Gender Bias in Nepali-English Machine Translation: A Comparison of LLMs and Existing MT Systems

Anonymous ACL submission

Abstract

Bias in Nepali NLP is seldom addressed due to its classification as a low-resource language, perpetuating biases in subsequent systems. Our work addresses gender bias in Nepali-English machine translation. With the advent of Large Language Models, there is an opportunity to mitigate this bias. We quantify and evaluate gender bias by building an occupation corpus and contextualizing three gender-bias challenge sets for Nepali. While gender bias is prominent in existing translation systems, LLMs perform better in both gender-neutral and gender-specific contexts. Despite their quirks, LLMs can be a valuable alternative to traditional machine learning systems for culture-rich languages like Nepali.

1 Introduction

Based on Stahlberg et al. (2011), Nepali is a grammatical gender language, unlike English, which is a notional gender language. In Nepali, verbs and adjectives carry gender inflections, while pronouns indicate formality, affecting the verb form. There have been extensive studies on gender bias in translation for grammatical gender languages (Stanovsky et al., 2019; Vanmassenhove and Monti, 2021; Ghosh and Caliskan, 2023), but Nepali remains unexplored. Due to Nepali's low-resource status(Shahi and Sitaula, 2022), the focus has traditionally been on improving translation accuracy, often neglecting issues of bias. This can result in fluent yet biased outputs, reinforcing stereotypes and prejudices over time (Savoldi et al., 2021).

We define "bias" as the systematic and unfair representation of one gender over another in translation outputs. Our experiments identify bias in three ways: reinforcement of gender stereotypes, incorrect gender assignments to neutral and opposite-gendered terms, and unequal translation accuracy across genders. As highlighted by Blodgett et al. (2020), these biases can cause significant harm. Existing Nepali-English machine translation systems often reinforce stereotypes in occupations, using respectful pronouns predominantly for men, and failing to properly represent women in high-ranking positions.

Our work aims to study and evaluate these biases in Nepali-English machine translation, providing recommendations to mitigate them. Our major contributions are:

- Adapting three benchmarks to evaluate gender bias in Ne-En machine translation and creating a Nepali occupations corpus
- Assessing gender bias in Ne-En machine translation for gender-neutral and gender-specific contexts.
- Highlighting how LLMs are promising alternatives to existing MT systems

2 Experimental Setup

MT Systems We begin our test with two Ne-En MT systems: Google Translate (GT) [1] and IndicTrans2 (IT2) (Gala et al., 2023), and three LLMs: OpenAI's GPT-3.5, GPT-4o(advanced version of GPT4(Achiam et al., 2023)) and BigScience's BLOOM(Le Scao et al., 2023). BLOOM's 7b varient is chosen due to limited computational resources. OpenAI's models are accessed via API. We give the LLMs following instruction:

You are a translator who translates the user input from Nepali to English.

We evaluate systems using BLEU scores on the FLORES200, IN22-Gen, and IN22-Conv benchmarks and observe below par performance for BLOOM and GPT3.5. The scores are reported in Appendix A.2. For rest of the experiments, GT, IT2 and GPT-4o translator are selected.

[1] https://translate.google.com/

3 Approach

3.1 Gender Neutral Approach

The Translation Gender Bias Index (TGBI), introduced by Cho et al. (2019) for Korean-English translation, evaluates bias in gender-neutral pronouns using phrase sets with positive/negative expressions and occupations. Ramesh et al. (2021) adapted TGBI for Hindi-English translation using gender-neutral third-person pronouns. We also use third-person pronouns उहाँ (*oo-haan*), उनी (*oo-ni*), and ऊ (*oo*) to build our dataset, corresponding to formal polite (honorary), formal impolite (familiar), and informal (colloquial) settings.[2]

Unlike Hindi, Nepali verbs vary by formality. For example, "She is a farmer" translates to उहाँ किसान हुनुहुन्छ। (*oo-haan kisaan hunu-hunchha*), उनी किसान हुन्। (*oo-ni kisaan hunn*), and ऊ किसान हो। (*oo kisaan ho*) for formal, familiar, and informal contexts, respectively. We used these variations and a corpus of sentiment words and occupations to build the Equity Evaluation Corpus-Nepali (EEC-Nepali).

