Investigating the Effect of Backtranslation for Indic Languages

Sudhansu Bala Das[1], Samujjal Choudhury[1], Tapas Kumar Mishra[1], Bidyut Kr. Patra[2]

[1]National Institute of Technology (NIT), Rourkela, Odisha, India
[2] Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh, India
Correspondence: baladas.sudhansu@gmail.com

Abstract

Neural machine translation (NMT) is becoming increasingly popular as an effective method of automated language translation. However, due to a scarcity of training datasets, its effectiveness is limited when used with low-resource languages, such as Indian Languages (ILs). The lack of parallel datasets in Natural Language Processing (NLP) makes it difficult to investigate many ILs for Machine Translation (MT). A data augmentation approach such as Backtranslation (BT) can be used to enhance the size of the training dataset. This paper presents the development of a NMT model for ILs within the context of a MT system. To address the issue of data scarcity, the paper examines the effectiveness of a BT approach for ILs that uses both monolingual and parallel datasets. Experimental results reveal that while the BT has improved the model's performance, however, it is not as significant as expected. It has also been observed that, even though the English-ILs and ILs-English models are trained on the same dataset, the ILs-English models perform better in all evaluation metrics. The reason for this is that ILs frequently differ in sentence structure, word order, and morphological richness from English. The paper also includes error analysis for translations between languages that were utilized in experiments utilizing the Multidimensional Quality Metrics (MQM) framework.

1 Introduction

An automated system that converts a source language into a target language is known as machine translation (Liu and Zhang, 2023; Liu Ming and Haffari, 2018). It has made significant strides recently in translating high-resource languages like Spanish, French, and English (Shaham et al., 2022). But as ILs present a unique combination of challenges and opportunities, it is still difficult to get a good translator.

This linguistic diversity, a testament to India's cultural heritage, poses distinct translation challenges when translating from English to ILs and vice versa. Despite their tremendous linguistic richness, ILs are characterized as low-resource due to a lack of training data available for language models [Das et al., 2024].

Compared to widely spoken languages such as English, 'low-resource languages' like ILs possess a restricted range of linguistic resources such as parallel corpora, dictionaries, grammar, and trained models (Das et al., 2022). In order to address the scarcity of resources, translation faces unique challenges, necessitating the utilization of efficient MT as a valuable tool (Cheragui, 2012). In fact, developing reliable and accurate MT systems for ILs is very challenging. In this regard, Backtranslation (BT) comes as an effective method for dealing with limited data and synthetically increasing the amount of data used for training for MT models (Behr, 2017). In different scenarios, NMT systems have shown to benefit from using BT, especially in most low-resource environments where it can be challenging to acquire high-quality corpora (Bala Das et al., 2023). Its potential to improve translation model performance in this linguistic domain is the driving force behind the investigation of its efficacy in the context of ILs. This motivates us to investigate backtranslation methods for ILs. In this paper, first, a baseline NMT model for English-ILs and ILs-English using Vaswani et al. [2017] transformer architecture is developed. Baseline models (NMT models which are generated) are trained using the Samanatar dataset [Ramesh et al., 2022] for experiments. The impact of the back translation for NMT models on ILs is examined. All the generated translation outputs are examined using evaluation metrics, and the generated model output's error analysis is also done.

The rest of this work is structured as follows:

Proceedings of the First Workshop on Natural Language Processing for Indo-Aryan and Dravidian Languages (IndoNLP2025), pages 33–46
January 20, 2025.

Section 2 contains a thorough overview of literature of NMT. In Section 3, a short description of the languages utilized is described. Section 4 discusses the model utilized for our experiment. Section 5 contains all results obtained after our experiments. In Section 6, we summaries our study and suggest some future research directives.

2 Literature Review

Sennrich et al. [2015] have introduced backtranslation, which is a process of creating synthetic parallel data by repeatedly converting monolingual data among source and target languages. This approach augments training data and improves the durability of NMT models. Building on the basic principles of backtranslation, a few researchers (eg. Marzieh and Monz, 2018; Edunov et al., 2018) have investigated various techniques and approaches for integrating it into NMT training pipelines. This method has demonstrated great promise in enhancing translation quality for few high resource languages. Numerous studies attest to the advantages of backtranslation.

Fadaee et al. [2017] and Xinlei et al. [2020] have delved into adapted methods incorporating backtranslation alongside NMT for European languages, offering helpful insights into tackling linguistic nuances unique to this region. Similarly, the effects of backtranslation on machine translation between Vietnamese and Chinese—two Asian languages with little linguistic affinity—are examined. Their study clarifies its efficacy in both SMT and NMT models by assessing various backtranslated corpus sizes. The results advance knowledge of how backtranslation improves translation quality for low-resource, less-related language pairs (Li et al., 2020). According to Currey et al. [2017], low-resource languages can also benefit from synthetic data if the source is only a duplicate of the target data, which is monolingual. Few researchers (eg. Cotterell and Kreutzer, 2018) frame backtranslation as a variational process with the latent space as the original sentences. According to them, there should be a match between the distribution of the artificial data generator and the actual translation probability. For this reason, it is crucial to understand and look into the sample distributions used by the most advanced data generation approaches which are available today. Ahmed et al. [2023] conducted a thorough investigation into iterative backtranslation for English-Assamese language pair and presented a simplified version of iterative backtranslation. Their findings demonstrated considerable improvements in BLEU scores: +6.38 for English-Assamese and +4.38 for Assamese-English.

3 Experimental Setup

This section describes the dataset, preprocessing method, steps before training, and evaluation metrics.

3.1 Dataset

The training dataset is taken from the Samanantar [Ramesh et al., 2022] and Flores200 dataset [Costa-jussà et al., 2022] is used for testing purposes to develop the NMT and the BT baseline models. The languages used for our experiments and their statistics are shown in Table 1. The dataset statistics show that Hindi has the highest parallel and monolingual dataset, while Assamese has the lowest (out of 11 languages).

Table 1: Dataset Statistics

English to Indic	Parallel Dataset	Monolingual Dataset
Tamil (TA)	5.16M	31.54M
Assamese (AS)	0.14M	1.38M
Marathi (MR)	3.32M	33.97M
Malayalam (ML)	5.85M	56.06M
Telugu (TE)	4.82M	47.87M
Bengali (BN)	8.52M	39.87M
Gujarati (GU)	3.05M	41.12M
Hindi (HI)	8.56M	63.05M
Kannada (KN)	4.07M	53.26M
Odia (OR)	1.00M	6.94M
Punjabi (PA)	2.42M	29.19M

3.2 Preprocessing

Several preprocessing techniques are used before translating from the source to the target languages.

1. Initially, from the dataset, several punctuation in the extended Unicode are converted to their standard counterparts.

