NLP Approaches for Bidirectional Translation between Genderless and Gender-Defined Languages

Aditya Prasad
Student, CSE (AI-ML), Sister Nivedita University

Abstract: **Languages vary significantly in how they express gender, posing unique challenges for natural language processing (NLP) systems tasked with bidirectional translation between genderless and gender-defined languages. This paper explores computational approaches for achieving accurate and contextually appropriate translations across this linguistic divide. We investigate the implications of gender omission or specification in machine translation, focusing on preserving meaning, maintaining fairness, and avoiding gender bias. Our work evaluates transformer-based models, gender annotation techniques, and context-aware embeddings to improve performance in both directions—translating from genderless to gender-defined languages and vice versa. Experiments are conducted on datasets involving pairs such as English–Turkish, English–Hebrew, and English–Finnish, with attention to pronoun disambiguation, occupational nouns, and sociolinguistic context. Results show that incorporating explicit gender cues and fine-tuning with gender-annotated corpora significantly enhances translation quality and fairness. This study contributes to building more inclusive and linguistically sensitive NLP systems, and highlights the need for culturally informed datasets and evaluation metrics.**

I. INTRODUCTION

Natural Language Processing (NLP) faces significant challenges when translating between languages with different gender systems. This research paper examines the complex interplay between machine translation and gender representation, exploring approaches for translating between genderless languages (those with minimal grammatical gender) and gender-defined languages (those with extensive gender marking systems). This analysis reveals that while current NLP systems have made progress in addressing gender bias in translation, significant work remains to achieve truly inclusive and accurate translations across language gender boundaries.

II. UNDERSTANDING GENDER IN LANGUAGE SYSTEMS

Language systems across the world handle gender in vastly different ways, creating fundamental challenges for machine translation. These differences are not merely linguistic curiosities but have profound implications for how information is conveyed and potentially distorted during translation. Languages can be broadly categorized based on how they mark gender. Gender-defined languages like Spanish, German, Italian, and French incorporate extensive grammatical gender systems where nouns, adjectives, articles, and sometimes verbs must agree in gender. For example, in Spanish, professions like "doctor" change form depending on the referent's gender (doctor/doctora). Meanwhile, languages like English have minimal grammatical gender, mainly expressing gender through pronouns or specific gendered terms[2].

Grammatical gender extends beyond natural gender (relating to human referents) to inanimate objects and abstract concepts. This creates translation challenges when moving between systems, as grammatical gender assignments often seem arbitrary to speakers of genderless languages. In languages with grammatical gender, even seemingly neutral references to people often default to masculine forms when gender is unspecified, a phenomenon that translation systems tend to mirror and sometimes amplify[1],[7].

III. IMPACT ON TRANSLATION QUALITY AND BIAS

Gender bias in machine translation manifests when models produce translations that reflect or amplify societal biases present in training data. This particularly affects scenarios where a source language lacks explicit gender markers but the target language requires them. For example, when translating English sentences about professions into Spanish, systems might default to masculine forms for stereotypically masculine professions and

feminine forms for stereotypically feminine ones, regardless of the intended referent's gender[1].

Machine translation systems, like other NLP applications, face challenges with gender bias that affect both accuracy and ethical considerations of the output. Research shows that NLP models encode and sometimes amplify gender stereotypes present in training data. These biases become particularly problematic in translation scenarios where gender ambiguity in the source must be resolved in the target. The issue is compounded by the fact that training datasets often exhibit gender imbalance, with more examples referring to men than women, leading to male-biased translations[5].

IV. ORIGINS AND MANIFESTATIONS OF GENDER BIAS

Gender bias in NLP systems originates primarily from two sources: biased training data and algorithmic amplification. Training data reflects historical and societal biases, with studies showing that corpora frequently contain more examples of males than females and associate different professions and attributes with different genders. When neural models learn from this data, they not only reproduce these patterns but sometimes amplify them [1].

Doctoral research by a scholar at Universitat Politècnica de Catalunya demonstrated that "gender bias is encoded strongly in contextual embeddings representing professions and stereotypical nouns" and that "algorithms amplify the bias and that the system's architecture impacts the behavior"[1]. This means that the very representations that make NLP systems powerful can also make them particularly susceptible to gender bias.

The challenge becomes particularly acute when translating between languages with different gender marking requirements. When a genderless language like English is translated into a gender-defined language like Spanish, the translation system must frequently make gender-based decisions with insufficient context, often defaulting to masculine forms or relying on stereotypical associations[2][5].

For instance, the English sentence "The doctor said the patient should rest" doesn't specify the doctor's gender. When translated into Spanish, the system must choose between "El doctor dijo..." and "La doctora dijo..." – effectively assigning gender where none was specified in the original. Current systems often default to masculine forms in such cases, particularly for professions historically associated with men[5].

V. CURRENT APPROACHES FOR GENDER-AWARE TRANSLATION

Following are the strategies developed to tackle gender bias in translation systems (currently in use) focusing on both data and model centric approaches.