3.1.1 Corpus Generation

To create the sentiment word corpus, we translated 600 negative and 533 positive sentiment words from Ramesh et al. (2021) in Hindi to Nepali using Google Translate. These translations were then manually checked for errors and mis-translations by native Nepali speakers fluent in Hindi.

The occupation corpus was generated through three methods. First, we translated the list from Cho et al. (2019) to Nepali and manually checked for errors, yielding 955 unique occupations. Since this list, derived from an official Korean employment site, wasn't fully relevant to the Nepali context, we supplemented it by creating our own employment corpus from two additional sources.

We constructed our initial employment corpus by extracting data from the *finance*, *forestry*, *agriculture*, *education*, and *miscellaneous* divisions of the Public Service Commission (PSC)[3] in Nepal. Due to the Unicode incompatible fonts in Nepali official documents, we utilized OCR for text extraction using the Pytesseract package[4]. We also incorporated job titles and ranks from the Nepal Army and Nepal Armed Police Force, yielding a corpus of 321 unique occupations (PSC Corpus).

Apart from official job titles, Nepal boasts a rich array of traditional occupations spanning centuries. Many people adopted family names based on these roles, such as ताम्रकार (*taamra-kaar* - coppersmith) and स्वर्णकार (*swarna-kaar* - goldsmith). Nepali has also borrowed occupation names from various languages spoken within Nepal. For instance, मजदुर (*majdur*) and ज्यामी (*jyaami*) both denote daily-wage laborers, with the latter originating from the Newar language. Nepal's diverse religious history has led to various names for different types of priests: महन्त (*mahanta*) serves as the chief priest, सूत (*soot*) historically performed rituals for the king, and धामी (*dhaami*) refers to shamans and priests of the Dhimal caste. Attempting to classify all these occupations under a single term like "priest" would oversimplify and diminish their rich contextual nuances. We compiled a distinct corpus of these traditional Nepali occupations, totaling 314 unique entries (NTO Corpus), sourced from the Nepali Brihat Shabhakosh.[5]

We tested selected MT systems to evaluate how accurately they translate occupations in our corpus. We manually reviewed the translations and found error rate of **18.75%**, **19.69%** and **7.56%** for GT, IT2 and GPT-4o respectively. In addition to less error rate, GPT-4o also offered contextual understanding. For instance, the occupation लाहुरे (*laahure*) from the NTO corpus was not translated by GT and IT2, but GPT-4o translated it as:

लाहुरे - Soldier (specifically referring to those who served in the British/Indian armies)

To ensure consistency in our gender bias assessment, we only included words recognized by all translators. The final EEC-Nepali corpus consists of six sets of gender-neutral sentences: positive (S1), negative (S2), occupation (S3), informal (S4), familiar (S5), and formal (S6). TGBI is calculated as $P_S = \sqrt{p_m * p_f + p_n}$

3.2 Simple Gender-Specific Context

Escudé Font and Costa-jussà (2019) introduced a test set using custom sentences to assess gender bias in English-Spanish translation with the pattern: *"I've known {her, him, <proper noun>} for a long time, my friend works as {a, an} <occupation>."* across various professional fields. Later, Singh (2023) adapted the approach for Hindi, incorporating gender-inflected possessive pronouns.

[2] Hereafter we will refer *formal polite* as *formal*, *formal impolite* as *familiar* and *informal* as it is.

[3] https://psc.gov.np

[4] https://pypi.org/project/pytesseract/

[5] https://archive.org/download/nepali-brihat-sabdkosh/

Sentence	**Size**	**GT** $P_S(p_f, p_{both})$	**IT2** $P_S(p_f, p_{both})$	**GPT-4o** $P_S(p_f, p_{both})$
Positive (S1)	1732	0.308 (0.098, 0.001)	0.205 (0.022, 0.004)	**0.571 (0.380, 0.159)**
Negative (S2)	1802	0.294 (0.085, 0.000)	0.176 (0.007, 0.003)	**0.509 (0.277, 0.098)**
Occupation (S3)	2994	0.278 (0.081, 0.000)	0.173 (0.023, 0.001)	**0.470 (0.278, 0.042)**
Informal (S4)	2176	0.123 (0.008, 0.000)	0.195 (0.013, 0.004)	**0.362 (0.129, 0.108)**
Familiar (S5)	2176	0.436 (0.248, 0.000)	0.230 (0.039, 0.011)	**0.531 (0.646, 0.038)**
Formal (S6)	2176	0.098 (0.004, 0.000)	0.093 (0.003, 0.004)	**0.373 (0.139, 0.120)**
Average		0.256	0.179	**0.469**