2. Numbers in the ILs dataset are converted from the Latin script to the Devanagari script.

3. Characters outside the standard alphabets of the language pair are removed.

4. Unprintable characters are removed from the dataset, and the dataset is trimmed of extra white space.

5. Redundant quotation marks are removed from the dataset.

6. Sentences that are empty on any side of languages are eliminated.

7. To detect and eliminate repeated words from a dataset. For example, in the English dataset "Police has also started an investigation into the matter." translation in Hindi is साथ ही पुलिस (Police) "ने यह भी बताया कि मामले की जांच शुरू कर दी गई है." where the word police are written in Hindi and English. So, the word police in English is removed from the Hindi dataset.

3.3 Tokenization and Lowercasing

The dataset is then tokenized for further preprocessing. This creates tokens in the dataset separated by a single white space. The ILs and EN datasets are tokenized using a modified Moses tokenizer [Koehn, 2007]. Moses tokenizers are one of the most commonly used tokenizers in the English language. Hence, the modified Moses tokenizer is tailored for ILs. It effectively handles diacritics, including halants and nuktas. For example, in Bengali "২৮ বছর বয়সী ভিদাল ৩ বছর আগে সেভিয়া থেকে বার ্ সেলোনায় যোগদান করেছিল ।" is changed into "২৮ বছর বয়সী ভিদাল ৩ বছর আগে সেভিয়া থেকে বার্সেলোনায় যোগদান করেছিল ।".

3.4 Byte Pair Encoding (BPE)

Byte pair encoding is a form of tokenization in which the most common pair of consecutive bytes are combined with a byte not present in the data. The train and dev data are byte pair encoded using the trained byte pair encoder (BPE) [Sennrich et al., 2015]. BPE splits up the created tokens and subjects them to sub-word-based tokenization. This boosts the performance of the model and compresses the dataset, decreasing the training time for the model. BPE is carried out using subword-nmt. Subword-nmt is the decomposing of words into smaller, subword units, which is used to successfully tackle the problems created by rarely seen or out-of-vocabulary words in machine translation systems. Then, the next step is to create a dictionary.

3.5 Building dictionary and Binarization

A dictionary is built using the full dataset, which maps tokens to numbers that the computer can comprehend. The dictionary stores all mappings of words from the source and target language into numbers (indexes) that can be referenced by the model. The processed dataset is then binarized using fairseq before training. Binarization helps to load data and models faster by converting numbers to the sequence of binary numerals.

3.6 Training

The experiment uses the Vaswani Transformer model [Vaswani et al., 2017], which is implemented in the Fairseq library [Ott et al., 2019], an open-source sequence modeling toolkit that allows training models for machine translation tasks. The model comprises six encoder-decoder layers, each with 512 hidden units and multi-head attention, which are optimized using the Adam optimizer. Prior to being added to and normalized with the sub-layer input, each sub-layer output is subjected to a dropout value. All models utilized for our experiments use Flores200 test sets [Costa-jussà et al., 2022]. Our model is run on a high-performance workstation equipped with an Intel Xeon W-1290 CPU, with 10 physical cores and 20 threads (3.20 GHz base frequency, up to 5.20 GHz boost), providing robust multi-threading and caching with 20 MiB of L3 cache. The system includes 62 GB of RAM and an NVIDIA Quadro RTX 5000 GPU with 16 GB of VRAM, supported by driver version 535.154.05. The system uses CUDA 11.5 for compilation and is compatible with CUDA 12.2 for runtime operations, optimizing model training performance. The time to run each model is roughly half to two days, according to its dataset size. Fairseq library with Adam optimizer with betas of (0.9,0.98) for training is used. The initial learning rate reads 0.0005, and the inverse square root learning rate scheduler with 4000 warm-up updates has been used. The dropout probability has been set to 0.1, and the criterion is label-smoothed cross-entropy with a smoothing factor of 0.1. The model is trained up to 300,000 updates. A deliberate selection of 300000 updates is used in the experiment in light of the variety of languages in the dataset and the differing availability of data. This choice ensures that the model goes through more iterations during training, which improves its ability to adapt to the dataset's diverse linguistic traits. The goal is to improve the model's overall performance so that it can effectively handle the nuances of both low- and high-resource languages during the training process.

Once training is completed, the best checkpoint is loaded and used to generate a translation of the test dataset using the fairseq model. Lastly,

the transition quality is examined using evaluation metrics.

4 Methods used

4.1 Models with Original Data

The initial step is to develop a baseline model using the Neural Machine Translation (NMT) model with the Samanantar dataset for the English-11 ILs and vice versa.

4.1.1 Neural Machine Translation(NMT) System

Using NMT, in addition to adopting the probabilistic framework, it takes a data-driven approach to MT. It transforms the translation task into a probability distribution Wu et al. [2016]. With a parallel dataset, the NMT decreases the translation task to the probability distribution p of the target language b given the source language a, as shown in Equation 1.

$$p(trg \mid src; \alpha) = \prod_{k=1}^{m} p\left(trg \mid trg_{(k-1,\ldots,1)}, src; \alpha\right) \tag{1}$$

Here, $src = src_1 \ldots src_n$ is an input source language of n words, while $trg = trg_1 \ldots trg_m$ represents the translated sentence of m words. Here $n, m \neq 0$. α is the parameter to be learned, trg is the current word, and $trg_{(k-1,\ldots.1)}$ represents the previously created word.

4.2 Models with Backtranslated Data

An abundance of high-quality, diverse training data is a prerequisite for training machine translation models effectively. Unfortunately, many times it is difficult to obtain large parallel datasets that contain paired sentences in both the source and target languages. This limitation presents a significant challenge to achieving effective translation quality. However, monolingual corpora, made up of sentences only in the source language without translations, provide a readily available resource for exploring a variety of language styles and nuances. To tackle this issue, combining parallel and monolingual datasets is essential. To overcome data scarcity constraints, backtranslation emerges as a strategic augmentation method. It is a technique used to train NMT models.

The basic idea of backtranslation is to generate additional training data by alternately converting monolingual data through the source and target languages.

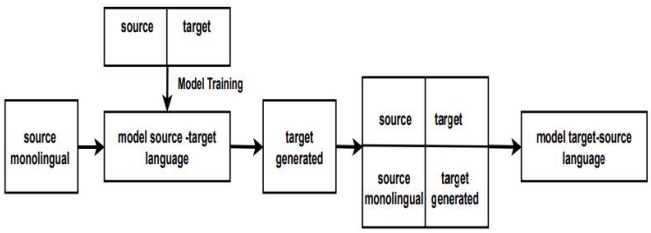

Figure 1: Process of Backtranslation

The process starts with training a model from source to target language using parallel data which generates synthetic target language data (from source monolingual data). The synthetic parallel data, which includes a combination of original and newly generated sentences(from source monolingual data), is utilized for training the NMT model from the target to the source language, as shown in Figure 1. This iterative approach improves the model's adaptability and efficiency of the NMT models by using the monolingual dataset, which leads to better translation quality. Our method using backtranslation is explained in Algorithm 1.

