

Case Study

Bias in Google Translate: A Case Study on Gendered Language Disparities in Hindi-to-English Translations

By

Monalisa Dutta

2333150

4BAENG

Christ University, Bannerghatta Road Campus

Date: 10.01.2025

Mozilla RCC Project

Table of Contents

Introduction	3
Case	4
Teaching Note	10
Conclusion	13
References	14

Introduction

With over 600 million Hindi speakers globally and Google Translate serving as a vital tool for bridging language barriers, it becomes critical to address the implicit biases embedded in automated translation systems. Gender bias in such systems, particularly in Hindi-to-English translations, raises significant concerns about fairness, inclusivity, and the trustworthiness of AI-powered tools. These biases reflect and reinforce stereotypes, such as associating professions like “doctor” with men and “nurse” with women, limiting their utility and accuracy for diverse users.

In an increasingly digital and multilingual world, where AI technologies are lauded for their role in enabling global communication, addressing language and cultural disparities becomes imperative. This case study examines the biases in Google Translate, focusing on its impact on Hindi-to-English translations, and emphasizes the importance of equitable design and evaluation in AI tools. By uncovering these biases, this study seeks to highlight the need for gender-neutral, culturally adaptive approaches in AI language systems to ensure inclusive and ethical technology.

Disclaimer: This case is a fictional scenario created for educational purposes. While the characters and their situation are hypothetical, the data and concerns about translation bias in Google Translate are real and based on verified research.

Case

It was a humid afternoon in Delhi when Priya, a young journalist working for a bilingual news outlet, found herself in a bind. She was tasked with translating a critical news report from Hindi to English for her publication's website. To save time, she turned to Google Translate. The original Hindi text was gender-neutral, referring to a local politician as "वह" (/vəhə/), meaning "they" or "he/she" depending on context. However, the English translation decisively rendered the politician as "he," sparking a cascade of misunderstandings in the newsroom and, ultimately, among readers. Priya was left grappling with the implications of the mistranslation and how it subtly perpetuated male-centric narratives in media coverage.

This case study looks at the pervasive issue of gender bias in automated translation systems through the lens of Priya's experience. It delves into how linguistic structures in gender-neutral languages like Hindi interact with algorithmic processes, leading to gendered mistranslations that reinforce societal stereotypes.

As Priya reviewed the mistranslation, she noticed a pattern that was not unique to her story. Automated translation systems often fail to recognize the nuanced contexts of gender-neutral languages. For instance, in Hindi, pronouns like "वह" (/vəhə/) and verbs are not inherently gendered. However, Google Translate frequently assigns masculine gender to these terms when converting them into English, reflecting an inherent algorithmic bias. The result is a distortion of

meaning that aligns with pre-existing societal norms, often prioritizing male-centric interpretations.

This issue is not merely theoretical. A 2020 study published in the *Journal of Artificial Intelligence Research* demonstrated that Google Translate frequently defaults to masculine pronouns in ambiguous contexts across several languages, including Hindi. For instance, sentences such as "वह डॉक्टर है" (/vəhə 'dɔːktər hɛ/, "They are a doctor") are translated as "He is a doctor," while professions traditionally associated with women are often rendered as feminine in translations. This reinforces stereotypes about gender roles in professional contexts, with potential repercussions for industries like media, education, and policymaking.

