ISSN 1812-0512 (Print) 2790-346X (online)

Wasit Journal for Human Sciences

Available online at: https://wjfh.uowasit.edu.iq

OPEN ACCESS

Gender Bias in Machine Translation: Analyzing Linguistic Choices in Automated Systems

May Mokarram Abdul Aziz

Dept. of English Language/College of basic Education University of Mosul/Iraq

ABSTRACT

The central purpose of this study is to analyze linguistic choices in automated systems, with a focus on gender bias in machine translation. The research applies Vinay and Darbelnet's model of translation strategies (1995) to identify variances between direct (literal) and oblique (adaptive) techniques in excerpts from Charles Dickens's Great Expectations translation into Arabic. The study examines how automated systems treat gendered language, emphasizing their leaning to sustain such biases, and deviate from the original text intent. Direct translation frequently results in gender errors and distorts the look of a character, especially for women. These findings highlight the necessity of advancing translation technologies that involve cultural comprehension, linguistic sensitivity, and moral responsibility to reduce disfigurement and improve the effectiveness of intercultural communication.

***Correspondence*:**
may.m.a@uomosul.edu.iq
Received: 26 August 2025
Accepted: 02 September 2025
Published: 01 November 2025

***DOI*:**
https://doi.org/10.31185/wjfh.Vol21.Iss4.1307

***Cite*:**
Abdul Aziz , . M. M. (n.d.). Gender Bias in Machine Translation: Analyzing Linguistic Choices in Automated Systems. Wasit Journal for Human Sciences, 21(4).
https://doi.org/10.31185/wjfh.Vol21.Iss4.1307

Keywords: Gender bias; Machine translation; Literary translation; Vinay and Darbelnet: direct and adaptive techniques

التحيز الجنسي في الترجمة الالية :تحليل الخيارات اللغوية في الانظمة الالية

ا.م. مي مكرم عبد العزيز

قسم اللغة الإنجليزية/كلية التربية الأساسية

جامعة الموصل/العراق

المُستخلص

الهدف الرئيسي من هذه الدراسة هو تحليل الخيارات اللغوية في الأنظمة الآلية، مع التركيز على التحيز الجنسي في الترجمة الآلية. يطبق البحث نموذج فيناي وداربلنت لاستراتيجيات الترجمة (1995) لتحديد الفروق بين التقنيات المباشرة (الحرفية) وغير المباشرة (التكيفية) في مقتطفات من ترجمة تشارلز ديكنز لرواية "آمال عظيمة" إلى العربية. تبحث الدراسة في كيفية تعامل الأنظمة الآلية مع اللغة المتحيزة جنسيًا، مؤكدةً ميلها إلى دعم هذه التحيزات، والانحراف عن مقصد النص الأصلي. غالبًا ما تؤدي الترجمة المباشرة إلى أخطاء في تحديد الجنس وتشويه مظهر الشخصية، وخاصةً عند النساء. تُسلّط هذه النتائج الضوء على ضرورة تطوير تقنيات الترجمة التي تُراعي الفهم الثقافي، والحساسية اللغوية، والمسؤولية الأخلاقية للحدّ من التشوهات وتحسين فعالية التواصل بين الثقافات.

الكلمات المفتاحية: التحيز الجنسي، الترجمة الآلية، الترجمة الأدبية؛ فيناي وداربلنت، التقنيات المباشرة والتكيّفية

1- Introduction

Machine Translation (MT) plays a role in global communication by providing fast and accurate translations to a large extent across languages. Nevertheless, its tackling of literary texts that reverse various cultural and contextual aspects is still limited. Literary texts present a crucial challenge for MT due to their reliance on distinctive linguistic pointers and cultural context; if not accurately rendered, translations may not provide the intended meanings.

The main issue in this context is gender bias, which manifests as stereotypical language and distortion of character depiction. This study investigates gender bias in MT translations in excerpts from Charles Dickens's Great Expectations, a novel full of gendered language. The main aim is to examine how MT systems treat gender language in literary texts and to assess whether they enhance or eliminate gender bias through linguistic choices and translation methods. By adopting Vinay and Darbelnet's (1995) model of direct and oblique translation strategies, this study reveals how automated systems handle gendered language in various contexts. The focus is not just on exploring literal and adaptive translation selections but also on investigating the influences of these selections on gender representation. The research assesses the cultural and gender-related details in the translated text and examines whether MT systems sustain such biases or stray from the original text's real meaning.

Recent local studies have also revealed the role of artificial intelligence in language-related education, such as the study published in the Wasit Journal for Humanities (2025) (Chaqmaqchee, O. & Naithel, O,2025) which mirrors the expanding academic focus in AI applications in linguistic and cultural contexts.