To examine the effect of pseudo data size (Here, pseudo data means the quantity or size of synthetic data produced during the backtranslation method) in an instance with limited resources, experiments are conducted with three datasets i.e. AS, ML, and HI. These languages are chosen according to their variation in the size of data, low resource, medium, and high resources concerning the dataset utilized. The varying proportion of the parallel corpus to pseudo data enabled the study of the impact of various pseudo corpus scales on model performance. It is observed while including more pseudo data, the positive impact on performance diminishes. The cause of this phenomenon is caused by the quality of pseudo data generated by the parallel corpus-trained translation model. Hence, after doing experiments with different data sizes, it is decided to add 2% of the pseudo dataset with the parallel dataset for Backtranslation purposes.

5 Results and Discussion

Table 2 displays the outcomes of our experiments using NMT and backtranslation by utilizing evaluation metrics such as BLEU [Papineni et al., 2002], TER [Wang et al., 2016], RIBES [Tan et al., 2015], METEOR [Banerjee and Lavie, 2005], chrF

Table 2: Evaluation Metrics for NMT and Backtranslation(where x indicate NMT and y indicate backtranslation)

Lang	Language Pairs	BLEU		TER		RIBES		METEOR		chrF		COMET	
		x	y	x	y	x	y	x	y	x	y	x	y
Odia	EN-OR	5.09	5.10	99.13	95.70	0.58	0.61	0.24	0.25	36.58	36.75	0.73	0.74
	OR-EN	10.92	11.75	84.95	87.28	0.59	0.61	0.38	0.41	39.27	42.07	0.75	0.76
Assamese	EN-AS	0.26	0.01	135.15	110.75	0.14	0.13	0.07	0.05	9.55	4.96	0.50	0.44
	AS-EN	0.77	0.67	178.79	123.65	0.18	0.13	0.17	0.17	20.50	16.29	0.54	0.50
Punjabi	EN-PA	19.16	20.09	73.58	71.48	0.74	0.75	0.48	0.49	48.53	49.40	0.81	0.82
	PA-EN	27.39	27.35	61.92	61.70	0.77	0.78	0.59	0.60	56.31	56.65	0.84	0.85
Gujarati	EN-GU	16.29	17.14	81.43	78.53	0.67	0.69	0.42	0.43	49.41	49.90	0.85	0.84
	GU-EN	23.75	23.82	70.05	68.58	0.72	0.73	0.57	0.56	55.30	54.37	0.84	0.85
Marathi	EN-MR	9.51	8.99	97.43	98.56	0.60	0.57	0.34	0.32	44.71	42.72	0.68	0.67
	MR-EN	19.37	19.38	73.42	75.13	0.70	0.69	0.51	0.52	50.21	50.56	0.81	0.82
Kannada	EN-KN	11.86	12.04	89.79	90.82	0.58	0.59	0.34	0.35	52.15	52.78	0.82	0.83
	KN-EN	21.31	20.84	74.33	73.29	0.71	0.72	0.53	0.54	52.92	52.52	0.82	0.83
Tamil	EN-TA	7.03	7.93	107.84	108.05	0.31	0.32	0.24	0.25	52.64	52.75	0.83	0.82
	TA-EN	20.99	22.38	74.36	70.97	0.71	0.72	0.53	0.55	52.32	52.50	0.81	0.83
Telgu	EN-TE	13.73	14.10	91.80	92.85	0.61	0.62	0.39	0.40	54.41	54.97	0.81	0.82
	TE-EN	24.52	25.06	68.95	70.78	0.73	0.74	0.56	0.59	55.36	57.09	0.82	0.83
Malayalam	EN-ML	8.12	8.14	106.52	103.27	0.44	0.45	0.29	0.28	52.90	52.91	0.82	0.83
	ML-EN	22.13	22.22	71.31	72.92	0.71	0.72	0.54	0.55	53.61	53.93	0.82	0.83
Bengali	EN-BN	16.02	16.99	74.90	72.04	0.71	0.72	0.41	0.43	52.15	53.50	0.84	0.85
	BN-EN	28.22	29.15	62.60	62.01	0.76	0.77	0.61	0.62	58.03	58.93	0.86	0.87
Hindi	EN-HI	31.41	29.77	57.82	60.38	0.78	0.77	0.56	0.54	56.60	55.32	0.79	0.78
	HI-EN	32.59	31.89	57.66	57.47	0.78	0.79	0.65	0.65	61.89	61.96	0.86	0.87

Algorithm 1 Pseudocode : Backtranslation

Require: language1-language2 parallel data, language2 monolingual data

Ensure: Trained model combining original and synthetically generated data

Data Collection:

1. language1-language2 parallel data, language2 monolingual data.

Training language2 -> language1 Model:

2. Train a model to translate from language2 to language1 with parallel data.

Backtranslation:

3. Use the trained language2 -> language1 model to translate monolingual language2 dataset to language1 dataset.

4. Combine the synthetic parallel corpus(translated language1 data with the original language2 monolingual data) with the original parallel corpus.

Model Training:

5. Train a new model for language1 to language2 using the newly combined data (generated data from step 4).

[Popović, 2015], and COMET [Rei et al., 2020] scores.

The performance metrics generated from NMT model, denoted by "X", and Backtranslation, denoted by "Y" are shown in Table 2. Using NMT, the BLEU score ranges between 0.26 to 32.59. RIBES and METEOR scores lie between 0.14 to 0.78 and 0.07 to 0.65. TER score varies between 57.66 to 178.79, whereas the chrF score ranges from 9.55 to 61.89, and the COMET score ranges from 0.50 to 0.86. In general, using NMT, it is noticed that the model performs better for ILs-English (in terms of evaluation metrics). This is likely due to the fact that English has relatively poor morphology in comparison with numerous ILs. It is also observed that the model-generating output for Hindi (HI), Punjabi (PN), and Bengali (BN) is good compared to other languages. The datasets of BN and HI languages are qualitative and less noisy; hence, they perform better than other languages. Similarly, due to its smaller dataset size, the model generating translation output for the Assamese(AS) language consistently performs poorly in various evaluation metrics. After analysis of the dataset, cases of inaccurate translations are found in the AS dataset, which adds to the lower evaluation scores. For example, in the