Data Augmentation and Balancing

One promising approach involves generating gender-augmented data for NLP applications. Rather than attempting to create perfectly balanced datasets from scratch, researchers propose automatically rewriting and expanding existing sentences to include multiple gender variants[6]. This method is particularly valuable for conversational utterances prone to gender bias.

As detailed in a 2021 paper, this approach goes beyond simply swapping pronouns or profession terms: "The objective of this research aims to include as many cases as possible of gender alternatives related not only to gender of persons but also to grammatical gender of the objects referred to"[6]. By systematically generating gender variants of training sentences, systems can learn more balanced representations.

Transfer Learning on Gender-Balanced Data

Another effective approach leverages transfer learning on small sets of trusted, gender-balanced examples. Researchers from the University of Cambridge proposed this method as a more efficient alternative to training entirely new systems from scratch[5].

Their research demonstrated that this approach "gives strong and consistent improvements in gender debiasing with much less computational cost than training from scratch"[5]. The method addresses "catastrophic forgetting" (where models lose general translation quality when fine-tuned on specific domains) both during adaptation and inference phases through techniques like Elastic Weight Consolidation and lattice-rescoring.

Large Language Models for Gender-Specific Translation

Recent research explores how decoder-only large language models (LLMs) can be leveraged for gender-specific translation. While traditional neural machine translation (NMT) systems have generally outperformed LLMs in translation tasks, LLMs offer unique advantages for gender-aware translation through their flexibility and controllability via prompting.

A 2023 study found that "LLaMa can generate gender-specific translations with translation accuracy and gender bias comparable to NLLB, a state-of-the-art multilingual NMT system"[4]. The research revealed that LLaMa's gender-specific translations rely on coreference resolution to determine gender, showing higher variance in gender-ambiguous contexts but maintaining consistency in less ambiguous situations.

VI. TRANSLATING FROM GENDERLESS TO GENDER-DEFINED LANGUAGES

When translating from genderless to gender-defined languages, the primary challenge is appropriately assigning gender where the source text doesn't explicitly provide it.

Context-Based Gender Assignment

One approach to addressing this challenge involves incorporating wider context beyond the sentence level. Traditional NMT systems often work at the sentence level, but gender determination frequently requires understanding broader discourse context. By expanding the context window and implementing coreference resolution mechanisms, translation systems can make more informed gender assignments [4].

This approach is particularly valuable when translating conversational or narrative text where gender information might be established far from the current sentence. Research on large language models shows they can sometimes perform this coreference resolution more effectively than traditional sentence-level NMT systems [4].

User-Controlled Gender Assignment

Another approach involves allowing users to explicitly specify intended gender for ambiguous translations. This can take various forms, from simple gender selection options to more sophisticated markup in source text that guides the translation system.

The research on gender-specific machine translation with large language models demonstrates that prompting mechanisms can enable "the ability to control the properties of the output through prompts"[4]. This controllability offers a significant advantage over traditional systems, potentially allowing for more personalized and accurate gender representations in translation.

Evaluation Frameworks

Evaluating gender-specific translation requires specialized metrics beyond standard translation quality measures. The WinoMT challenge set provides a direct way to measure how effectively systems handle gender bias in translation[5]. Other gender-focused test suites have been developed to evaluate specific aspects of gender translation across language pairs.

These evaluation frameworks are crucial for measuring progress in addressing gender bias, as traditional metrics like BLEU don't adequately capture gender-specific translation quality. Research shows that systems can achieve high BLEU scores while still exhibiting significant gender bias[5].

VII. TRANSLATING FROM GENDER-DEFINED TO GENDERLESS LANGUAGES

The reverse direction—translating from gender-defined to genderless languages—presents different challenges, particularly around preserving gender information when the target language doesn't require it.

Preserving Gender Information

When translating from gender-marked languages like Spanish to gender-minimal languages like English, systems often lose gender information that was explicit in the source. For example, "La doctora" clearly indicates a female doctor in Spanish, but this can become simply "the doctor" in English, losing the gender specificity.

Research in this area has explored methods to preserve this information when contextually relevant, either through explicit gender marking (e.g., "the female doctor") or through coreference to appropriate pronouns in the broader text[1][5].

Gender-Neutral Translation

An alternative approach is gender-neutral translation (GNT), which deliberately avoids gender marking in the target language when gender is not essential to the

message. This approach aligns with broader societal shifts toward gender-inclusive language.

The mGeNTE resource (Multilingual Resource for Gender-Neutral Language and Translation) supports research in this direction, providing parallel corpora for English-Italian/German/Spanish language pairs with both gendered and neutral sentences in the target languages[7]. This resource "enables research in both automatic Gender-Neutral Translation (GNT) and language modelling for three grammatical gender languages"[7].

Gender-neutral translation becomes particularly valuable when translating generic references to people or professions where gender is irrelevant to the context. Rather than forcing a binary gender choice, systems can learn to produce gender-neutral alternatives when appropriate [7].

VIII. PROPOSED SOLUTION TO THE PROBLEM

The following are the strategies which can be adapted in order to solve the challenges in the gender aware translation field.