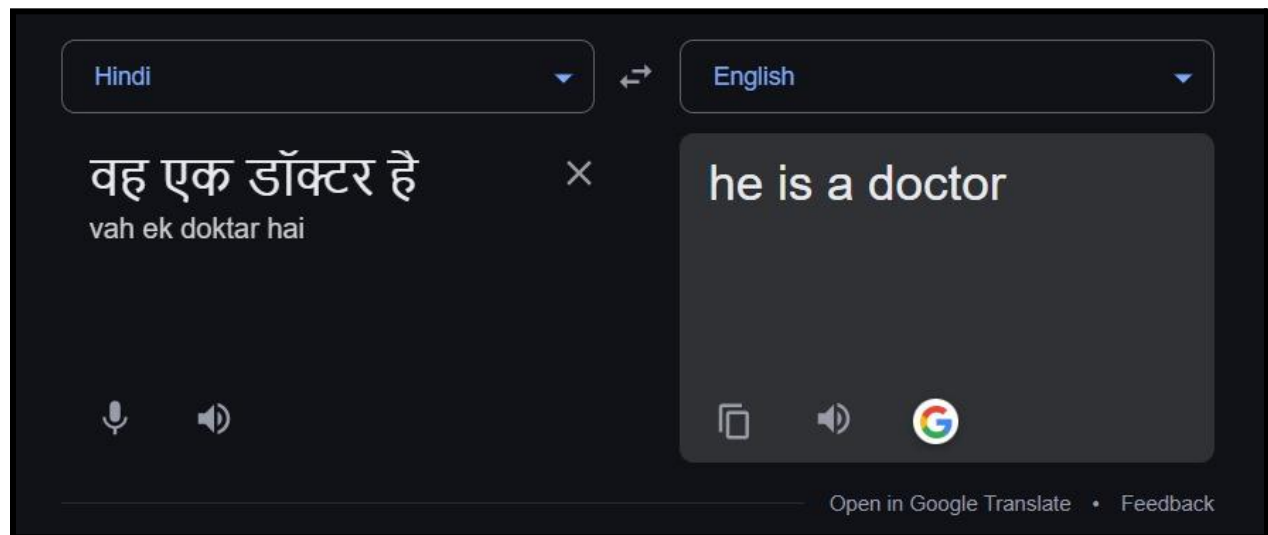


Image 1: Print screen of Google Translate's result for "वह डॉक्टर है" (/vəhə 'dɔːktər hɛ/).

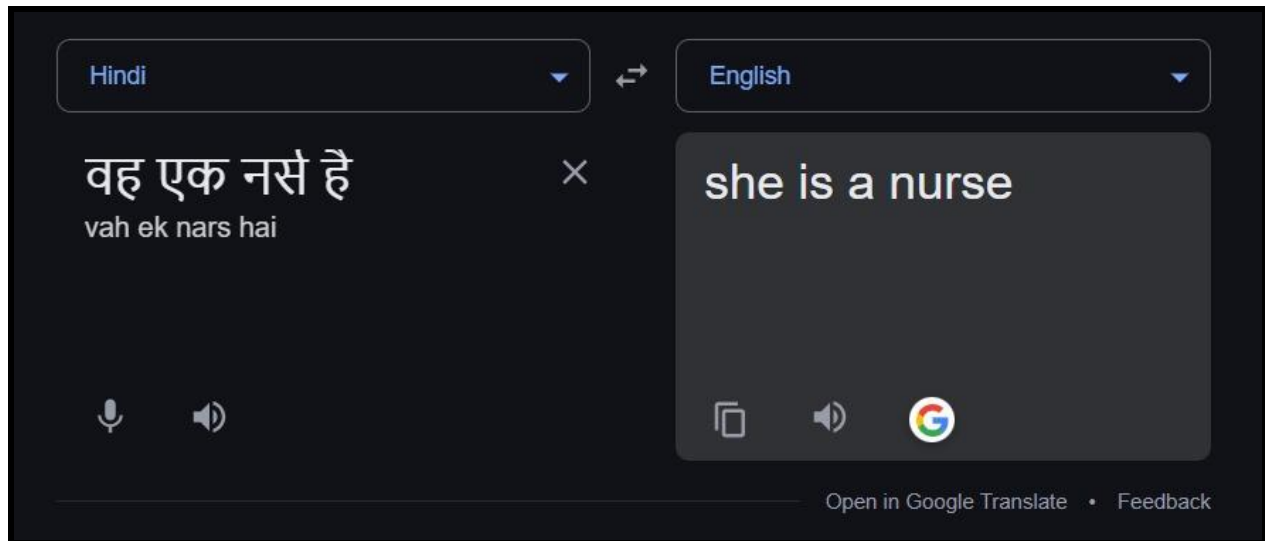


Image 2: Print screen of Google Translate’s result for “वह एक नर्स है” (/vəhə ekə nərsə hɛ/)

Priya’s newsroom experience serves as a representation of the broader societal implications of biased translations. When gender-neutral statements are rendered with masculine defaults, it subtly validates male dominance in cultural narratives. This effect becomes more pronounced in regions like India, where societal structures already grapple with gender disparities.