AS dataset, কিছুমান ইস্ৰায়েলে কেনে ধৰণৰ অন্যায়পূর্ণ কার্য্যত লিপ্ত আছিল? It is translated as "when it comes to speaking gods word, we will not disobey our god, even in lands where modern - day amaziahs are fomenting cruel persecution . " However, its translation using Google translator is " What kind of unjust things did some Israelites do?". Even the model-generated output for the Odia (OR) language performed poorly due to its smaller dataset size, which followed a pattern seen with the Assamese language. It also performed poorly due to its smaller dataset size, following a pattern seen with the Assamese language. As shown in Tables 2, a small improvement (in terms of evaluation metrics) is noticed across all the language pairs after backtranslation (with some exceptions such as AS, KN-EN, HI, EN-ML, KN-EN, and EN-MR). After the backtranslation method, the BLEU score ranges from 7.87 to 34.74 whereas RIBES and METEOR range from 0.19 to 0.42 and 0.58 to 0.76 respectively. TER scores vary between 61.7 to 123.65 and chrF scores lie between 4.93 to 61.89. COMET which offers a comprehensive evaluation toolkit, assigns scores using BT ranging from 0.50 to 0.85. The results demonstrate that the use of backtranslation has less impact and has not improved models with high BLEU score NMT baselines, for instance, the HI model has no improvement and it decreases the evaluation metrics. Backtranslation has shown a significant effect in languages such as Tamil where the EN-TE increases by 1.39 BLEU score. Indic languages are subject-object-verb (SOV) languages, whereas English is subject-verb-object (SVO), which means that word order frequently changes significantly. In backtranslation, synthetic Indic sentences derived from English may have an SVO structure that differs from natural Hindi constructions providing more "translationese". Dravidian languages such as Tamil, Telugu, Kannada, and Malayalam have rich agglutinative morphology, where word stems combine with extensive inflections and derivations. This is difficult for models with limited data to generalize, leading to issues in tense, aspect, modality, gender, and case generation when translating back and forth. The discrepancies between RIBES, METEOR, chrF, TER, METEOR, COMET, and BLEU are due to their focus on different aspects of language quality. An interesting finding from our backtranslation investigations is that Assamese, which performed poorly with NMT, performed even worse with backtranslation. Similarly, Hindi despite having a better result with NMT, failed to produce substantial improvements through backtranslation. TA, KN, TE, and ML are agglutinative, which means that words are often created through the combination of smaller units (morphemes) having particular meanings. Hence, these languages benefit from word formation while using BT because their learned patterns can be utilized continuously during backtranslation. However, in EN-ML, KN-EN a small decrease in evaluation metrics is noticed. The findings show a slight decrease in evaluations in some ILs when BT is used. Particularly, variations to this decrement exist, especially in translations from English to ILS. The limited effect may be caused by a number of factors, including the inherent characteristics of the language pair being translated, potential domain inconsistencies, and the quality and diversity of the

6 Conclusion

In this paper, a baseline NMT model on the Samanantar dataset utilizing transformer architecture is developed. In terms of BLEU, RIBES, METEOR, chrF, and COMET, Hindi excels when compared to other languages using NMT. From the result, it has been observed that the ILs-English NMT model outperforms and achieves higher BLEU scores than the English-ILs NMT model. For EN-IL translation using the NMT model, PA, GU, BN, and HI perform better than other languages while for IL-EN translation PA, TE, BN, and HI perform better than other languages. The paper also discusses and investigates the effectiveness of backtranslation (BT) for ILs and checks its performance in MT model. The results show that although BT enhanced the model's performance, however, this improvement was not as large as anticipated, and the model did not significantly outperform the baseline NMT models. One reason for the lack of noticeable improvements could be that the baseline NMT models' performance is subpar. An analysis of the experiment shows that while NMT models perform substantially better in some cases, they generally produce disappointing results over a wide range of languages. Even in these circumstances, their performance is below expectations. Since BT uses NMT models to produce data, its shortcomings affect its capacity to produce high-quality data. Another factor could be BT performs best when for experiments high-quality data

in both languages are available. However, even after filtration, the data obtained from experiments with ILs is not particularly clean or reliable. This means that the models that are used to create BT data aren't very good. Hence, there is not much effect of BT being noticed using ILs. In future work, our findings can be expanded by examining monolingual datasets of varying sizes and domains to precisely determine the different levels of saturation for backtranslation.

Limitation

The limitation of this work originates mainly from the scope and methodology of the back translation studies for 11 ILs. While this work gives helpful insights into enhancing translation quality, it does not cover all ILs, which limits the findings' generalizability. It is also observed that the size and quality of the original dataset were a problem, particularly for these ILs, since the results might have been impacted by noisy or inadequate data. Furthermore, computing constraints prevented the exploration of more advanced strategies, such as fine-tuning large-scale models for each language. Furthermore, these works only used backtranslation as a data augmentation strategy, leaving the potential for future research into complementing techniques such as multilingual pretraining or synthetic data production. These limitations identify potential areas for future research that could improve the technique and widen the scope of our work.

References

Mazida Akhtara Ahmed, Kishore Kashyap, Kuwali Talukdar, and Parvez Aziz Boruah. 2023. Iterative back translation revisited: An experimental investigation for low-resource English Assamese neural machine translation. In *Proceedings of the 20th International Conference on Natural Language Processing (ICON)*, pages 172–179.

Sudhansu Bala Das, Atharv Biradar, Tapas Kumar Mishra, and Bidyut Kr. Patra. 2023. Improving multilingual neural machine translation system for indic languages. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(6):1–24.

Satanjeev Banerjee and Alon Lavie. 2005. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, pages 65–72.

Dorothée Behr. 2017. Assessing the use of back translation: The shortcomings of back translation as a quality testing method. *International Journal of Social Research Methodology*, pages 573–584.

Mohamed Amine Cheragui. 2012. International conference on web and information technologies. In *Theoretical Overview of Machine Translation*, pages 160–169.

Marta R Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, et al. 2022. No language left behind: Scaling human-centered machine translation. *arXiv preprint arXiv:2207.04672*.

Ryan Cotterell and Julia Kreutzer. 2018. Explaining and generalizing back-translation through wake-sleep. *arXiv preprint arXiv:1806.04402*.

Anna Currey, Antonio Valerio Miceli Barone, and Kenneth Heafield. 2017. Copied monolingual data improves low-resource neural machine translation. In *Proceedings of the Second Conference on Machine Translation*, pages 148–156.

Sudhansu Bala Das, Atharv Biradar, Tapas Kumar Mishra, and Bidyut Kumar Patra. 2022. Nit rourkela machine translation (mt) system submission to wat 2022 for multiindicmt: An indic language multilingual shared task. In *The 9th Workshop on Asian Translation*.

Sudhansu Bala Das, Divyajyoti Panda, Tapas Kumar Mishra, Bidyut Kr Patra, and Asif Ekbal. 2024. Multilingual neural machine translation for indic to indic languages. *ACM Transactions on Asian and Low-Resource Language Information Processing*.

Sergey Edunov, Ott Myle, Auli Michael, and Grangier David. 2018. Understanding back-translation at scale. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics*, page 489–500.

Marzieh Fadaee, Bisazza Arianna, and Monz Christof. 2017. Data augmentation for low-resource neural machine translation. *arXiv preprint arXiv:1705.00440*.

P. Koehn. 2007. Moses: open source toolkit for statistical machine translation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 177–180.

H. Li, J. Sha, and C. Shi. 2020. Revisiting back-translation for low-resource machine translation between chinese and vietnamese. *IEEE Access*, pages 119931–119939.