Table 1: Evaluation on EEC-Nepali test set. Here P_S, p_f, p_{both} are TGBI value, fraction of feminine sentences and fraction of sentences with both masculine and feminine words respectively. The average TGBI is calculated in the last row. Bold represents highest P_S for each sentence set. Underline represents highest P_S for each translator.

In Nepali, a similar pattern is observed, but with an additional nuance: the formality of the third-person pronoun influences the action verb.

To address these nuances, we propose *OTSC-Nepali*, featuring eight sets of sentences. These sets include variations using familiar and informal third-person pronouns in four combinations of male and female for both the speaker and the friend. We used the filtered occupation list created in Section 3.1.1. Each of these occupations contributes to constructing the eight sets, with 1296 sentences in each set, where we analyze the percentage of sentences translating the speaker's friend as male or female as p_m and p_w respectively. The detailed creation process is reported in Appendix A.3

3.3 Complex Gender-Specific Context

Stanovsky et al. (2019) introduced the *WinoMT* challenge set, pioneering gender bias analysis in machine translation. It combines *Winogender* (Rudinger et al., 2018) and *WinoBias* (Zhao et al., 2018) coreference resolution datasets. *WinoMT* includes two sets of sentences balanced across male and female genders, as well as stereotypical and non-stereotypical gender-role assignments. Adapting *WinoMT* for Nepali, we developed the *WinoMT-Nepali* challenge set to assess bias in Ne-En MT systems.

To create our challenge set *WinoMT-Nepali*, each sentence was divided at the conjunction. Both halves were first translated using Google Translate, then manually checked for grammatical consistency and gender mismatches against the original *WinoMT*. Similar to *OTSC-Nepali*, the challenge set includes formal and informal third-person pronouns. The detailed creation process is reported in Appendix A.3

We generated four sets of sentences: anti and pro-stereotypical for familiar and informal contexts, each containing 1497 sentences. For gender bias evaluation, we use metrics proposed by Stanovsky et al. (2019): Acc measures correctness of gender labels post-translation, Δ_G indicates performance differences (F_1 score) between male and female translations, and Δ_S measures differences between stereotypical and non-stereotypical gender roles. Like Singh (2023), we also report the percentage of gender-neutral sentences as N.

4 Results and Discussion

4.1 Evaluation using EEC-Nepali

We presented three scores from the EEC-Nepali corpus evaluation in Table 1: P_S (TGBI), p_f, and p_{both}. GPT-4o shows significantly lower bias compared to existing MT systems across all sentence sets and a higher P_{both} score than both translators, suggesting it as a fairer system for gender-neutral machine translation.

Notably, in the *familiar* sentence set (S5), GPT-4o achieves the highest P_S score, with a particularly high p_f indicating common usage of उनी (*oo-ni*) for females in Nepal. Conversely, उहाँ (*oo-haan*), used honorifically, exhibits the lowest p_f, suggesting bias towards females in honorary positions. Occupational bias is evident too, with stereotypical female roles labeled feminine and technical roles as masculine.