The study addresses the following questions:

1. How do MT systems administrate gendered language, for instance pronouns, descriptors, and metaphors, within Great Expectations?

2. Do the outputs from MT explain conventional gender standards, oppose them, or coordinate with updated cultural biases?

3. Which translation techniques—direct (literal) or oblique (adaptive)—are mainly applied by MT systems when handling gendered elements?

4. How do these translation decisions impact the representation of characters and the comprehension of the narrative in the target language?

The increasing use of machine translation tools in translation work has generated considerable interest among scholars, particularly in the gender bias. Stanovsky et al. (2019) created a multilingual criterion known as Wno MT to assess gender bias in MT systems. The study explains that even the best systems in Google, Amazon, and Microsoft fail to render gendered pronouns correctly, where clarification from the context is required. It showed restrictions in the accurate gender identification of current MT models.

Costa-jussà and de Jorge (2020) support an exhaustive analysis of gender bias in neural machine translation (NMT), investigating methods for its determination, and suggest strategies to mitigate it. They encouraged the incorporation of sociolinguistic elements in the training model to reduce gender bias while maintaining a high level of rendering.

Zhao et al. (2018) enhance the perspective by investigating bias in words used in MT systems. They argue that gender bias comes from its pre-trained language rules. The study indicates that correcting biases at the embedding level reduces them in the final translations.

These research findings affirm the consistency problem of gender bias in various MT platforms and highlight the significance of applying proper evaluation metrics and model enhancements. Furthermore, investigating gendered language structures, pronoun usage, and gendered language frameworks is necessary for understanding and tackling gender bias.

2- Literature Review

2.1. Machine Translation and Literary Texts

Neural machine translation (NMT) has widely enhanced machine translation using deep learning, making it more precise. However, the translation of literary works remains challenging due to their figurative language, references, and enduring cultural significance (Venuti, 2012, pp. 1-15). Bassnett (2014, pp. 12-30) and Munday (2016, pp. 145-168) affirm the literary translation requirement for a clear understanding of the style, passions, and culture of the source text, while retaining tone, rhythm, and symbolism. Moorkens (2020, pp. 191-205) affirms the ethical and practical defiance of using MT for literary works and refers to the co-creator role of the translator in this process. Recently, Castilho (2021, pp. 112–130) emphasizes that post-editing NMT output for literary texts needs as much effort as rendering from scratch for their difficulty of stylistic nuance. Likewise, Wu et al. (2022, pp. 55–72) point out that MT systems still tend to exhibit ambiguity, irony, and cultural allusions. Hu and Zhang (2023, pp. 201–218) argue that literary MT needs more than syntactic agreement, affirming the requirement for affection and cognitive mapping. Lee and Park (2024, pp. 34–50) illustrate through empirical testing that current NMT models, though useful in primary drafts, are unable to maintain tone and character advancements. Sennrich and Haddow (2020, pp. 77–94) observe that transformer-based models have improved fluency; however, they still encounter difficulties in handling stylistic conventions of literary genres and references. Koehn (2021, pp. 56–73) emphasizes that adapting MT systems for literature rendition depends on specialized corpora and stylistic

harmonization beyond standard MT approaches.

2.2. Gender Bias in Machine Translation

Gender bias in Machine Translation (MT) is not only a sensitive topic but also a linguistic and moral challenge in rendition studies. Addressing gender bias is necessary for improving machine translation accuracy.

Hovy and Spruit (2016, pp. 591–598) affirm that gender bias in MT systems mirrors long-rooted societal ideas that exist in data. They warn against the uncritical employment of biased corpora without care and call for assessing metrics and ethical design in NLP.

Stanovsky et al. (2019, pp.1-7) present empirical findings that frequently employed MT

Systems try to render gender-specific pronouns and functions precisely, especially for women or non-binary individuals. This has concrete results for reinforcing clarity and the formation of gender identities, particularly in vocational or public settings.

Costa-jussà & de Jorge (2019, pp.1-10)and Prates et al.(2020,pp 1-1) state that when rendering from languages with no marked gender (such as Turkish, Finnish) into languages that employ grammatical gender (such as English, German), MT systems often default to masculine terms or enhance outdated stereotypes. Moryossef, Goldberg, and Dagan (2019, pp. 354–358) introduce black-box context injection methods to reduce gender bias in occupational renderings, showing marked improvements in gender agreement without retraining the model. Saunders et al. (2020, pp.1-10) observe that while fine-tuning and post-editing can reduce bias, full improvement requires major changes that need to be made to the model's structure and data selection. Larson and Jojic (2020, pp. 1–9) also share this view, arguing that solutions must go beyond post-editing or token balancing.