Consider a scenario where a local government document, originally written in Hindi, is translated for an international audience. If phrases describing leadership roles default to masculine pronouns, it erases the presence of women leaders, further entrenching gender biases. In Priya’s case, her article was meant to highlight the achievements of a female politician. The mistranslation not only misrepresented her gender but also undermined the essence of the story, potentially skewing public perception.

Moreover, these mistranslations can exacerbate issues of inclusivity. Language shapes thought, and repeated exposure to gendered biases in translations can reinforce harmful stereotypes in the

collective consciousness. For young readers or viewers consuming translated content, such biases can subtly influence career aspirations, perpetuating cycles of inequality.

The root of these biases lies in the data and algorithms that power translation systems. Priya's experience reflects the limitations of machine learning models trained predominantly on datasets that mirror existing societal biases. Since Hindi texts in training datasets often implicitly or explicitly favor masculine interpretations, translation algorithms replicate these patterns without questioning their validity.

In addition, automated systems lack contextual awareness. A human translator might infer the gender of a subject based on accompanying information or cultural cues, but machines rely solely on patterns within their training data. For example, if Priya had been translating a historical text mentioning "महान नेता" (/mə'ɦɑ:n 'ne:tɑ/, "great leader"), the system would likely default to "he" as the pronoun, irrespective of whether the leader was male or female. This rigidity underscores the ethical challenge of creating systems that align with evolving societal values and linguistic integrity.

As Priya sat in the newsroom pondering her mistranslation, she couldn't help but wonder: Who bears responsibility for these biases? Is it the developers who create these systems, the organizations that deploy them, or society at large for tolerating such inequities? The answer remains elusive, as addressing bias in translation systems requires a multi-stakeholder approach.

Potential solutions include diversifying training datasets to include balanced representations of genders, enabling user feedback mechanisms to refine translations, and fostering collaboration between linguists and technologists. However, implementing these measures is resource-intensive and requires sustained commitment. The challenge, then, is not merely

technical but also ethical and societal: How do we ensure that technology serves as a tool for equity rather than perpetuating existing divides?

Keywords: Linguistic Diversity, Google Translate, Ethical Development, Inclusivity, Automated Translation.

Learning Objectives:

- Identify and analyze the presence of gender bias in automated translation systems, specifically in Hindi-to-English translations using Google Translate.
- Evaluate the implications of biased AI tools on societal perceptions, inclusivity, and ethical AI development.
- Explore the technical and cultural factors contributing to translation biases and discuss potential solutions to address them.
- Understand the role of linguistics, gender studies, and computer science in developing more equitable AI systems.

Discussion Questions:

1. What were the specific challenges Priya faced while preparing her STEM presentation that exposed the gender bias in Google Translate?
2. How do the gendered translations in Priya's examples reinforce stereotypes, and what impact could this have on her credibility as a speaker?
3. What small but impactful changes could Google Translate incorporate to address the biases Priya discovered?

4. How can Priya's findings about translation bias be used to design AI tools that are more inclusive for culturally and linguistically diverse audiences?
5. Beyond Priya's experience, how might such biases in AI translation systems affect individuals in STEM fields, especially in professional and academic settings?

Teaching Note

Case Overview

This case study examines gender bias in automated translation systems, focusing on a real-world example involving a Hindi-to-English mistranslation. It highlights how machine learning algorithms often default to masculine pronouns in gender-neutral contexts, reflecting biases in training data. The narrative revolves around Priya, a journalist who encounters this issue while translating an article. By exploring the societal, ethical, and technological implications of such biases, the case highlights the challenges of ensuring inclusivity in AI systems. The study aims to equip learners with an understanding of the intersection between natural language processing (NLP), machine learning, and social equity.

Learning Objectives

- **Understand bias in AI systems:** Analyze how biases in training data affect the performance of NLP algorithms, specifically in gender-neutral languages.
- **Explore gender neutrality in linguistic structures:** Examine the challenges that languages with context-based gender markers (e.g., Hindi) pose for machine translation.
- **Evaluate the societal impact of AI decisions:** Assess how mistranslations influence public perceptions, reinforce stereotypes, and affect inclusivity.
- **Discuss ethical responsibilities in AI development:** Identify strategies for addressing and mitigating bias in machine learning models.
- **Apply interdisciplinary approaches:** Integrate concepts from linguistics, ethics, and technology to propose solutions for improving translation accuracy.

Discussion Questions

1. What are the technical limitations of current NLP algorithms in handling gender-neutral languages, and how can these be addressed?
2. How do biases in training datasets propagate into AI systems, and what measures can developers take to mitigate them?
3. In what ways do mistranslations reinforce societal stereotypes, and how can AI systems be made more inclusive?
4. What ethical considerations should guide the development and deployment of automated translation systems?
5. How can interdisciplinary collaboration between linguists, data scientists, and ethicists improve the design of NLP tools?

Key Teaching Concepts

- **Bias in Machine Learning:** Understanding how training data influences AI behavior and perpetuates societal biases.
- **Gendered Linguistics:** Exploring the complexities of gender-neutral languages in machine translation.
- **Algorithmic Accountability:** Exploring the ethical and societal implications of biased AI outputs.
- **Dataset Diversity:** Highlighting the importance of inclusive and representative datasets for training AI models.
- **Problem-Solving:** Leveraging insights from multiple fields to address technological challenges.

Teaching Approach and Methodology

- **Case Study Analysis:** Divide students into small groups to analyze Priya's story and identify the underlying biases and technical issues in the translation system.
- **Role-Playing Activity:** Assign roles (e.g., developer, linguist, journalist) to students and have them propose solutions for improving translation accuracy.
- **Data Exploration:** Provide a sample dataset and task students with identifying potential biases using Python or another analytical tool.
- **Debate:** Host a debate on whether ethical considerations should outweigh technical feasibility in AI development.

Conclusion

The intersection of language and technology holds immense potential to bridge cultural and linguistic divides. However, the presence of biases in automated translation systems, particularly regarding gender representation in Indian languages like Hindi, reveals significant gaps in fairness and inclusivity. These biases not only perpetuate societal stereotypes but also hinder efforts toward gender equality and alienate marginalized communities from digital technologies. The case of gendered mistranslations in Hindi-English conversions underscores the importance of addressing these challenges through ethical AI practices.

This case study, framed through Priya's experience, highlights the need for a concerted effort from developers, linguists, and policymakers to ensure linguistic integrity in AI systems. By embedding inclusivity and cultural sensitivity as foundational principles in AI development, technology can transform from a source of inequity to a tool for empowerment. Achieving linguistic equity in AI is not just a technical challenge but a societal imperative, essential for fostering trust, inclusivity, and accountability in an increasingly interconnected world.

References

- Blodgett, Su Lin, et al. "Language (technology) is power: A critical survey of" bias" in nlp." arXiv preprint arXiv:2005.14050 (2020).
- Garg, Nikhil, et al. "Word embeddings quantify 100 years of gender and ethnic stereotypes." *Proceedings of the National Academy of Sciences* 115.16 (2018): E3635-E3644.
- Hovy, Dirk, and Shannon L. Spruit. "The social impact of natural language processing." *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 2016.
- Prates, Marcelo OR, Pedro H. Avelar, and Luís C. Lamb. "Assessing gender bias in machine translation: a case study with google translate." *Neural Computing and Applications* 32 (2020): 6363-6381.
- Singh, Muskaan, Ravinder Kumar, and Inderveer Chana. "Machine translation systems for Indian languages: review of modelling techniques, challenges, open issues and future research directions." *Archives of Computational Methods in Engineering* 28.4 (2021): 2165-2193.
- Ghosh, Sourojit, and Srishti Chatterjee. "11 Misgendering and Assuming Gender in Machine Translation When Working with Low-Resource Languages." *Gendered Technology in Translation and Interpreting: Centering Rights in the Development of Language Technology* (2024): 274.
- Sun, Tony, et al. "Mitigating gender bias in natural language processing: Literature review." arXiv preprint arXiv:1906.08976 (2019).
- Fitria, Tira Nur. "Gender bias in translation using google translate: Problems and solution." *Language Circle: Journal of Language and Literature* 15.2 (2021).

Lopez-Medel, Maria. "Gender bias in machine translation: an analysis of Google Translate in English and Spanish." (2021).