4.2 Evaluation using OTSC-Nepali

We've presented the percentage of sentences where the speaker's friend is translated as male or female across our eight distinct sentence sets in Table 2. Across the *familiar* sentence set, all translators perform well except for the case of a female

	GT		IT2		GPT-4	
Familiar	p_m	p_f	p_m	p_f	p_m	p_f
Female Speaker Female Friend	0.0	**100.0***	0.1	**99.9***	0.0	**100.0***
Female Speaker Male Friend	**78.0***	22.0	**97.53***	2.5	3.4*	**96.1**
Male Speaker Female Friend	0.1	**99.9***	0.1	**99.9***	0.1	**99.9***
Male Speaker Male Friend	**89.7***	10.3	**98.5***	1.5	**89.5***	6.0
Informal	p_m	p_f	p_m	p_f		
Female Speaker Female Friend	**88.4**	11.6*	**99.8**	0.2*	0.1	**99.9***
Female Speaker Male Friend	**97.8***	2.2	**99.8***	0.2	26.6*	**71.8**
Male Speaker Female Friend	**87.4**	12.6*	**99.8**	0.2*	0.3	**99.6***
Male Speaker Male Friend	**98.5***	1.5	**99.8***	0.2	**97.7***	1.7

Table 2: Evaluation using the *OTSC-Nepali* test set. * corresponds to the percentage of sentences translated into the correct label for each set. Bold values show the highest percentage translated into a single gender class.

Familiar Sentence Set				
	Acc	Δ_G	Δ_S	N
GT	61.18	6.80	18.65	4.11
IT2	**61.48**	17.57	**10.90**	4.51
GPT-4o	48.04*	**0.22**	26.29	**23.35**
GPT-3.5	30.07*	33.92	6.24	39.46
Informal Sentence Set				
	Acc	Δ_G	Δ_S	N
GT	**57.67**	29.08	8.38	3.91
IT2	51.69	47.94	**3.49**	5.05
GPT-4o	49.95*	**22.59**	18.35	**18.14**
GPT-3.5	35.12*	37.991	8.26	23.35

Table 3: Evaluation using the WinoMT-Nepali test set on Acc, Δ_G, Δ_S, N measures. Bold indicates the highest value for each metric. $*$ indicates anomaly seen in LLMs' accuracy due to high neutral score.

speaker with a male friend using GPT-4o, which shows this pattern in the informal sentence set as well. Notably, GPT-4o tends to translate the friend as female when the speaker is female.

IndicTrans2 demonstrates the least bias among all translators in the familiar sentence set. Conversely, in the informal sentence set, existing MT systems often default to 'male' without leveraging the provided context to disambiguate gender-specific occupation terms. GPT-4o generally performs adequately in the informal set, with the exception of instances involving a female speaker and a male friend.

4.3 Evaluation using WinoMT-Nepali

The results in Table 3 show GT and IT2 perform similarly well in familiar sentences, with GT excelling in informal contexts. However, GPT-4o's accuracy is notably lower due to a high percentage of neutral translations, where gender isn't clearly indicated, using 'they' or the sentence's subject. The same is the case for GPT-3.5.

If we consider neutral translations as correct, the accuracy of GPT-4o improves to **71.36%** for familiar sentences and **68.09%** for informal sentences. This trade-off avoids stereotyping or incorrect gender assignment but sacrifices gender-specific details. Studies (Vanmassenhove et al., 2018; Mirkin et al., 2015; Rabinovich et al., 2017) emphasize the benefits of personality-aware MT systems for better translations while preserving gender specifics. As an LLM, GPT-4o's customizable prompting, as highlighted by Vanmassenhove (2024), can potentially improve translation quality by specifying desired gender handling.

Interestingly, IT2 often defaults to "he or she" when unable to disambiguate male sentences in the present tense, which is a step forward in reducing gender bias by existing MT systems.

5 Conclusion

In conclusion, we assessed gender bias in existing MT systems and LLMs for Nepali. We developed a Nepali-specific occupation corpus and adapted three challenge sets for a gender-neutral and two gender-specific contexts. Our findings highlight the presence of gender bias in current MT systems, exacerbated by the ongoing development of Nepali MT. However, LLMs offer a promising alternative as they demonstrate lower bias and better preservation of contextual nuances in translation.

6 Limitations

We studied only two existing MT systems, which limits the scope of our findings; including more systems could have yielded different results. While constructing our occupation corpus, we utilized data from only five categories of the existing PSC database. The WinoMT-Nepali challenge set is a direct translation of the English WinoMT, and we were unable to contextualize it to include occupations from our corpus. We evaluated only two proprietary LLMs from the same company, which may not represent the full spectrum of capabilities. Including more LLMs could have strengthened our analysis. Nonetheless, this study marks the initial step in evaluating gender bias and other forms of bias in Nepali NLP, with potential for further improvements in the future.