Bender et al. (2021, pp.610-615) illustrate that MT training data originates from sources shaped by social aspects, causing translations to recreate the existing biases present in the original texts. Savoldi et al. (2021, pp.1-13) reveal that even advanced AI systems exhibit gender bias, which misrepresents female topics, especially in sentences that involve ambiguous pronouns or titles of position, where male renderings are dominant. Habash and Faraj (2022, pp. 767–793) state that Arabic–English NMT produces complex structures for gender, causing inaccurate gender agreement and masculine defaults in rendering job titles and rules. Garcia et al. (2022) argue that developing genuinely fair MT systems demands deliberate integration of sociolinguistic aspects rather than data-based systems, emphasizing that tackling gender bias requires interdisciplinary cooperation from feminist linguistics, sociolinguistics, and moral philosophy.

2.3. Translation Model of Vinay and Darbelnet

The translation model of Vinay and Darbelnet (1958/1995) distinguishes between two main categories: direct (literal) and oblique (free or adaptive). The direct approach involves calque, borrowing, and literal translation, which is proper when languages engage similar structures. By contrast, oblique translation, including equivalence, transposition, modulation, and adaptation, is employed when a direct rendering fails to maintain sense or style. This model is adopted to analyze how translators address cultural nuances, coherence, and rhetorical choices. Literary texts like " Great Expectations" employ oblique strategies for figurative and gendered subtleties. This study applies their structure to investigate how machine translation systems execute or ignore these strategies in translating gendered language. The model of Vinay and Darbelnet provides a structured approach for detecting both linguistic changes and cultural shifts between source and target texts (Munday, 2016,

pp. 88–92). The direct and oblique strategies of discrimination are beneficial for analyzing literary discourse where metaphor, idiomatic, and stylistic equivalence prevail (Baker & Saldanha, 2019, pp. 234–239). Modern scholarship also confirms that this model, although developed for French–English rendition, can be impact-adapted to various language pairs with salient structural and cultural differences through socio-pragmatic integration (Chesterman, 2016, pp. 51–55).

2.4. Gender Representation in Great Expectations

Dickens's Great Expectations offers fertile ground for studying gendered language and character evolution through figures like Pip, Estella, and Miss Havisham, who embody varied representations of masculinity and femininity. Michie (1992, pp. 101–125) and Flint (1995, pp. 45–67) examine Dickens's interaction with gender expectations of Victorian culture, making the novel a prime case study for evaluating machine translation of gendered discourse within Victorian cultural parameters. Estella's coldness and emotional control, for instance, conform to societal norms for femininity (Michie, 1992, pp. 101–125), while Pip's precarious masculinity, governed by guilt and ambition, shows the interplay between class hierarchy and gendered selfhood (Flint, 1995, pp. 45–67). Translating these elements requires sensitivity to subtle, context-bound features such as irony and narrative voice. Van Rijsselbergen and Lefever (2021, pp. 215–233) say that machine translation tools may continue gender stereotypes by overlooking contextual complexities in literary works.

2.5. Cultural Adaptation and Translation Ethics

Spivak (1993, pp. 179-181) states that ethical translation in postcolonial contexts must retain the rhetorical and cultural features of the source text. Venuti (1995, pp. 20-25) prioritizes maintaining foreignness over domestication, which strengthens cultural narratives. Tymoczko (2007, pp. 190-194) emphasizes cultural adaptation to ensure clarity while respecting the source context. Pym (2010, pp. 7-10) and Baker & Maier (2011, pp. 1-3) highlight that translation ethics focuses on assuring cultural identity, reducing bias, and reinforcing precise and just representation. García and Pena (2011, p. 45) emphasize that without careful supervision, MT may perpetuate linguistic biases. Angelelli (2014, pp. 63-65) and Chesterman (2017, pp. 120-124) discuss translators' ethical obligation to balance source fidelity with audience comprehension. Moorkens et al. (2018, pp. 222-225), Bielsa & Bassnett (2020, pp. 54-57), and Wu & Zhang (2020, pp. 15-18) stress the need for cultural awareness, reflexivity, and precise adaptation to ensure ethical translation across contexts. Chang, Prabhakaran, and Ordonez (2019, n.p.) situate MT bias within the broader context of fairness in NLP, emphasizing that meaningful correction requires both technical and socio-cultural interventions. Ciora, Iren, and Alikhani (2021, 23 pp.) demonstrate how hidden gender bias appears in Turkish–English MT models, showing reliance on stereotypical assumptions. Bender et al. (2021, pp. 610-615) emphasize the importance of moral factors in MT when processing sensitive material. Bansal (2022, ~30 pp.) illustrates the origins of bias in NLP and strategies for reduction relevant to MT.