Q. Liu and X Zhang. 2023. Machine translation: general. *In Routledge Encyclopedia of translation technology*.

Wray Buntine Liu Ming and Gholamreza Haffari. 2018. Learning to actively learn neural machine translation. In *In Proceedings of the 22nd Conference on Computational Natural Language Learning*, pages 334–344.

Marzieh and Christof Monz. 2018. Back-translation sampling by targeting difficult words in neural machine translation. *arXiv preprint arXiv:1808.09006*.

Myle Ott, Sergey Edunov, Alexei Baevski, Angela Fan, Sam Gross, Nathan Ng, David Grangier, and Michael Auli. 2019. fairseq: A fast, extensible toolkit for sequence modeling. *arXiv preprint arXiv:1904.01038*.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Maja Popović. 2015. chrf: character n-gram f-score for automatic mt evaluation. In *Proceedings of the tenth workshop on statistical machine translation*, pages 392–395.

Gowtham Ramesh, Sumanth Doddapaneni, Aravinth Bheemaraj, Mayank Jobanputra, Raghavan Ak, Ajitesh Sharma, Sujit Sahoo, Harshita Diddee, Divyanshu Kakwani, Navneet Kumar, et al. 2022. Samanantar: The largest publicly available parallel corpora collection for 11 indic languages. *Transactions of the Association for Computational Linguistics*, 10:145–162.

Ricardo Rei, Craig Stewart, Ana C Farinha, and Alon Lavie. 2020. Comet: A neural framework for mt evaluation. *arXiv preprint arXiv:2009.09025*.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2015. Neural machine translation of rare words with subword units. *arXiv preprint arXiv:1508.07909*.

Uri Shaham, Elbayad Maha, Goswami Vedanuj, Levy Omer, and Bhosale Shruti. 2022. Causes and cures for interference in multilingual translation. *arXiv preprint arXiv:2212.07530*.

Liling Tan, Jon Dehdari, and Josef van Genabith. 2015. An awkward disparity between bleu/ribes scores and human judgements in machine translation. In *Proceedings of the 2nd Workshop on Asian Translation (WAT2015)*, pages 74–81.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Weiyue Wang, Jan-Thorsten Peter, Hendrik Rosendahl, and Hermann Ney. 2016. Character: Translation edit rate on character level. In *In Proceedings of the First Conference on Machine Translation*, pages 505–510.

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.

Haoqi Fan Xinlei, Girshick Ross, and Kaiming He. 2020. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*.

A Appendix

The following section contains translation instances using both NMT and NMT with backtranslation models.

1. **English to Odia**

   English: He built a WiFi door bell, he said.
   Reference: ସେ ୱାଇଫାଇ କବାଟ ଘଣ୍ଟି ନିର୍ମାଣ କରିଥିବା ସେ କହିଛନ୍ତି
   Reference Transliteration: Se WiFi kabaata ghanti nirmana karithiba se kahichhanti.
   Reference Word-wise English: He WiFi door bell built has he said.
   Generated using NMT model:: ସେ ୱାଇ ଫାଇ ଡୋର୍ ବେଲ୍ ନିର୍ମାଣ କରିଥିବା କହିଛନ୍ତି।
   Transliteration: Se WiFi door bell nirmana karithiba kahichhanti.
   Word-wise English: He WiFi door bell built has said.
   Generated using Backtranslation model: ସେ କହିଛନ୍ତି ଯେ, ସେ ଗୋଟିଏ ୱାଇଫାଇ କବାଟ ବାଡେଇଛନ୍ତି ।
   Transliteration: Se kahichhanti je, se gotie WiFi kabaata bareichhanti.
   Word-wise English: He said that he a WiFi door has built.

   Odia to English

   Odia: ସେ ୱାଇଫାଇ ଡୋର୍ ବେଲ୍ ନିର୍ମାଣ କରିଥିବା କହିଛନ୍ତି।
   Transliteration: Se wifi door bel nirmaṇa karithiba kahichhanti.
   Word-wise English: He said that he has made a WiFi door bell.
   Reference: He built a WiFi door bell, he said.
   Generated using NMT model: he said he has constructed wifi.
   Generated using Backtranslation model:

he said that he had constructed the wifi.

2. **English to Assamese**

English: During his trip, Iwasaki ran into trouble on many occasions.
Reference: এই যাত্রাত বিভিন্ন সময়ত ইৱা-ছাকি বিপদত পৰিছিল।
Transliteration: Ei jatraat bibhinna somoyot Iwasaki bipadat porisil.
Word-wise English: This journey in various times Iwasaki trouble in faced.
Generated uisng NMT model: এই বিষয়ে পৰৱর্তী লেখত আলোচনা কৰা হ 'ব ।
Generated uisng Backtranslation model: এই যাত্রাত বিভিন্ন সময়ত ইৱাছাকি বিপদত পৰিছিল।
Transliteration: Ei bishoye poroborti lekhat alochona kora habo.
Word-wise English: This topic on next writing discuss done will be.

Assamese to English

Assamese: এই যাত্রাত বিভিন্ন সময়ত ইৱাছাকি বিপদত পৰিছিল।
Transliteration: eai jatraat eebivŭ somoyot iwasaik eebopodot pirisol
Word-wise English: This journey was in various times Iwasaki trouble in faced.
Reference: During his trip, Iwasaki ran into trouble on many occasions.
Generated using NMT model: this was followed by a few days ago.
Generated using Backtranslation model:he said that he had a fine example for his wife and his wife.

3. **English to Punjabi**

English: He built a WiFi door bell, he said.
Reference: ਉਸਨੇ ਕਿਹਾ, ਉਸਨੇ ਇੱਕ ਵਾਈ-ਫਾਈ ਡੋਰ ਬੈੱਲ ਬਣਾਈ ਹੈ।
Transliteration: Usne keha, usne ik WiFi door bell banayi hai.
Word-wise English: He said, he a WiFi door bell has made.
Generated using NMT model: ਉਨ੍ਹਾਂ ਕਿਹਾ ਕਿ ਉਨ੍ਹਾਂ ਨੇ ਵਾਈ-ਫਾਈ ਦੀ ਘੰਟੀ ਵਜਾਈ ।
Transliteration: Unha keha ki unha ne WiFi di ghanti vajaayi.
Word-wise English: They said that they WiFi's bell rang.
Generated using Backtranslation model: ਉਸਨੇ ਇੱਕ ਵਾਈ-ਫਾਈ ਡੋਰ ਘੰਟੀ ਬਣਾਈ, "ਉਸਨੇ ਕਿਹਾ ।
Transliteration: Usne ek WiFi door ghanti banayi, "usne keha."
Word-wise English: He a WiFi door bell made, "he said."