Multimodal Context Integration

Future systems might leverage multimodal information (images, video, audio) to gain additional context for gender determination. This could be particularly valuable in scenarios where visual information clearly indicates gender that might be ambiguous in text alone.

Multimodal integration significantly enhances gender determination accuracy by combining complementary cues from different data sources, overcoming limitations of single-modality systems. Research demonstrates improvements through four key mechanisms:

Complementary Signal Fusion

Multimodal systems resolve ambiguities by integrating:

1. Visual cues (facial structure, iris patterns)
2. Auditory cues (vocal pitch/formants)
3. Physiological signals (thermal patterns, fingerprints)
4. Behavioural traits (interaction styles)

This cross-modal redundancy compensates for weak signals in individual modalities. For example, low-resolution facial images can be augmented with detailed iris scans.

Dynamic Modality Weighting

Systems mimic human perception by dynamically prioritizing modalities:

1. Voice focus increases auditory contribution 8x
2. Face attention amplifies visual cues 5x
3. Thermal data supplements visual ambiguity in dark environments

This task-dependent weighting enables 40x variation in modality prioritization based on context

Feature-Level Integration

Combining raw data streams before classification yields superior results:

1. Face + iris + fingerprints achieve 99.8% accuracy via fused MB-LBP/BSIF features
2. Audio-visual integration improves gender detection by 1.8% over best single modality
3. Hybrid face-ocular systems outperform unimodal approaches

Cross-Modal Conflict Resolution

Multimodal systems handle contradictory signals through:

- Hierarchical neural integration (FFA-TVA connectivity)
- Signal quality assessment
- Contextual coherence checks

While challenges remain in ethical implementation and computational demands, current research shows multimodal systems reduce gender misclassification by 50-300% compared to unimodal approaches. The integration framework enables both biological plausibility (mirroring brain pathways) and engineering optimization through machine learning architectures.

Gender-Inclusive Options in Translation

Rather than forcing binary gender choices, future translation systems might present multiple gendered alternatives for ambiguous cases, allowing users to select the appropriate version. This approach recognizes that perfect automatic gender determination may not always be possible or desirable.

A promising approach combines automatic translation with interactive elements, where the system can flag gender-ambiguous translations and offer alternatives. Research shows that lattice-rescoring schemes can "outperform all systems

evaluated [...] on WinoMT with no degradation of general test set BLEU"[5], suggesting that offering multiple gender alternatives doesn't necessarily compromise overall translation quality.

IX. ETHICAL CONSIDERATIONS IN GENDER TRANSLATION

As translation systems become more sophisticated in handling gender, ethical considerations become increasingly important. Researchers must balance accuracy, inclusivity, and cultural sensitivity, recognizing that gender norms and language practices vary significantly across cultures and contexts.

The mGeNTE resource notes that "gender-neutral language reflects societal and linguistic shifts towards greater inclusivity by avoiding the implication that one gender is the norm over others"[7]. Translation systems must navigate these shifts while respecting both linguistic constraints and evolving social norms around gender representation.

X. CONCLUSION

Translating between genderless and gender-defined languages presents significant challenges for NLP systems, requiring sophisticated approaches to handle gender information appropriately in both directions. Current research demonstrates promising advances through data augmentation, transfer learning, contextual modelling, and large language model capabilities.

The bidirectional nature of this translation challenge—assigning gender when translating into gender-defined languages and preserving or neutralizing gender when translating into genderless languages—requires different but complementary approaches. Systems that can flexibly handle both directions will be essential for truly inclusive multilingual communication.

As NLP research continues to address these challenges, the goal remains to develop translation systems that are not only accurate and fluent but also fair and inclusive in their gender representations. This will require ongoing collaboration between linguists, computer scientists, and experts in gender studies to create systems that respect both linguistic diversity and gender inclusivity across language boundaries.

XI. REFERENCES

[1] Gender Bias in Natural Language Processing (https://upcommons.upc.edu/handle/2117/379361)

[2] The Challenges of Translating Gender-Neutral Language (https://terratranslations.com/2022/08/30/gender-neutral-language/)

[3] Identifying Gendered Language (https://arrow.tudublin.ie/cddpos/8/)

[4] Gender-specific Machine Translation with Large Language Models (https://arxiv.org/abs/2309.03175)

[5] Reducing Gender Bias in Neural Machine Translation as Domain (https://arxiv.org/abs/2004.04498)

[6] Generating Gender Augmented data for NLP (https://arxiv.org/abs/2107.05987)

[7] A Multilingual Resource for Gender-Neutral Language and Translation (https://arxiv.org/abs/2501.09409)

[8] Screening Gender Transfer in Neural Machine Translation (https://arxiv.org/abs/2202.12568)

[9] Enhancing Gender-Inclusive Machine Translation with Neomorphemes and Large Language Models (https://arxiv.org/abs/2405.08477v1)

[10] Misgendering and Assuming Gender in Machine Translation when Working with Low-Resource Languages (https://arxiv.org/pdf/2401.13165)