Moniz and Parra Escartín (Eds., 2023, n.p.) propose ethical and legal structures for responsible MT development, complementing prior discussions on morality and fairness.

3. Methodology

This research employed a qualitative design concentrated on a descriptive and comparative approach to analyze texts. It aims to explore how gender bias appears in machine translation (MT) systems of literary works, such as Charles Dickens's "Great Expectations." The analysis adopted Vinay and Darbelnet's (1995) framework for translation procedures, which differentiates between direct (literal) and oblique (adaptive) methods.

3.1 Corpus Selection

The source item includes selected gender-relevant excerpts from "Great Expectations," involving passages that contain descriptions of gendered pronouns, metaphors, and character portrayals.

3.2 Translation Platforms

Two widely used machine translation services—Google Translate and Microsoft Translator were employed to produce Arabic translations of the selected excerpts. Arabic is a language that marks gender and has unique morphological characteristics in verbs, adjectives, and pronouns, supporting a distinct viewpoint for examining the representation of gender in translation.

3.3Analytical Framework

Vinay and Darbelnet's model (1995) presents a framework for analyzing translations, investigating whether MT systems employ direct or oblique strategies, concentrating on gendered language. Instances of mistranslation, gender bias, or character misrepresentation were identified and categorized.

3.4 Data Analysis Procedure

The analysis compared the source text with each MT output. The main aspects investigated included:

- Gender aligned in pronouns and verbs.

-Lexical choices that signify gender stereotypes or neutralization attempts.

4. Data Analysis

This study analyzes gender bias in machine-translated excerpts from Dickens's Great Expectations using Vinay and Darbelnet's model. Several gender-related excerpts were chosen from the source text and translated into Arabic using two widely used machine translation platforms: Google Translate and Microsoft Translator, to examine how each handled gendered language. The analysis revealed cases such as gendered pronouns, figurative expressions, and character representation reflect an overreliance on direct translation. None of the systems consistently applied oblique strategies, resulting in biased renderings of gendered meaning.

5. Discussion

This study examined how gender bias appears in Arabic translations of "Great Expectations" using machine translation (MT) platforms (Google Translate and Microsoft Translator). The outcomes were analyzed via Vinay and Darbelnet's model, which consists of two basic techniques :

-Direct Translation: It is a literal translation that strictly follows the original text's structure and words, repeating its shape as much as possible.

-Oblique Translation: It is a flexible technique using adaptable approaches like

adaptation, modulation, and equivalence when direct translation deforms the meaning.

1-Overuse of Direct Translation

Machine translation tools frequently rely on literal translations, which result in grammatical errors and gender inconsistency.

Example:

Original: “She was proud and cold.” Google Translate (Arabic): "كان فخورًا وباردًا"

Analysis:

Although the original sentence indicates a feminine subject(Estella), Google Translate fails to reflect this. The translation employs the masculine verb “كان” for “was”; the translation mistakenly uses the masculine words “فخورًا” (proud) and “باردا” (cold). This mistake indicates reliance on direct translation without adopting modulation to fulfill proper gender harmony in Arabic. Such mistakes can misdirect readers about the subject’s gender, particularly in literary contexts where character depiction is integral to sense. It also explains that machine translation systems frequently ignore contextual signals from surrounding text, depending on virtual grammatical structures. Microsoft Translator presents some contradictions in its outputs, but it maintains gender consistency and adapts idiomatic expressions impressively. While it sometimes employed alternative strategies, these were not used consistently.

2-Misuse of Lexical Choices (Cultural Bias)

Some machine-translated outputs detect culturally biased additions or mirror the absence of some lexical elements from the source text.

Example:

Original: “I have no heart.” (said by Estella)

Arabic MT Output: "ليس لدي قلب كامرأة"

Analysis:

In this machine translation, the addition "كامرأة" does not exist in the source text; however, it appears in the Arabic translation. This addition presents a cultural stereotype because of biased training data and mirrors a misuse of calque, a technique of direct borrowing that results in unintended gender bias. Such additions alter the reader's mental image of the character, leading to societal presumptions that did not exist in the original narrative. It also shows that the MT system may lessen meaning from common patterns in its corpus instead of sticking strictly to the semantic content of the original.