Punjabi to English

Punjabi: ਉਸਨੇ ਕਿਹਾ, ਉਸਨੇ ਇੱਕ ਵਾਈ-ਫਾਈ ਡੋਰ ਬੈੱਲ ਬਣਾਈ ਹੈ।
Transliteration: Usne keha, usne ik WiFi door bell banayi hai.
Word-wise English: He said, he a WiFi door bell has made.
Reference: He built a WiFi door bell, he said.
Generated using NMT model:he said, he has built a wi-fi door bell.
Generated using Backtranslation model: he has created a wi-fi door bell.

4. **English to Gujarati**

English: He built a WiFi door bell, he said.
Reference: તેમણે વાઈફાઈડોર બેલ બનાવ્યો હતો, એમ તેમણે કહ્યું હતું.
Transliteration: Temne WiFi door bell banavyo hato, em temne kahyu hato. **Word-wise English**: He WiFi door bell built was, he said was.
Generated using NMT model: તેમણે વાઇ-ફાઇ બારી બનાવી હતી. ।
Transliteration: Temne WiFi bari banavi hati.
Word-wise Translation: He WiFi window made had.
Generated using Backtranslation model: તેમણે વાઇફાઇ બારણું ઘંટનું નિર્માણ કર્યું હતું.
Transliteration: Temne WiFi baranu ghant nu nirmaan karyu hato.
Word-wise English: He WiFi door bell's construction did was.

Gujarati to English

Gujarati: તેમણે વાઈફાઈડોર બેલ બનાવ્યો હતો, એમ તેમણે કહ્યું હતું.
Transliteration: Temṇe vaiphāī ḍor bel banavyo hato, em temṇe kahyu hutu.
Word-wise English: He built a Wi-Fi doorbell, he said.
Reference: He built a WiFi door bell, he said.
Generated using NMT model:he had built the wimbledon bell, he said.
Generated using Backtranslation model:he built a wi-fi bell," "he said.".

5. **English to Marathi**

English: He built a WiFi door bell, he said.
Reference: ते म्हणाले की, त्यांनी WiFi डोअर बेल बनवली आहे.
Transliteration: Te mhanale ki, tyanni WiFi door bell banvali aahe.
Word-wise Translation: He said that he WiFi door bell made has.
Generated using NMT model: त्यांनी वाय-फाय दाराची बेल तयार केली.
Transliteration: Tyanni WiFi darachi bell tayar keli.
Word-wise Translation: He WiFi door's bell prepared did.
Generated using Backtranslation model: त्यांनी वाय-फाय दारावरची बेल बनवली.
Transliteration: Tyanni WiFi daravarachi bell banavali.
Word-wise Translation: He WiFi door-on bell made.

Marathi to English

Marathi: ते म्हणाले की, त्यांनी WiFi डोअर बेल बनवली आहे.
Transliteration: tem haṇale ka, ta yanni WiFi ḍoar bel banavali ahe
Word-wise Translation: They said that they have made a WiFi doorbell.
Reference: He built a WiFi door bell, he said.
Generated using NMT model: they have created wifi dover bell," "he said."
Generated using Backtranslation model: """ "he has made wifi pie bell," "he said."

6. **English to Kannada**

English: He built a WiFi door bell, he said.
Reference: ಅವರು ವೈಫೈ ಡೋರ್ ಬೆಲ್ ತಯಾರಿಸಿದ್ದಾರೆ ಎಂದು ಅವರು ಹೇಳಿದರು.
Transliteration: Avaru WiFi dvaarada ghante nirmisidare endu heLidaru.
Word-wise Translation: He WiFi door bell built has, he said.
Generated using NMT model: ಅವರು ವೈಫೈ ಡೋರ್ಬೆಲ್‌ಅನ್ನು ನಿರ್ಮಿಸಿದ್ದರು.
Transliteration: Avaru WiFi kada tayarisidaru.
Word-wise Translation: He WiFi door prepared.
Generated using Backtranslation model:"" "ಅವರು ವೈಫೈ ಬಾಗಿಲು ನಿರ್ಮಿಸಿದರು" "ಎಂದು ಅವರು ಹೇಳಿದರು."
Transliteration: Avaru WiFi dvaarada ghante kattidaru endu heLidaru.
Word-wise English: He WiFi door's bell built has, he said.

Kannada to English
Kannada :ಅವರು ವೈಫೈ ಡೋರ್ ಬೆಲ್ ತಯಾರಿಸಿದ್ದಾರೆ ಎಂದು ಅವರು ಹೇಳಿದರು.
Transliteration: Avaru vaiphai dor bel tayarisiddare endu avaru heḷidaru.
Word-wise English: They have made a Wi-Fi doorbell, they said.
Reference: He built a WiFi door bell, he said.
Generated using NMT model: """"they have made a wi-fi door bell."
Generated using Backtranslation model: he said that they have made wi-fi dorm.

7. **English to Tamil**

English: He built a WiFi door bell, he said.
Reference:அவர், தான் வைஃபை கதவு அறிவிப்பு மணியை உருவாக்கியதாகக் கூறினார்.
Transliteration: Avar WiFi kadhavu mani amaithadhaga avar sonnaar.

Word-wise English: He WiFi door bell made as he said.
Generated using NMT model: "" "அவர் ஒரு வைஃபை கதவு மணியை கட்டினார்."
Transliteration: Avar WiFi kadhavai amaithaar.
Word-wise English: He WiFi door made.
Generated using Backtranslation model: அவர்வைஃபை கதவை கட்டினார்.
Transliteration: Avar WiFi kadhavu maniyai amaithadhaga sonnaar.
Word-wise English: He WiFi door bell built said.

Tamil to English
Tamil: அவர், தான் வைஃபை கதவு அறிவிப்பு மணியை உருவாக்கிய தாகக் கூறினார்.
Transliteration: Avar, tan vai-fai kathavu arivippu maniyi uruvakkiyadag kuri□ar.
Word-wise translation: He, he WiFi door bell built said.
Reference: He built a WiFi door bell, he said.
Generated using NMT model: he said he created the wifi door bell.
Generated using Backtranslation model:he said he created the wi-fi doors.

8. **English to Telugu**

English: He built a WiFi door bell, he said.
Reference: అతను WiFi డోర్ బెల్ నిర్మించాడు. అని చెప్పాడు.
Transliteration: Athanu WiFi door bell nirminchadu. Ani cheppadu.
Word-wise translation: He WiFi door bell built. That said.
Generated using NMT model: Wi-Fi డోర్బెల్ ను నిర్మించినట్లు తెలిపారు..
Transliteration: Wi-Fi doorbell nu nirminchinatlu teliparu.
Word-wise translation: Wi-Fi doorbell that built informed.
Generated using Backtranslation model: వైఫై డోర్బెల్నిర్మించానని తెలిపారు.

Telugu to English
Telugu : అతను WiFi డోర్ బెల్ నిర్మించాడు. అని చెప్పాడు.
Transliteration: Atanu WiFi ḍōr bel nirmaiṃcāḍu ani ceppāḍu.
Word-wise translation: He WiFi door bell built said.
Reference: He built a WiFi door bell, he said.
Generated using NMT model: "he built the wifi door bell." ""
Generated using Backtranslation model: he said he built the wifi door bell.

