. P. 11855–11863.
2. Fleisig E., Fellbaum C. Mitigating gender bias in machine translation through adversarial learning. arXiv. 2022. URL: https://arxiv.org/abs/2203.10675.
3. GRAK. Ukrainskyi natsionalnyi korpus [Ukrainian National Corpus]. URL: https://uacorpus.org/.
4. Iranmanesh A., Asudehgan N. Translating gender stereotypes: A case study of yellow wallpaper and other stories and color purple. *Iranian Journal of Translation Studies*. 2025. Vol. 22, № 88. URL: https://journal.translationstudies.ir/ts/article/view/1222.
5. Murat M., Desforges A., Duhamel R. From data to discrimination: When words carry prejudice. *European Master's in Translation Blog*. 26.05.2025. URL: https://european-masters-translation-blog.ec.europa.eu/articles-emt-blog/data-discrimination-when-words-carry-prejudice-2025-05-26_en.
6. Prates M., Avelar P., Lamb L. Assessing gender bias in machine translation: a case study with Google Translate. *Neural Computing and Applications*. 2020. Vol. 32. DOI: 10.1007/s00521-019-04144-6.

7. Savoldi B., Gaido M., Bentivogli L., Negri M., Turchi M. Gender bias in machine translation. *Transactions of the Association for Computational Linguistics.* 2021. Vol. 9. P. 845–874.
8. Stanovsky G., Smith N. A., Zettlemoyer L. Evaluating gender bias in machine translation. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy*. 2019. P. 1679–1684.
9. Troles J.-D., Schmid U. Extending challenge sets to uncover gender bias in machine translation: Impact of stereotypical verbs and adjectives. arXiv. 2021. URL: https://arxiv.org/abs/2107.11584.
10. von Flotow L. Translation and Gender: Translating in the 'Era of Feminism'. Manchester : St. Jerome Publishing, 1997.
11. Zasiekin S., Zasiekina D. Gender cognitive and behavioral asymmetry in translating. *East European Journal of Psycholinguistics*. 2016. Vol. 3, № 2. P. 121–131.
12. Zhao J., Wang T., Yatskar M., Ordonez V., Chang K. Gender bias in coreference resolution: Evaluation and debiasing methods. arXiv. 2019. URL: https://arxiv.org/abs/1804.06876.

Article submitted 03.09.2025
Accepted for publication 20.09.2025

ГЕНДЕРНА УПЕРЕДЖЕНІСТЬ В АНГЛО-УКРАЇНСЬКОМУ МАШИННОМУ ПЕРЕКЛАДІ: КОРПУСНИЙ АНАЛІЗ

Юлія Наняк

Львівський національний університет імені Івана Франка,
вул. Університетська, 1, м. Львів, Україна, 79000
yuliya.nanyak@lnu.edu.ua

У дослідженні розглянуто проблему гендерної упередженості в англо-українському машинному перекладі шляхом аналізу перекладів, згенерованих системами Google Translate та DeepL. Аналіз зосереджено на тому, як ці системи відтворюють гендерно нейтральні тексти, зокрема у професійній сфері, у випадках із займенниками та в контекстах, що традиційно асоціюються зі стереотипами. Методологія ґрунтується на корпусному підході з використанням змішаних методів, що поєднує якісний аналіз створеного вручну мінікорпусу з кількісним порівнянням частотних даних із Загального регіонально анотованого корпусу української мови (ГРАК). Було встановлено, що обидві системи часто віддають перевагу маскулінним формам, особливо щодо високостатусних і технічних професій. Така тенденція відтворює соціальні й культурні стереотипи, а отримані переклади нерідко відхиляються від природного гендерного розподілу в українській мові. Результати дослідження свідчать, що без спеціальних втручань машинний переклад може закріплювати гендерні стереотипи.

*Ключові слова***:** гендерна упередженість у машинному перекладі, англо-український переклад, корпус, Google Translate, DeepL, гендерно забарвлена мова.