A more suitable strategy would be adaptation, an oblique translation strategy that preserves the original sense while preventing culturally loaded explanations. This matter ensures the author's neutrality while maintaining cultural sensitivity in the target language. Specifically, inserting (امراة)emphasizes a presumption that the heart is related to femininity, which distorts the intention of the author. While Estella affirms her lack of affective capacity in absolute terms, the MT system reinterprets it as a gendered statement, thereby distorting the character and the narrative.

3-Failure in Translating Metaphorical and Figurative Language.

Machine translation systems frequently face challenges in representing metaphorical and figurative language.

Example :

Original: “A wax figure who had stopped melting.”

Arabic MT Output: "تمثال من الشمع توقف عن الذوبان"

Analysis:

Although the rendering is literal, it fails to capture the metaphorical symbol of the original significance and emotional depth. The portrayal of Miss Havisham as a frozen, ghostly figure loses its effect. It indicates a missed chance to apply equivalence or adaptation indirect strategies that could have presented the metaphor in a way that resonates culturally and emotionally. The inability to achieve this results in a diminished narrative tone and character representation.

4-Softening of Female Authority

Sometimes, machine translation systems tend to lessen the assertive tone of female characters, unintentionally mirroring gendered politeness standards that change the source text’s dynamic force.

Example:

Original: “Well? You can break his heart.” (spoken by Miss Havisham)

Arabic MT Output: "حسنًا؟ يمكنك أن تكسري قلبه، إذا أردتِ"

Analysis:

In the original English sentence, Miss Havisham’s utterance is short, direct, and carries a commanding tone. It presents her as a woman in control of Estella’s behavior. However, the Arabic machine translation produces an optional, polite touch by adding “إذا أردتِ” , which does not exist in the source text. This addition lessens the authoritative tone and turns the intention from a manipulative directive to a polite proposition. Such turns may show gendered biases in MT systems, where female speech is automatically explained with culturally formed presumptions about how women should speak politely and indirectly. This rendering is a misuse of modulation, where the machine changes the tone without contextual reasoning. A more appropriate Arabic translation would be: "حسنًا؟ يمكنك أن تكسري قلبه" keeping the brevity and the manipulative power of the original.

5. Misuse of Direct (Literal) Translation Causing Gender Errors

Machine translation tools frequently ignore gender concord in Arabic because they rely too much on a literal translation.

Example :

Original: “She replied bitterly.”

Google Translate (Arabic): "رد بمرارة"

Analysis:

In the Google translation, the verb "رد" is in the masculine form, which wrongly represents the subject "she." In the original utterance. The correct Arabic translation should be "ردت بمرارة" to ensure proper gender agreement. This error arises from employing literal translation without making the necessary

modifications to verb forms that rely on gender in Arabic. This matter affirms a serious problem with MT systems ignoring critical grammatical rules in gendered languages, resulting to distorted and biased representations.

6. Gender Stereotyping in Lexical Choice

Subtle gender bias in machine translation systems is sometimes exhibited in the selection of culturally laden adjectives, even when the grammar is correct.

Example:

Original: “Estella is hard and proud.” Microsoft Translator (Arabic): "إستيلا قاسية ومغرورة"

Analysis:

Although the translation is grammatically accurate, the adjectives "قاسية" and "مغرورة" have culturally negative connotations in Arabic when employed to describe women. Dickens' original prescription appears to be the self-confidence of Estella, instead of a flawed ethical one. A more appropriate translation would have employed a neutral alternative like "متكبرة" for (proud), maintaining the original tone without upholding stereotypes. It expresses a lack of applying equivalence or adaptation, both indirect strategies that could have effectively conveyed the intended message will steering clear of culturally biased interpretations.

7. Inappropriate Softening of Female Authority

Machine translation applications frequently change the tone of confident female dialogue by incorporating politeness markers absent in the original text.

Example:

Original: Miss Havisham said, “Come back at once"

Arabic MT Output: "!من فضلك، عد فورًا"

Analysis:

The addition of “من فضلك” (please) introduces a polite tone that is not present in the original English sentence. The command of Miss Havisham is direct and urgent, yet the translation reframes it in a softer, more deferential manner. It reflects a misapplication of modulation, where the tone is adjusted based on stereotypical assumptions about how women speak. Instead of maintaining the assertiveness of character, the MT system modifies the tone in a manner that introduces gender bias through unnecessary politeness. Instead of keeping the character's assertiveness, the MT system modifies the tone in a way that presents gender bias through unnecessary politeness.

8. Loss of Figurative Meaning in Gendered Metaphors

Machine translation systems frequently have difficulty preserving figurative language, especially when it explains gendered or emotional variation.

Example :

Original: “She looked like a ghost, frozen in time.”