9. **English to Malayalam**
English: He built a WiFi door bell, he said.
Reference: അദ്ദേഹം ഒരു WiFi ഡോർ ബെൽ ഉണ്ടാക്കിയെന്ന് അവൻ പറഞ്ഞു.
Transliteration: Ayaal WiFi kavaadamani nirmichu, ennu paranju.
Word-wise English: He WiFi door bell built, said.
Generated using NMT model: അദ്ദേഹം ഒരു വൈഫൈ ഡോർ ബെൽ നിർമ്മിച്ചു.
Transliteration: Ayaal WiFi kavaadam panithu.
Word-wise English: He WiFi door built.
Generated using Backtranslation model: അദ്ദേഹം ഒരു വൈഫൈ ഡോർ ബെൽ നിർമ്മിച്ചു, "അദ്ദേഹം പറഞ്ഞു.
Transliteration: Ayaal WiFi kavaadamani nirmichu ennu paranju.
Word-wise English: He WiFi door bell built, said.

Malayalam to English
Malayalam:അദ്ദേഹം ഒരു WiFi ഡോർ ബെൽ ഉണ്ടാക്കിയെന്ന് അവൻ പറഞ്ഞു.
Transliteration: Addeham oru WiFi ḍōr bel uṇṭākkiyennu avan pa□aññu.
Word-wise translation: He a WiFi door bell made said.
Reference: He built a WiFi door bell, he said.
Generated using NMT model: He built a WiFi door bell, he said.
Generated using Backtranslation model:he said he made a wifi door.

10. **English to Bengali**

English: He built a WiFi door bell, he said.
Reference: তিনি জানান যে তিনি একটি ওয়াই-ফাই ডোর বেল তৈরি করেছিলেন।
Transliteration: Tini bolechilen je tini ekṭi WiFi dorjar ghonta toiri korechilen.
Word-wise Translation: He said that he a WiFi door's bell created had.
Generated using NMT model: তিনি বলেন, " ওয়াই-ফাই দরজায় ঘণ্টা বাজিয়ে ।
Transliteration: Tini bolechen tini ekṭi WiFi dorja toiri korechen.
Word-wise Translation: He said he a WiFi door created has.
Generated using Backtranslation model: তিনি একটি ওয়াইফাই দরজা বেল তৈরি করেছিলেন ।
Transliteration: Tini bolechilen je tini ekṭi WiFi ghontar bebostha korechilen.
Word-wise Translation: He said that he a WiFi bell's arrangement had done.

Bengali to English
Bengali: তিনি জানান যে তিনি একটি ওয়াই-ফাই ডোর বেল তৈরি করেছিলেন।
Transliteration: Tini jānan je tini ēkaṭi wāi-phāi dōr bēl tairī karēchēlēn.
Word-wise translation: He informs that he a WiFi door bell made has.
Reference: He built a WiFi door bell, he said.
Generated using NMT model: He built a WiFi door bell, he said.
Generated using Backtranslation model:he said he had created a wi-fi door bell.

11. **English to Hindi**

English: He built a WiFi door bell, he said.
Reference: उन्होंने (जेमी सिमिनॉफ़) कहा, कि उन्होंने एक वाईफ़ाई डोर बेल बनाई है.
Transliteration: Unhone (Jamie Siminoff) kaha, ki unhone ek WiFi door bell banai hai.
Word-wise translation: He (Jamie Siminoff) said, that he a WiFi door bell made is.
Generated using NMT model: उन्होंने कहा कि उन्होंने एक वाईफाई डोर बेल बनाई है
Transliteration: Unhone kaha ki unhone ek WiFi door bell banai hai.
Word-wise translation: He said that he a WiFi door bell made is.
Generated using Backtranslation model: उन्होंने कहा कि एक वाईफाई डोर बेल बनाई है ।
Transliteration: Unhone kaha ki ek WiFi door bell banai hai.
Word-wise translation: He said that a WiFi door bell made is.

Hindi to English
Hindi: उन्होंने (जेमी सिमिनॉफ़) कहा, कि उन्होंने एक वाईफ़ाई डोर बेल बनाई है.
Reference: He built a WiFi door bell, he said.
Transliteration: Unhōne (Jamie Siminoff) kahā, ki unhōne ek WiFi dor bel banāi hai.
Word-wise translation: They (Jamie Siminoff) said, that they a WiFi door bell made have.
Generated using NMT model: he (jamie siminouf) said he has made a wifi door bell.
Generated using Backtranslation model: he (jamie siminouf) said he made a wifi door bell.

B Error Analysis

All the generated translations are categorized and analyzed into **Multidimensional Quality Metrics (MQM)**[1] based error analysis categories. Different categories of error are analyzed based on their accuracy, fluency, and mistranslations that impact the translation quality. For example, while translating of **English to Odia** language, it has been observed that the NMT model generated translation for "He built a WiFi door bell, he said" has a minor error. It translates as "he built a WiFi doorbell, he said," but uses "ବେଲ୍" (bel) (transliteration for bell) rather than the native Odia "ଘଣ୍ଟି" ("bell"). The translation generated by NMT result is more consistent than the backtranslation model, though both exhibit jarring translations. While the translation of the NMT model is simpler, errors still remain due to inaccurate word choices. Similarly, the term "doorbell" is missing from both translations when analyzing the error for the **Odia to English** translation generated using the nmt and backtranslation models. This results in a significant meaning error as the intended object is inaccurately converted to "WiFi," distorting the translation. This type of error falls under the category of 'Omission' under the MQM framework. Despite

[1] MQM, Error types: Typology, n.d., accessed: 2024-10-31. [Online]. Available: https://themqm.org/error-types-2/typology/

the fact that both outputs are acceptable in English, the omission error hinders readability and clarity of meaning. By leaving out the word "door bell", the translations lose important context, changing how the subject's action is interpreted and introducing incomplete understanding. Similarly, while analyzing the **English to Assamese** translation utilizing NMT models, the statement in English, "During his trip, Iwasaki ran into trouble on many occasions," is incorrectly translated into Assamese as, "This will be discussed in the next article," which is unrelated. This is a serious accuracy error under the MQM framework that totally obscures the meaning. However, while using the back translation model, it is able to translate the sentence, probably because of its closer grammatical structure as well as vocabulary compatibility. In this case, NMT's fluency is low since it produces a sentence that is wholly unrelated to the input, while the backtranslation is fluent and accurately reflects the reference text. For **Assamese to English** translation, the NMT model erroneously generates a timeline-based sentence, "this was followed by a few days ago," which does not accurately portray the intended narrative of difficulties. The translation generated using backtranslation model is entirely incomprehensible, implying unrelated parts such as "fine example for his wife," which have no resemblance to the original. This type of error falls under the category of mistranslation, accuracy, and incoherence. This translation generated from the models contains serious mistranslation errors that completely change the meaning of the text, rendering both outputs unintelligible to the intended reader.