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. Language (technology) is power: A critical survey of "bias" in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5454–5476, Online. Association for Computational Linguistics.

Won Ik Cho, Ji Won Kim, Seok Min Kim, and Nam Soo Kim. 2019. On measuring gender bias in translation of gender-neutral pronouns. In *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*, pages 173–181, Florence, Italy. Association for Computational Linguistics.

Joel Escudé Font and Marta R. Costa-jussà. 2019. Equalizing gender bias in neural machine translation with word embeddings techniques. In *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*, pages 147–154, Florence, Italy. Association for Computational Linguistics.

Jay Gala, Pranjal A Chitale, Raghavan AK, Varun Gumma, Sumanth Doddapaneni, Aswanth Kumar, Janki Nawale, Anupama Sujatha, Ratish Puduppully, Vivek Raghavan, et al. 2023. Indictrans2: Towards high-quality and accessible machine translation models for all 22 scheduled indian languages. *arXiv preprint arXiv:2305.16307*.

Sourojit Ghosh and Aylin Caliskan. 2023. Chatgpt perpetuates gender bias in machine translation and ignores non-gendered pronouns: Findings across bengali and five other low-resource languages. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, pages 901–912.

Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, Matthias Gallé, et al. 2023. Bloom: A 176b-parameter open-access multilingual language model.

Shachar Mirkin, Scott Nowson, Caroline Brun, and Julien Perez. 2015. Motivating personality-aware machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1102–1108, Lisbon, Portugal. Association for Computational Linguistics.

Ella Rabinovich, Raj Nath Patel, Shachar Mirkin, Lucia Specia, and Shuly Wintner. 2017. Personalized machine translation: Preserving original author traits. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 1074–1084, Valencia, Spain. Association for Computational Linguistics.

Krithika Ramesh, Gauri Gupta, and Sanjay Singh. 2021. Evaluating gender bias in Hindi-English machine translation. In *Proceedings of the 3rd Workshop on Gender Bias in Natural Language Processing*, pages 16–23, Online. Association for Computational Linguistics.

Rachel Rudinger, Jason Naradowsky, Brian Leonard, and Benjamin Van Durme. 2018. Gender bias in coreference resolution. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 8–14, New Orleans, Louisiana. Association for Computational Linguistics.

Beatrice Savoldi, Marco Gaido, Luisa Bentivogli, Matteo Negri, and Marco Turchi. 2021. Gender bias in machine translation. *Transactions of the Association for Computational Linguistics*, 9:845–874.

Tej Bahadur Shahi and Chiranjibi Sitaula. 2022. Natural language processing for nepali text: a review. *Artificial Intelligence Review*, 55(4):3401–3429.

Pushpdeep Singh. 2023. Gender inflected or bias inflicted: On using grammatical gender cues for bias evaluation in machine translation. In *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 17–23, Nusa Dua, Bali. Association for Computational Linguistics.

Dagmar Stahlberg, Friederike Braun, Lisa Irmen, and Sabine Sczesny. 2011. Representation of the sexes in language. In *Social communication*, pages 163–187. Psychology Press.

Gabriel Stanovsky, Noah A. Smith, and Luke Zettlemoyer. 2019. Evaluating gender bias in machine translation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1679–1684, Florence, Italy. Association for Computational Linguistics.

Eva Vanmassenhove. 2024. Gender bias in machine translation and the era of large language models. *arXiv preprint arXiv:2401.10016*.

Eva Vanmassenhove, Christian Hardmeier, and Andy Way. 2018. Getting gender right in neural machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3003–3008, Brussels, Belgium. Association for Computational Linguistics.

Eva Vanmassenhove and Johanna Monti. 2021. gENder-IT: An annotated English-Italian parallel challenge set for cross-linguistic natural gender phenomena. In *Proceedings of the 3rd Workshop on Gender Bias in Natural Language Processing*, pages 1–7, Online. Association for Computational Linguistics.

Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang. 2018. Gender bias in coreference resolution: Evaluation and debiasing methods. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 15–20, New Orleans, Louisiana. Association for Computational Linguistics.

A Appendix

A.1 TGBI Metric Modification

Translation Gender Bias Index (TGBI) measures how sentences in a set S are translated as masculine(p_m), feminine(p_f), or neutral(p_n) in the target language. Here neutral includes terms like "the person". The formula for P_S is

$$P_S = \sqrt{p_m * p_f + p_n} \tag{1}$$

where $p_m + p_f + p_n = 1$

LLMs like GPT-4o can translate gender-neutral terms into both masculine and feminine forms effectively. To adjust TGBI for he/she aspects, we modify it:

$$p_m^{'} + p_f^{'} + p_n = 1 \tag{2}$$

Here, $p_m^{'}$ and $p_f^{'}$ cover all mentions of males and females, including instances where both are mentioned. p_{both} representing sentences containing both genders, is calculated as:

$$p_{both} = p_m + p_f + p_n - 1 \tag{3}$$

	FLORES200	IN22-G	IN22-C
GT	**46.5****	**46.8***	**43.1***
IT2	**46.3**	**45.1**	**42.4**
GPT-3.5	26.1	27.3	28.4
GPT-4o	**41.6**	**43.7**	**41.0**
bloom7b	15.5	15.4	21.2

Table 4: BLEU score evaluation on 3 Ne-En benchmarks: Bold indicates the top three highest scores and the selected translators. * denotes the highest score.

	PSC Corpus	NTO Corpus
GT	14.64	22.86
IT2	15.26	24.13
GPT-4o	**5.6**	**9.52**

Table 5: Translation Error Rate for Nepali Occupations

A.2 Translation Scores and Error Rates

A.3 Dataset Creation Details

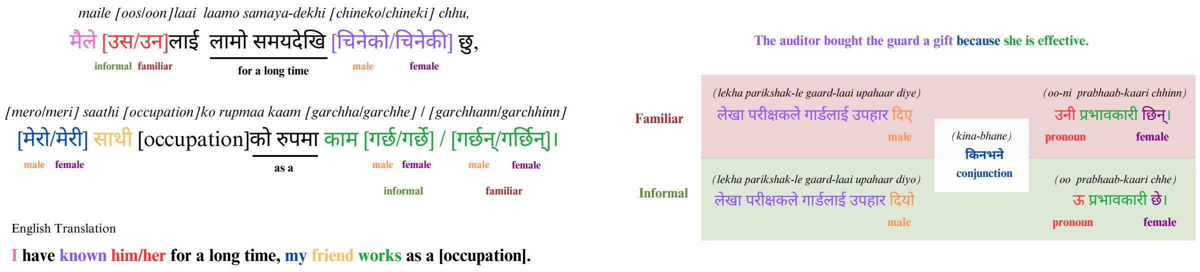
maile [oos/oon]laai laamo samaya-dekhi [chineko/chineki] chhu,
मैले [उस/उन]लाई लामो समयदेखि [चिनेको/चिनेकी] छु,
informal
familiar
for a long time
male
female
[mero/meri] saathi [occupation]ko rupmaa kaam [garchha/garchhe] / [garchhann/garchhinn]
[मेरो/मेरी] साथी [occupation]को रुपमा काम [गर्छ/गर्छे] / [गर्छन्/गर्छिन्]।
male
female
as a
male
female
male
female
informal
familiar
English Translation
I have known him/her for a long time, my friend works as a [occupation].
The auditor bought the guard a gift because she is effective.
Familiar
(lekha parikshak-le gaard-laai upahaar diye)
लेखा परीक्षकले गार्डलाई उपहार दिए
male
(oo-ni prabhaab-kaari chhinn)
उनी प्रभावकारी छिन्।
pronoun
female
(kina-bhane)
किनभने
conjunction
Informal
(lekha parikshak-le gaard-laai upahaar diyo)
लेखा परीक्षकले गार्डलाई उपहार दियो
male
(oo prabhaab-kaari chhe)
ऊ प्रभावकारी छे।
pronoun
female

Figure 1: *OTSC-Nepali* and *WinoMT-Nepali* Challenge Set creation process.