Google Translate (Arabic): "بدت كأنها شبح مجمد في الزمن"

Analysis:

Although the translation is grammatically correct, it lacks the emotional and symbolic depth of the original. The phrase “frozen in time” serves as a powerful metaphor for Miss Havisham’s emotional paralysis and enduring trauma. The literal rendering in Arabic feels flat and mechanical. A more culturally resonant adaptation, such as "كأنها شبح توقف به الزمن" (as if a ghost suspended in time), would better convey the intended meaning. This failure clarifies the neglect of equivalence by the MT system, a subtle strategy necessary for maintaining metaphorical richness, particularly in showing female psychological experiences and identities.

9. Gender Ambiguity in Neutral or Ambiguous Contexts

In certain situations where the subject's gender is unclear in English, machine translation services often default to using masculine forms in Arabic, resulting in gender distortion.

Example :

Original: “The child was quiet.”

Microsoft Translator (Arabic): "كان الطفل هادئًا"

Analysis:

The above translation is grammatically accurate, yet employing the masculine adjective"هادئًا." seems confused. If the context refers to the child being female (Estella), the translation should be "هادئة" to clarify correct gender alignment in Arabic. The reason behind this mistake is related to a literal translation strategy, which fails to show contextual alludes. A more accurate translation requires a modulation, an indirect method that modulates grammatical structures, relies on character context. The lack of this modulation results in gender mismatch, especially in texts where the identity of characters is important to the story.

10. Inconsistent Treatment of Female Characters

Machine translation tools can produce cultural misunderstandings that alter the personification of female characters occasionally, even when translations are grammatically accurate.

Example:

Original: “Estella smiled coldly.”

Google Translate (Arabic): "ابتسمت إستيلا ببرود"

Analysis:

The above translation is correct, yet the term "ببرود" (coldly) indicates culturally negative connotations in Arabic, especially when signifying a woman. It may come across as arrogance instead of restrained emotions. According to the context, this phrase refers to Estella’s emotional distance instead of rudeness. It explains how a direct translation can unintentionally provide cultural bias. Applying modulation or adaptation could have described emotional distance in a culturally neutral way, preserving the character’s intended picturing without maintaining stereotypes.

11. Emotional Attribution to Female Characters

Although translations look grammatically accurate, machine translation systems can skip delivering emotional tone and character diversity due to cultural differences in proper word selection.

Example:

Original: “Biddy was thoughtful and kind.” Google Translate (Arabic): "كانت بيـدي متفكرة ولطيفة"

Analysis:

The above rendering is grammatically accurate, yet in Arabic, the word "متفكرة" usually connotes deep, abstract reflection, often linked with formal or masculine situations. It turns the image of Biddy away from Dickens’s intention of passion, intelligence, and kindness. An appropriate translation is: "كانت بيـدي حكيمة وحنونة" seems to more precisely reflect the emotional depth of Biddy's character. This mistake arises from a focus on literal translation rather than modulation or equivalence to convey the emotional and gendered subtleties of character descriptions in Arabic. The inability of MT tools to recognize the gendered implications of emotional vocabulary results in deformations of female character portrayals.

12. Weakening of Female Critique

Machine translation tools sometimes lessen the emotional power and critical tone of female characters, particularly when they express judgment, anger, or authority.

Example:

Original: “I am what you have made me.” (spoken by Miss Havisham to Pip)

Microsoft Translator (Arabic): "أنا ما جعلتني عليه"

Analysis:

In the original, Miss Havisham made an explicit and emotional accusation against Pip and held him responsible for what she had become (or, symbolically, on community or other characters). The English structure affirms blame and responsibility: “you have made me,” referring to deliberate shaping by another, describing bitterness.

The Arabic translation “أنا ما جعلتني عليه” is structurally inaccurate and confused. It reduces the sounds of accusative and makes the sentence more submissive. This softening influence weakens Miss Havisham’s critical tone, lessening her emotional authority. The translation mirrors a possible gendered bias in MT systems by neutralizing female expressions of blaming.

A proper proposed Arabic rendering would be:

"أنا ما صنعتَه بي" or "أنا ما جعلتني عليه أنت", both maintain the accusatory tone and direct blame in a way that preserves Miss Havisham’s emotional force and critique in agreement with the original narrative.

6. Findings

The study analyzed 12 machine-translated sentences from “Great Expectations” (translated into

Arabic) to investigate gender bias, using Vinay and Darbelnet's model of translation strategies. Each example was classified based on whether it used direct translation (e.g., literal or calque) or oblique translation (e.g., modulation, equivalence, or adaptation) and whether the strategy led to gender bias in the output. They analyzed using two machine translation (MT) platforms (Google Translate and Microsoft Translator), and the following patterns emerged:

1. Overuse of Literal Translation (Direct Strategy): 66.7%, 8 out of 12 cases relied on literal or direct translation. In these cases, MT tools failed to adjust gender agreement or tone.