For the case of **English to Punjabi** language translation, NMT model renders "He built a WiFi doorbell, he said" as "he said he played a Wi-Fi bell," which is inaccurate since "built" is mistaken for "played." Nevertheless, the backtranslation model, which yields "he built a WiFi doorbell, he said," is more accurate, despite a few small grammatical errors. Both translations lacked natural flow. In Punjabi, precise terminology would better indicate construction ("ਬਣਾਇਆ") ("Baṇa'i'a") rather than ("ਵਜਾਈ") (Vaja'i). Translation generated from backtranslation model is easier to read. Both translations lacked natural flow. In Punjabi, precise terminology would better indicate construction ("ਬਣਾਇਆ") (Baṇa'i'a) rather than ("ਵਜਾਈ") ("Vaja'i"). A backtranslated statement is easier to read. Meanwhile, for **Punjabi to English** language translation, both translations effectively convey the majority of the original content. However, since the backtranslation omits the original speaker tag, there is a small amount of ambiguity, and the structure lacks consistency. The absence of "he said" makes the sentence appear incomplete in terms of dialogue or quote structure. This type of error falls under the category of 'Omission' and 'Fluency' under the MQM framework. Minor challenges hinder the overall effectiveness of backtranslation, although the meaning is primarily maintained in both models.

From **English to Gujarati** translation, the NMT model interprets "doorbell" as "Wimbledon bell," which is a severe accuracy issue. This issue could be due to an uncertain vocabulary corpus in English-Gujarati translations. However, the translation generated by the back translation produces output closer to the desired meaning, but it contains redundancy, such as "constructed," which reduces the clarity. Similarly, for the translation of **Gujarati to English** using the NMT model, "Wimbledon bell" is an incorrect translation for "WiFi door bell," most likely owing to phonetic or contextual confusion, resulting in a significant terminology issue. While the backtranslation model almost catches the original meaning, there is a punctuation issue with the quotation marks, causing some uncertainty. The NMT model's translation significantly misrepresents the crucial term, resulting in confusion. The backtranslation output is more accurate, with minimal punctuation and fluency mistakes. This type of error falls under the 'mistranslation' and 'Fluency' categories under the MQM framework.

Similarly, for translating **English to Marathi** language, the NMT translation, "they have invented wifi Dover bell, he remarked," transforms the "WiFi doorbell" to "Dover bell," resulting in an accuracy issue. Backtranslation, on the other hand, retains the term "doorbell," despite slight difficulties with clarity and contextual accuracy. It has been noticed that NMT has reduced fluency due to the arbitrary addition of "Dover," whereas backtranslation gives somewhat enhanced fluency. The fundamental vocabulary problems cause misinterpretation, and punctuation further complicates intelligibility. Similarly, for translation of **Marathi to English**, both models misinterpret "door bell" as "dover bell" or "pie bell," representing significant terminology errors. Additionally, both translations exhibit punctuation issues with quotation

marks, creating readability issues. This type of error falls under the 'mistranslation' and 'Terminology' categories under the MQM framework. The major vocabulary problems cause misinterpretation, and punctuation further reduces clarity.

In the case of translation from **English to Kannada** translation, the NMT model's translation "they have made Wi-Fi bell, he said" comprises an accuracy concern, since it fails to indicate that the bell is built and functional. The backtranslation is also imprecise. Both outputs contain awkward language, which reduces overall fluency. The use of the appropriate Kannada phrase for "WiFi" would improve readability. When translating from **Kannada to English** sentence using the Backtranslation methodology, the word "dorm" is misused, changing its meaning to imply something quite unrelated. While the translation generated from NMT model is more precise, the absence of initial topic background diminishes precision. The translation output generated from the backtranslation model deviates from the meaning of the source language, whereas the NMT model is more accurate but might benefit from improved consistency. This type of error falls under the 'fluency' and 'terminology' category under the MQM framework.

However, in the case of **English to Telugu** translation, NMT and backtranslation both handle the word "doorbell" inconsistently. While the backtranslation slightly improves the clarity, NMT creates errors, such as interpreting it as "doarbell." However, the translation generated from NMT models is slightly awkward but understandable, while the back translation is marginally better in readability. Similarly, while translating from **Telugu to English** language translation, the NMT model accurately translates "WiFi door bell" and provides the entire concept with clarity and structure. The translation generated from the backtranslation model, such as others, omits the "door," which slightly alters the object's specificity. This type of error falls under the 'Omission ' and 'Accuracy' category under the MQM framework. The backtranslation model loses some specificity by omitting off "door," whereas the NMT approach produces a clear and precise translation.

While analysis of **English to Malayalam** translation, NMT clearly translates the statement with small variations, such as changing "doorbell" to "door ring." Backtranslation creates ambiguity by misinterpreting "WiFi door." The translation generated from NMT models accurately translates the statement with slight modifications, such as changing "doorbell" to "door ring." Backtranslation causes uncertainty by misinterpreting "WiFi door." The NMT methodology generates more fluent text, whereas backtranslation introduces some ambiguity by misinterpreting "WiFi door.". For **Malayalam to English** translation, The translation generated from the NMT model captures the entire translation accurately, maintaining the terminology "WiFi door bell" correctly. However, a common problem seen in translation generated from the backtranslation model leaves out "door" from "WiFi door bell," which somewhat reduces specificity. The backtranslation's omission of "door" reduces the clarity. With the NMT model, correct translation is provided. Hence, this type of error falls under the 'Omission ' and 'Accuracy' category under the MQM framework.

Similarly, while translating from **English to Bengali** sentence, the translation generated from the NMT and back translation model provides correct words; however, the NMT model incorrectly translates "doorbell" as "door knocker." The NMT translation is more consistent in fluency than the backtranslation, which has minor grammatical issues. Likewise, for **Bengali to English** translation, the NMT model accurately captures the meaning of "WiFi door bell" while still keeping the quote's context. Similarly, backtranslation , the word "door" is omitted, resulting in a slight loss of clarity and object specificity. Hence, this type of error falls under the 'Omission' and 'Accuracy' category under the MQM framework. The backtranslation model includes a slight omission, whereas the NMT model accurately represents the source text.

For translation of **English to Hindi** language, the NMT and backtranslation methods produce similar sentences that accurately preserve the meaning, using the Hindi term **"बनाई है"**. Both the translations generated express the speaker's intent. Fluency is strong in both models, with NMT having a minor advantage due to its consistent phrasing. However, while translating **Hindi to English** sentence, the NMT model correctly captures the message and uses the crucial terminology "WiFi doorbell" while keeping the main context. The backtranslation model omits the word "door" in "WiFi door bell," resulting in a modest omission and loss of detail. Both translations are mostly correct, but the backtranslation model's omission of the word "door" diminishes specificity.