Examples include:

"She was proud and cold." → incorrect masculine gender

"She replied bitterly." → wrong verb conjugation

"Biddy was thoughtful and kind." → mistranslation of emotion

2. Gender Agreement Errors in Arabic: 41.7%

5 out of 12 examples had incorrect gender forms due to literal translation.

These errors show a lack of syntactic adaptation to Arabic's gendered grammar.

Examples:

"The child was quiet." → rendered in masculine form regardless of context

"She was proud…" → masculine adjective used

3. Tone Modulation Leading to Bias: 33.3%

4 out of 12 examples involved inappropriate tone softening in female speech.

MT tools frequently add politeness markers or soften the assertive language.

Examples:

"Come back at once!" → translated with "من فضلك" (please)

"You can break his heart." → rendered as "hurt" rather than "break"

"I am what you have made me."—rendered as "انا ماجعلتني عليه" which reduces the blaming tone and shifts Miss Havisham's critical voice into a more submissive one.

4. Biased Lexical Choice (Cultural Stereotyping): 25%

3 examples introduced culturally biased or stereotypical words for female characters.

Examples:

"Estella is hard and proud." → "مغرورة" has moral judgment

"I have no heart." → added "كامرأة" (as a woman), not in the source

5. Misuse of Figurative Language Affecting Characterization: 33.3%

Four examples lost the emotional/metaphorical depth of the original text due to literal translation.

For example:

"Frozen in time" metaphor

"Wax figure" → literal but emotionless Arabic output.

7. Conclusion

This study has explored the presence of gender bias in machine translation (MT) systems through a qualitative analysis of selected excerpts from Charles Dickens's " Great Expectations", translated into Arabic using leading MT tools. Applying Vinay and Darbelnet's translation model, the research identified a consistent overreliance on literal (direct) translation strategies that fail to account for the grammatical and cultural complexities of gender representation in Arabic. The analysis revealed that literal translation often leads to gender agreement errors, distortion of character portrayal, and subtle reinforcement of gender stereotypes—especially when translating female characters. Furthermore, modulation and adaptation were either misapplied or avoided, resulting in softened tones, altered emotional content, and the introduction of culturally biased lexical choices.

Findings showed that 66.7% of the examples relied on literal translation, while over 40% included gender agreement mistakes. Such patterns highlight a systemic shortcoming in current MT systems: the lack of gender-sensitive and context-aware processing, particularly for languages with grammatically gendered structures like Arabic. These results stress the urgent need for integrating literary, cultural, and gender awareness into MT algorithms. Improvements should include better training data, context-based parsing, and targeted adaptation strategies to ensure accurate and equitable translation outcomes. This is especially vital in literary translation, where subtle shifts in tone and characterization can significantly alter meaning and reinforce or challenge societal norms.

References

Angelelli, C. V. (2014). The Sociological Turn in Translation and Interpreting Studies. Amsterdam: John Benjamins.

Baker, M. (2018). In other words: A Course Book on Translation (3rd ed.). London: Routledge.

Baker, M., & Maier, C. (2011). Ethics in Interpreter & Translator Training: Critical

perspectives. The Interpreter and Translator Trainer.

Baker, M., & Saldanha, G. (2019). Routledge Encyclopedia of Translation Studies (3rd ed.). London: Routledge.

Bansal, R. (2022). A survey on Bias and Fairness in Natural Language Processing. arXiv.

Bassnett, S. (2014). Translation Studies (4th ed.). London: Routledge.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can language Models be too big? Proceedings of FAccT '21.

Bielsa, E., & Bassnett, S. (2020). Translation in Global News (2nd ed.). London: Routledge.

Castilho, S. (2021). Challenges in Post-editing Literary Texts: Human Effort and Machine Creativity.

Chang, K.-W., Prabhakaran, V., & Ordonez, V. (2019). Bias and Fairness in Natural Language Processing. EMNLP-IJCNLP Tutorial Abstracts. Association for Computational Linguistics.

Chaqmaqchee, O. & Naithel, O. (2025). Artificial Intelligence in English Language Education in Iraq: A Review Study of Interventions and Perceptions. Wasit Journal for Humanities, 20(3), 1124–1147.doi https://doi.org/10.31185/wjfh.Vol21.Iss3.1080

Chesterman, A. (2016). Memes of Translation: The Spread of Ideas in Translation Theory (2nd ed.). Amsterdam: John Benjamins.

Ciora, C., Iren, N., & Alikhani, M. (2021). Examining Covert Gender Bias: A case Study in Turkish and English Machine Translation models.

Costa-jussà, M. R., & de Jorge, P. (2019). Mitigating Gender Bias in Machine Translation: Literature Review. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics.

Dickens, C. (1861). Great Expectations. London: Chapman and Hall.

Flint, K. (1995). The woman reader, 1837–1914. Oxford: Oxford University Press.

Garcia, A., Sun, M., & Tiedemann, J. (2022). Gender-Inclusive Machine Translation: A Feminist Computational Approach. Journal of Computational Linguistics, 48(1), 95–121.

Habash, N., & Faraj, R. (2022). Gender Bias and Morphological Richness in Arabic-English Neural Machine Translation. Natural Language Engineering.

Hovy, D., & Spruit, S. L. (2016). The Social Impact of Natural Language Processing. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers).

Hu, Y., & Zhang, L. (2023). Cognitive-Emotive Aspects of Literary Translation and Neural Machine Translation. Journal of Literary Translation Studies.

Koehn, P. (2021). Neural machine translation. Cambridge: Cambridge University Press.

Larson, S., & Jojic, O. (2020). Gender Bias in Natural Language Processing: A Survey. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts.

Lee, J., & Park, S. (2024). Evaluating Narrative Voice and Tone in NMT literary Outputs. Digital Translation Review.

Michie, E. (1992). Outside the Pale: Cultural Exclusion, Gender Difference, and the Victorian Woman Writer. Ithaca, NY: Cornell University Press.

Michie, E. (1992). Sororophobia: Differences Among Women in Literature and Culture. Oxford: Oxford University Press.

Moniz, H., & Parra Escartín, C. (Eds.). (2023). Towards Responsible Machine Translation: Ethical and Legal Considerations in Machine Translation. Springer.

Moryossef, A., Goldberg, Y., & Dagan, I. (2019). Filling Gender and Number Gaps in Neural Machine Translation Using Black-Box Context Injection. In Proceedings of the 57th Annual Meeting of the Association

for Computational Linguistics.

Moorkens, J. (2020). The Use of Machine Translation in Literary Texts: A Matter of Ethics.

Munday, J. (2016). Introducing Translation Studies: Theories and Applications (4th ed.). London: Routledge.

Prates, M. R., Avelar, F. C., & Lamb, L. C. (2020). Assessing Gender Bias in Machine Translation: A Case Study with Google Translate. In the NeurIPS 2018 Workshop on Ethics in NLP. Montréal: NeurIPS Conference Proceedings.

Pym, A. (2010). Exploring Translation Theories. London: Routledge.

Saunders, D., Ruder, S., & Søgaard, A. (2020). Reducing gender Bias in Neural Machine Translation as a Domain Adaptation problem.

Savoldi, B., Anastasopoulos, A., Belinkov, Y., & Tiedemann, J. (2021). Gender Bias in Machine Translation. In Findings of the Association for Computational Linguistics: EMNLP 2021. Punta Cana: Association for Computational Linguistics.

Sennrich, R., & Haddow, B. (2020). Low-Resource Machine Translation with Pre-Trained Language Models. Machine Translation.

Spivak, G. C. (1993). The Politics of Translation. In Outside in the Teaching Machine. London: Routledge.

Stanovsky, G., Smith, N. A., & Zettlemoyer, L. (2019). Evaluating Gender Bias in Machine Translation. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics.

Toral, A. (2022). Literary Translation and Neural Machine Translation: Overcoming the Limits of Literalism. Translation Spaces, 11(2), 145–162.

Tymoczko, M. (2007). Enlarging Translation, Empowering Translators. London: Routledge.

Van Rijsselbergen, D., & Lefever, E. (2021). Gender Bias in Machine Translation: A Review of Current Research and Future Directions. Language Resources and Evaluation.

Venuti, L. (2012). The Translator's Invisibility: A History of Translation (2nd ed.). London: Routledge.

Vinay, J. P., & Darbelnet, J. (1995). Comparative Stylistics of French and English: A Methodology for Translation (J. C. Sager & M.-J. Hamel, Trans.). Amsterdam: John Benjamins. (Original work published 1958).

Wu, H., Yang, J., & Liu, X. (2022). Interpreting Ambiguity and Allusion in Literary Neural MT: A Semantic-Pragmatic Perspective. Language and Translation Technology.

Wu, Y., & Zhang, X. (2020). Cultural Adaptation in Translation: Balancing Fidelity and Acceptability. Meta: Journal des traducteurs.

Author Biography: