


ARTICLE

Statistical machine translation for Indic languages

Sudhansu Bala Das¹ , Divyajyoti Panda¹, Tapas Kumar Mishra¹ and Bidyut Kr. Patra²

¹National Institute of Technology (NIT), Rourkela, Odisha, India and ²Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh, India

Corresponding author: Sudhansu Bala Das; Email: baladas.sudhansu@gmail.com

(Received 25 December 2022; revised 4 March 2024; accepted 26 March 2024; first published online 3 June 2024)

Special Issue on ‘Natural Language Processing Applications for Low-Resource Languages’

Abstract

Statistical Machine Translation (SMT) systems use various probabilistic and statistical Natural Language Processing (NLP) methods to automatically translate from one language to another language while retaining the originality of the context. This paper aims to discuss the development of bilingual SMT models for translating English into fifteen low-resource Indic languages (ILs) and vice versa. The process to build the SMT model is described and explained using a workflow diagram. Samanantar and OPUS corpus are utilized for training, and Flores200 corpus is used for fine-tuning and testing purposes. The paper also highlights various preprocessing methods used to deal with corpus noise. The Moses open-source SMT toolkit is being investigated for the system’s development. The impact of distance-based reordering and Morpho-syntactic Descriptor Bidirectional Finite-State Encoder (msd-bidirectional-fe) reordering on ILs is compared in the paper. This paper provides a comparison of SMT models with Neural Machine Translation (NMT) for ILs. All the experiments assess the translation quality using standard metrics such as BiLingual Evaluation Understudy, Rank-based Intuitive Bilingual Evaluation Score, Translation Edit Rate, and Metric for Evaluation of Translation with Explicit Ordering. From the result, it is observed that msd-bidirectional-fe reordering performs better than the distance-based reordering model for ILs. It is also noticed that even though the IL-English and English-IL systems are trained using the same corpus, the former performs better for all the evaluation metrics. The comparison between SMT and NMT shows that across various languages, SMT performs better in some cases, while NMT outperforms in others.

Keywords: Machine translation; low-resource languages; evaluation metrics; Statistical Machine Translation; Indic languages

1. Introduction

Technology reaches new heights through its journey from the origins of ideas to their full-scale practical implementation. One such journey is heading toward the elimination of language barriers in order to establish seamless social communication in every domain.

In this context, advances in related fields such as Natural Language Processing (NLP), Artificial Intelligence-based Language Modeling (LM), and Machine Learning contribute significantly to the development of a faultless automatic machine translation (MT) system. Regardless of the variety of heuristic methods used to preserve both contextual and lexical interpretations, obtaining sufficient adequacy and fluency of the translation model (TM) remains difficult. It requires a large high-quality parallel corpus (translation pairs in source and target languages) (Chapelle and Chung 2010). Hence, MT systems prove to be quite efficient with adequate training for high-resource languages, that is, languages with a large quantity of corpora available and having massive

digital footprint across the globe (Dorr, Hovy, and Levin 2004). On the contrary, it becomes very complicated for low-resource languages suffering from limited recognition and scanty digital presence. Such imbalance often leads to poor-quality translation. Therefore, MT systems need to understand the syntax (rules to combine words), semantics (meaning of words and combinations), and morphology (rules to cover morphemes—smallest meaningful units—into words) of such low-resource languages adequately (Somers 2011).

Statistical MT (SMT) is one of the popular methods proposed to solve these problems. It is a way of translation wherein a statistical-based learning algorithm is applied to a large bilingual corpus that helps the machine learn the translation. This method also enables the machine to translate sentences not encountered by the machine during its training and testing. Unlike rule-based and example-based MT approaches, SMT does not require human intervention, and it can build a translation system directly from training data (Lopez 2008).

On the other hand, Neural MT (NMT) is performed using a neural network (NN) (Stasimioti 2020). In contrast to SMT, NMT does not need any TMs, language models, or reordering models. Instead, it makes use of a single sequence model to produce words one at a time. On the basis of the source sentence and the previously formed sequence in the target language, predictions are made. NMT is a deep learning-based technique for MT that operates using word vector representations and a massive NN.

Whether NMT can replace SMT is still debated, with sufficient evidence supporting both sides. Eventually, the experiment of Ziemski, Junczys-Dowmunt, and Pouliquen (2016) on the corpus of the United Nations (consisting of fifteen low-resource languages) emphasizes the fact that SMT and NMT can excel in different situations. From the result of his experiment, it is evident that the performance of SMT is better than that of NMT for the majority of cases, as measured by the BiLingual Evaluation Understudy (BLEU) score. In comparison with SMT, NMT training typically takes a longer time. Research has also shown that when there is a domain incompatibility between testing and training data, SMT performance is superior to that of NMT (Wang, Tu, and Zhang 2018; Mahata *et al.* 2018). Long sentences are another area where SMT excels (Toral and Sánchez-Cartagena 2017). Moreover, many researchers (Zhou *et al.* 2017; Wang *et al.* 2017; Castilho *et al.* 2017; Lohar *et al.* 2019) have pointed out various disadvantages of NMT over SMT on low-resource languages, like Indic languages (ILs), such as the fact that NMT requires more corpus and resources than SMT.

English (EN) and ILs are languages with less parallel data, which motivates us to work with ILs. This research examines the effectiveness of SMT systems on low-resource language pairs, of which many are rarely worked on. The corpus used in the experiment for all fifteen ILs is tested for the first time for building SMT models. These languages are Assamese (AS), Malayalam (ML), Bengali (BN), Marathi (MR), Gujarati (GU), Kannada (KN), Hindi (HI), Oriya (OR), Punjabi (PA), Telugu (TE), Sindhi (SD), Sinhala (SI), Nepali (NE), Tamil (TA), and Urdu (UR).

Our main goal is to develop SMT model for low-resource languages, that is, ILs, which can serve as a baseline system. The following are the summary of our work's main contributions:

- This work is the first attempt to build an SMT model with the Samanantar and OPUS Dataset for fifteen IL-EN and EN-IL pairs (both directions), including the Indo-Aryan and Dravidian groups. A realistic assessment of translation quality is evaluated using different evaluation metrics such as BLEU, Metric for Evaluation of Translation with Explicit Ordering (METEOR), Translation Edit Rate (TER), and Rank-based Intuitive Bilingual Evaluation Score (RIBES).
- Various data filtration methods are investigated in order to clean the data and improve translation quality.

- A comparison of distance and Morpho-syntactic Descriptor Bidirectional Finite-State Encoder (msd-bidirectional-fe) reordering models is examined to determine the optimal configuration of components in the TMs.
- SMT models are compared to NMT models to determine how well they perform in translation tasks.

This paper is arranged as follows. Subsection 1.1 gives insight into SMT, while Subsections 1.2 and 1.3 elucidate alignment and reordering models, respectively. In Section 2, prominent works on SMT and NMT using ILs are described. The experimental framework, including an overview of the dataset and methodology with the workflow of the SMT model, is explained in Section 3. Results are presented in Section 4 followed by the conclusion and future work in Section 5.

1.1 SMT

SMT is dependent on statistical methods (Koehn *et al.* 2007; Zens, Och, and Ney 2002; Hearne and Way 2011). It is a data-driven technique that makes use of parallel-aligned corpora. It utilizes Bayesian theory to maximize the likelihood of source-to-target language translation ($P(Tl|Si)$, where Tl is the target language and Si is the source input) using equation (1):

$$P(Tl | Si) \propto P(Tl)P(Si | Tl) \quad (1)$$

SMT is divided into three stages: the LM $P(Tl)$, the TM $P(Si|Tl)$, and the decoder model (DM) (Kumawat and Chandra 2014).

The LM relies on the n -gram model to compute the probability of a sentence. It assigns the probability of a single token to the last n tokens that come before it in the sentence and estimates the likelihood of the token. It assumes that the likelihood is independent of any tokens before the last n tokens before it. The chain rule aids in breaking down the sentence into conditional probability products.

$$\begin{aligned} P(s) &= P(w_1, w_2, w_3, \dots, w_n) \\ &= P(w_1)P(w_2|w_1)P(w_3|w_1w_2)P(w_4|w_1w_2w_3) \dots P(w_n|w_1w_2 \dots w_{n-1}) \\ &= P(w_1)P(w_2|w_1)P(w_3|w_1w_2)P(w_4|w_1w_2w_3) \dots P(w_n|w_1w_2 \dots w_{n-k}) \end{aligned} \quad (2)$$

where $P(s)$ is the probability of the sentence s , consisting of tokens w_1, w_2, \dots, w_n , assuming a k -gram model. It utilizes the bilingual parallel corpus of the desired language pair. This is accomplished by calculating the likelihood of tokens extracted from sentences.

The TM is the second phase of SMT, which computes the conditional probability $P(Si | Tl)$, trained from the corpus. The DM is the final and most crucial phase of SMT. It assists in the selection of words with the highest probability to be translated by maximizing the likelihood, that is, $P(Tl)P(Si | Tl)$.

1.2 Alignment model

In MT, an alignment model is used to determine the relationship between phrases or words in the original language and the language being translated. It is important for SMT systems as they help to determine which phrases or words in the source text are associated with particular words or phrases in the destination language.

Giza++ is one such library that is used for alignment. It has different alignment methods such as intersection grow, grow-diag, union, srctotgt, and tgttosrc. Girgzdís *et al.* (2014) used the Gold Standard as well as the first one million sentences from the DGT-TM 2013 corpus (Steinberger *et al.* 2013) and investigated the effect of these alignment models on them. The models were evaluated using the BLEU score. Among all of the methods examined, the “grow-diag-final-and”

method received the highest BLEU score, indicating its outstanding efficiency in precisely aligning words and improving translation quality.

The grow-diag-final-and model starts with the intersection of the alignments from source to target and target to source, then two steps are used to add additional alignment points (Costa-jussà *et al.* 2003) :

grow-diag: For every neighboring point to the alignments measured, if either source or target word is not aligned already but is present in the union of the alignment, then the neighboring point is included in the alignment.

final: If any phrase pairs are unaligned but present in the union, add the point to the alignment.

It assesses the likelihood of word-to-word alignment for each source and target word in each sentence. To produce a good-quality word alignment, the alignment is produced using a series of successive estimations. The alignment method establishes a connection between the target and source words. This step is followed by reordering.

1.3 Reordering model

Reordering is the process of restructuring the word order of one natural language sentence to make it more similar to the word order of another natural language sentence. It is a critical task in the translation of languages with different syntactic structures. During training, the Moses system determines various reordering possibilities with every phrase. Some of the reordering strategies are msd-bidirectional-fe and distance-based reordering.

1.3.1 msd-bidirectional-fe reordering

msd-bidirectional-fe reordering is an approach utilized during translation to improve the order of words in the target language sentence. It is an improved approach that considers the place of each word within the sentence. It utilizes a set of rules to detect same-side dependencies and bidirectional reordering inside a fixed range, allowing for easier reordering while preserving translation coherence and fluency. This method is especially useful for ILs, which possess complex word orders as a result of their rich morphological and syntactic structure. The advantage of this reordering is that it can provide more accurate and fluent translations by identifying the dependencies within words in both left-to-right and right-to-left directions.

1.3.2 Distance based reordering

Distance reordering is a more straightforward method that reorders words based on their location in the source sentence using a fixed set of rules. It needs to consider the sentence's syntactic structure or the specific dependencies within words, which otherwise would result in less accurate translations. It assigns a linear cost to reordering distance, implying that the movement of phrases over long distances is more expensive (Kumawat and Chandra 2014). The advantage of distance-based reordering is that it is computationally effective as it distributes reordering costs according to relative distances instead of creating complex structures. However, it may not be as good at capturing long-term interdependence or crucial word order changes as the bidirectional model.

2. Related work

In the year 2009, EN-to-HI SMT system has been created by (Ramanathan *et al.* 2009) using morphological and syntactic preprocessing in the SMT model. In their work, the suffixes in HI language are segmented for morphological processing before rearranging the EN source sentences as per HI syntax.

In 2010, research has been conducted by Zbib *et al.* (2010) at MIT, USA, using the grammatical structures in SMT with the Newswire corpus for Arabic to EN language to give better translation results.

Work on KN-to-EN SMT system by Shiva Kumar, Namitha, and Nithya (2015) shows a remarkable feat with 14.5 BLEU score on using Bible corpus on 20,000 sentences. The results are supported by Papineni *et al.* (2002).

Kaur and Josan (2011) presented a TM based on SMT for EN to PA with their own corpus containing 3844 names in both languages with BLEU and word accuracy as 0.4123 (with range 0-1) and 50.22%, respectively.

Nalluri and Kommaluri (2011) has created “enTel,” an SMT-based EN to TE MT system, using the Johns Hopkins University Open Source Architecture (Li *et al.* 2009). For the purpose of training the translation system, TE parallel dataset from the Enabling Minority Language Engineering (EMILLE) is used for their work.

In the year 2014, an SMT Framework for SI-TA MT System has been created by Pushpananda, Weerasinghe, and Niranjan (2014). In their work, the result of SMT-dependent translation between language pairs, including TA-SI and SI-TA has been shown. Outcomes of the experiments using the SMT model give more noticeable results for the SI-TA than the TA-SI language pair. For languages closely related, SMT shows remarkable results.

In 2017, a survey has been conducted by Khan, Anwar, and Durrani (2017) on the IL-EN language MT models revealing the importance of SMT over eight languages that is, HI, BN, GU, UR, TE, PA, TA, and ML. In their work, EMILLE corpus (Nalluri and Kommaluri 2011) is used and Moses SMT model is preferred to make the TMs, with out-of-vocabulary (OOV) words transliterated to EN. In their work, the evaluation using BLEU, NIST and UNK counts as metrics reveals the overall SMT performance as satisfactory (PA-EN and UR-EN models as the best and the HI-EN and GU-EN models as the worst).

An EN-BN SMT system has been presented by Islam, Tiedemann, and Eisele (2010). In their work, a transliteration module is presented to handle OOV words and a preposition handling module has been added to address the systematic grammatical distinctions between EN and BN. BLEU, NIST and TER scores has been used to check the effectiveness of their system.

Nowadays, NMT is widely appreciated for its advancement in the development of MT with remarkable improvement in quality. Hence, many researchers have compared both techniques for low and high-resource languages.

Toral and Sánchez-Cartagena (2017) performed a thorough evaluation using SMT and NMT systems for nine language directions along a variety of dimensions. In their experiment, SMT systems perform better than the NMT for long sentences.

Recently, Castilho *et al.* (2017) used automatic metrics and expert translators to conduct a thorough quantitative and qualitative comparison of NMT and SMT. SMT shows better according to their experiments.

The comparison of NMT and SMT for the NE using the Nepali National Corpus with 6535 sentences has been shown by Acharya and Bal (2018). The researchers have proved in their experiments that the SMT model performs better than the NMT-based system with a small corpus with a 5.27 BLEU score.

Jiao Liu (Liu 2020) provided an extensive comparison and study of the use of cohesive devices by several MT systems in both SMT and NMT circumstances for the Chinese to the EN language. When compared to SMT, the NMT system is more effective in processing (a) additive devices that render EN sentences cohesive, (b) adjectives, (c) and pronouns, indicating better quality of translation.

Abujar *et al.* (2021) developed a BN-EN MT model on AmaderCAT corpus using Sequence-to-Sequence (seq2seq) architecture, a special class of Recurrent Neural Networks to develop the translation system, and has achieved a BLEU score of 22.3.

In the year 2021, translation of EN and HI-to-TA languages using both SMT and NMT has been presented by Ramesh *et al.* (2020). The disadvantages of NMT have been shown in their

experiments such as the occurrence of numerous errors by NMT when interpreting domain terms and OOV phrases. NMT frequently constructs inaccurate lexical choices for polysemous words and occasionally counters reordering mistakes while translating words and domain terms. The translations that have been generated by the NMT models mostly include repetitions of previously transcribed words, odd translations, and many unexpected sentences having no correlation with the original sentence.

In the year 2022, WAT2022 (Workshop on Asian Translation 2022) organizes (hosted by the COLING 2022) various works on Indic languages (Nakazawa *et al.* 2022).

Laskar *et al.* (2022) investigated a transliteration-based method in the MNMT model for two ILs, AS and BN, and achieved BLEU scores of 1.10 for EN to AS and 3.50 for AS to EN. One more research paper related to five ILs (Das *et al.* 2022) also used the MNMT approach with filtration data and achieved an 8.00 BLEU score for NE to the EN language.

3. Experimental framework

This section gives an overview of the dataset and languages used for the experiment and describes the process of building the SMT model.

3.1 Language preference

India is a multilingual nation where people from various states use a variety of regional tongues. Such diversity of language brings difficulty in communicating with one another for information exchange. Further, limitations in public communication also bring inconvenience to share feelings, thoughts, opinions, and facts, as well as to deal with business purposes. Moreover, there are many resources available on the internet in EN but many Indians struggle to take benefit of those due to language barriers (Das *et al.* 2024).

Hence, it is crucial to have an easy translation solution for regional languages to support effective communication and to help utilising global resources. To make it possible, technological innovation is continuing to find out efficient methods for a flawless translation using machines, because it is impractical to have human translators everywhere. For MT, an enormous amount of resources is required for training with a proper knowledge base (rules) for better efficiency so as to fulfill the demand of a flawless translation solution. For translation, understanding the meaning of words is important, but words are not enough to constitute a language as a whole. They must be used in sentence construction that adheres to strict grammar rules and every language is having its own writing style. In the work, fifteen commonly spoken languages (over various regions of India) are chosen. Table 1 describes the languages used for the experiments with their linguistic features (What are the Top 200 Most Spoken Languages).

3.2 Dataset

OPUS and Samanantar datasets for model building and Flores 200 for fine-tuning and testing are utilized. Samanantar is the largest corpus collection for ILs (Ramesh *et al.* 2022). The collection includes more than 45 million sentence pairs in EN and 11 ILs. The Samanantar Corpus has been used for AS, ML, BN, MR, GU, KN, HI, OR, PA, TE, and TA for the experiments.

OPUS is a large resource with freely available parallel corpus, including data from many domains and covering over 90 languages (Tiedemann 2012). The OPUS corpus is used for SI, SD, UR, and NE. Table 2 gives statistics of the dataset used in the experiments.

FLORES-200 (Costa-jussà *et al.* 2022) corpora are multilingual parallel corpora with 200 languages, that are used as human-translated benchmarks. They consist of two corpora, labeled “dev” (997 lines) and “devtest” (1,012 lines). The “dev” corpus has been used for fine-tuning, and the “devtest” corpus has been used for testing. These fifteen languages are datasets of WAT2022.

Table 1. Linguistic features of languages used in MT experiments

Languages	Script	Word order	Family	Number of speakers (in millions)	Writing direction
Assamese (AS)	Assamese	SOV	Indo-European	15	Left to right
Malayalam (ML)	Malayalam	SOV	Dravidian	38	Left to right
Bengali (BN)	Bengali	SOV	Indo-European	265	Left to right
Marathi (MR)	Devanagari	SOV	Indo-European	95	Left to right
Gujarati (GU)	Gujarati	SOV	Indo-European	60	Left to right
Kannada (KN)	Kannada	SOV	Dravidian	36	Left to right
Hindi (HI)	Devanagari	SOV	Indo-European	615	Left to right
Oriya (OR)	Oriya	SOV	Indo-European	38	Left to right
Punjabi (PA)	Perso-Arabic, Gurmukhi	SOV	Indo-European	125	Right to left Left to right
Telugu (TE)	Telugu	SOV	Dravidian	93	Left to right
Sindhi (SD)	Devanagari Perso-Arabic	SOV	Indo-European	25	Left to right Right to left
Sinhala (SI)	Sinhala	SOV	Indo-European	17	Left to right
Nepali (NE)	Devanagari	SOV	Indo-European	24	left to right
Tamil(TA)	Tamil	SOV	Dravidian	81	left to right
Urdu (UR)	Urdu	SOV	Indo-European	170	right to left
English (EN)	Roman	SVO	Indo-European	1,132	left to right

3.3 Methodology

Fig. 1 describes the process of building of the SMT model. The steps for building the SMT model are described below.

- (1) **Pretraining Preprocessing Stage I:** Some of the punctuation in the extended Unicode were converted to their standard counterparts. The accented characters were removed. Numbers in the ILs were converted from ENG to Indic scripts. Characters outside the standard alphabets of the language pair, extra spaces, and the unprintable characters were removed from the corpus. The corpus is then tokenized using a modified Moses tokenizer (Koehn *et al.* 2007), which involves dividing a character sequence into smaller tokens, such as words, punctuation, and numerals, based on a given character sequence and a specified document unit. Finally, redundant quotation marks were removed from the corpus. The modified Moses tokenizer is tailored for ILs. It deals with diacritics like halants and nuktas effectively. For example, in AS সমৰ ্থন is modified into সমৰ্থন.
- (2) **Training Truecasing Model:** In Moses, the training procedure utilizes word and segment occurrences to draw connections between the target and source languages. The language and TMs are trained on the training dataset and binarized. GIZA++ grow-diag-final-and alignment is used for word alignments, which start with the intersection of the two alignments and then add the additional alignment points. A truecaser model (a model which changes the words at the beginning of the sentence were changed to the most common

Table 2. Parallel corpus statistics

English to Indic	Parallel corpus (sentences)
Assamese (AS)	0.14M
Malayalam (ML)	5.85M
Bengali (BN)	8.52M
Marathi (MR)	3.32M
Gujarati (GU)	3.05M
Kannada (KN)	4.07M
Hindi (HI)	8.56M
Oriya (OR)	1.00M
Punjabi (PA)	2.42M
Telugu (TE)	4.82M
Sindhi (SD)	1.95M
Sinhala (SI)	8.68M
Nepali (NE)	3.35M
Tamil (TA)	5.16M
Urdu (UR)	8.95M

casing) is trained on the train dataset. The Moses truecaser is used for the same. The train, dev, and test datasets are truecased using the trained truecase model.

- (3) **Pretraining Preprocessing Stage II:** The train, dev, and test datasets are cleaned using `clean-corpus-n.perl` which removes empty lines and redundant space characters. GIZA++ cannot handle parallel sentences where the ratio of the number of tokens on one side over the other is more than 9:1. Such translations are indicative of too short or too long translations and may indicate incorrect translations. Lines that violate the 9:1 sentence ratio limit of GIZA++ are also removed from the corpus.
- (4) **Training Language and TMs:** In Moses, the training procedure utilizes word and segment occurrences to draw connections between the target and source languages. The language and TMs are trained on the training dataset and binarized. After training the language model, the next step is to train TMs. The proposed work utilizes the GIZA++ (Costa-jussà *et al.* 2003) grow-diag-final-and alignment. Two different reordering methods that is, msd-bidirectional-fe reordering and distance-based reordering are also explored, and their efficiency is checked in terms of ILs in this paper.
- (5) **Fine-tuning:** It is the process of determining the best parameters for a TM when it is used for a specific purpose. It uses a TM to translate all fifteen ILs source language phrases in the fine-tuning set. Then, it compares the model's output to a set of references (human translations) and adjust the variables to improve translation quality. This procedure is repeated several times until the translation quality is optimized. The model is fine-tuned on the preprocessed Flores-200 dev dataset.
- (6) **Translation:** The final model is filtered based on the test dataset and then used to translate the preprocessed test dataset from the source to the target language.

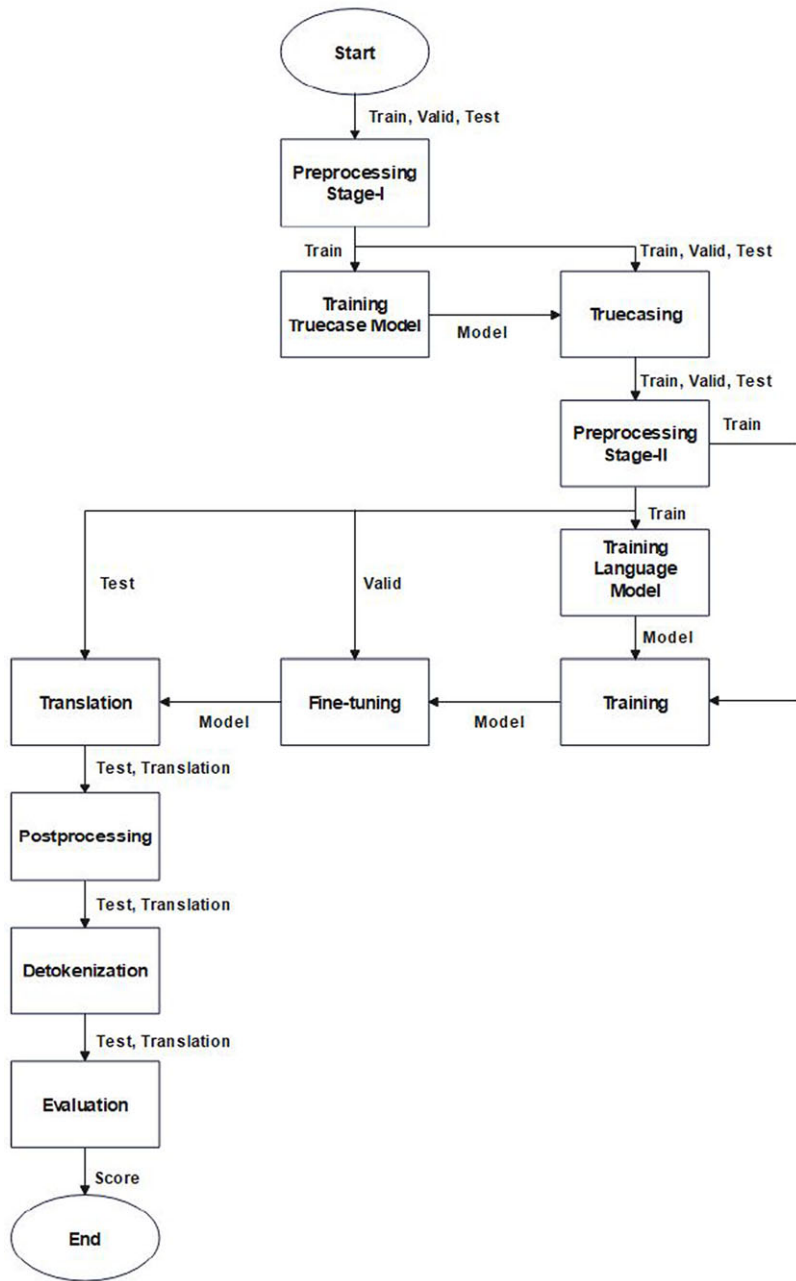


Figure 1. Workflow of Statistical Machine Translation(SMT) model.

- (7) **Postprocessing and Detokenization:** Redundant quotation marks were removed, and the translation file is detokenized using the Moses detokenizer.
- (8) **Evaluation:** The evaluation metrics used for the experiments are TER (Snover *et al.* 2006), METEOR (Banerjee and Lavie 2005), RIBES (Wołk and Koržinek 2016), and BLEU (Papineni *et al.* 2002). Evaluation is done on the Flores-200 devtest dataset. TER can assume any non-negative value, with higher scores being worse. RIBES and METEOR

range between 0 and 1 whereas BLEU ranges from 0 to 100 where higher scores reflect better accuracy.

3.4 Neural machine translation (NMT) system

The procedures for NMT are similar to those used for SMT. Initial data processing includes Preprocessing Stage 1, truecasing, preprocessing Stage 2, and then some additional steps such as concatenation of processed data, vocabulary creation and pruning (vocabulary size is 16,000), binarization and model training, detokenization, and finally the calculation of evaluation metrics. The experiment for building the NMT model utilizes the Transformer model (Vaswani *et al.* 2017) implemented under the Fairseq library (Ott *et al.* 2019), an open-source sequence modeling toolkit that enables researchers to train custom models for MT tasks. The model comprises six encoder-decoder layers, each with 512 hidden units and optimized with Adam optimizer. The dropout (Srivastava *et al.* 2014) value of 0.3 is applied to each sublayer output before it is added to and normalized with the sublayer input. The model is trained for 300,000 updates to ensure that the training process is completed in a reasonable time frame, balancing model convergence with computational efficiency. The model uses label-smoothed cross entropy as the evaluation criterion. Label smoothing is an effective technique for improving model performance and is a parameter of the label-smoothed cross-entropy loss. It is done so that the model is not overconfident in its prediction and acquires better ways to represent the target language. The maximum number of tokens in a batch is set as 4,096. This series of steps aim to create an NMT model that includes the required preprocessing, training, detokenization, and finally evaluation stages. The best model is used to calculate the evaluation metrics.

4. Results and discussion

Tables 3 and 4 display the evaluation metrics of all the fifteen ILs to EN and vice versa utilizing SMT (with and without fine-tuning) incorporating distance and msd-bidirectional-fe reordering models, respectively.

For EN-IL and IL-EN language using SMT with distance reordering (with and without fine-tuning), the BLEU scores lie between 1.03 to 15.41. The RIBES score lies between 0.08 and 0.63 whereas the METEOR score lies between 0.02 and 0.28. TER ranges between 72.42 and 172.38. From the result, it is observed that the SMT models using the distance reordering techniques are giving better results for languages such as BN, PA, UR, and HI than other languages in terms of all evaluation metrics as shown in Table 3.

It is also noticed that fine-tuning typically decreases the efficacy of the models across all languages, with notable exceptions being SD where fine-tuning helps in increasing efficacy, and AS, where no significant change is observed. Upon further inspection, it is found that the translations present in the two corpora are not accurate. For instance, in the AS corpus, “ (“ যিৰিমিয়া ৩৮ : ৪ , ১৭ , ১৮) যিৰিমিয়াৰ দৰে আমিও লোকসকলক এক অপ্ৰীতিকৰ বাৰ্তা ঘোষণা কৰোঁ । ” is translated as “So if we take some false step before we are aware of it and receive needed counsel from Gods Word, let us imitate Baruchs maturity, spiritual discernment, and humility. Galatians 6: 1,” but is translated by Google NMT as “(Jeremiah 38:4, 17, 18) Like Jeremiah, we declare an unpleasant message to the people.” Similarly, in the SD corpus, “... کُل آن جو قبيلي كهڙي ڪر، ” is literally translated in the dataset as “But fill me with the old familiar Juice,” but is translated by Google NMT as “Tell me, which tribe. ...” Therefore, fine-tuning on a more accurate corpus increases the efficacy of the models. For other languages, the difference in linguistic characteristics (differences in the grammar, language structure, or vocabulary) of Samanantar and Flores-200 dev datasets may explain the decrease in efficacy of the models.

Table 3. Evaluation metric result of SMT with and without fine-tuning using distance-based reordering model

Lang	Language pairs	Without fine-tuning				With fine-tuning			
		BLEU	RIBES	METEOR	TER	BLEU	RIBES	METEOR	TER
AS	EN-AS	1.91	0.52	0.09	89.00	2.17	0.50	0.08	88.80
	AS-EN	3.60	0.45	0.11	92.12	2.84	0.42	0.10	95.10
BN	EN-BN	6.64	0.63	0.19	78.90	4.31	0.58	0.13	87.89
	BN-EN	13.08	0.61	0.27	74.84	6.29	0.51	0.16	84.67
MR	EN-MR	4.89	0.49	0.13	86.14	3.43	0.45	0.10	93.48
	MR-EN	8.49	0.54	0.20	79.20	4.63	0.48	0.13	86.34
PA	EN-PA	6.59	0.59	0.23	82.05	5.23	0.56	0.19	85.97
	PA-EN	11.76	0.58	0.23	84.73	7.60	0.52	0.17	94.37
ML	EN-ML	3.79	0.27	0.08	97.67	2.05	0.23	0.06	122.01
	ML-EN	4.59	0.43	0.12	85.55	1.84	0.27	0.06	92.43
OR	EN-OR	3.92	0.59	0.14	84.00	5.25	0.58	0.15	86.78
	OR-EN	6.41	0.52	0.17	82.25	2.22	0.39	0.11	172.38
SD	EN-SD	1.29	0.39	0.08	88.45	1.59	0.41	0.09	87.93
	SD-EN	2.48	0.35	0.09	88.31	2.53	0.38	0.09	89.57
UR	EN-UR	9.43	0.62	0.24	75.58	6.34	0.56	0.19	81.26
	UR-EN	11.35	0.61	0.23	76.75	7.07	0.54	0.18	87.66
HI	EN-HI	13.09	0.63	0.28	72.42	8.64	0.57	0.22	79.89
	HI-EN	15.41	0.64	0.28	72.66	5.38	0.49	0.14	88.45
GU	EN-GU	7.62	0.56	0.16	81.93	5.82	0.52	0.14	88.96
	GU-EN	10.14	0.59	0.21	77.41	3.56	0.45	0.10	89.00
KN	EN-KN	5.06	0.40	0.11	91.28	3.35	0.35	0.09	107.35
	KN-EN	7.17	0.51	0.16	81.94	3.67	0.41	0.10	88.45
TA	EN-TA	2.78	0.16	0.05	107.69	1.86	0.16	0.05	111.99
	TA-EN	2.64	0.31	0.07	90.22	1.03	0.08	0.02	97.10
TE	EN-TE	5.76	0.43	0.13	89.76	3.33	0.38	0.09	104.57
	TE-EN	8.16	0.52	0.22	80.06	4.18	0.43	0.13	87.38
SI	EN-SI	4.05	0.36	0.11	107.68	2.60	0.31	0.07	111.99
	SI-EN	6.64	0.48	0.19	116.48	3.39	0.11	0.11	104.95
NE	EN-NE	6.00	0.58	0.16	82.89	4.00	0.55	0.14	91.65
	NE-EN	8.29	0.53	0.19	79.27	5.25	0.49	0.13	85.09

Table 4. Evaluation metric result of SMT with and without fine-tuning using msd bidirectional reordering

Lang	Language Pairs	Without fine-tuning				With fine-tuning			
		BLEU	RIBES	METEOR	TER	BLEU	RIBES	METEOR	TER
AS	EN-AS	2.52	0.52	0.10	88.64	2.52	0.52	0.09	95.55
	AS-EN	4.48	0.48	0.13	89.11	3.89	0.48	0.11	90.00
BN	EN-BN	8.23	0.63	0.20	80.96	4.94	0.58	0.14	88.98
	BN-EN	14.54	0.62	0.28	75.53	7.09	0.52	0.17	84.78
MR	EN-MR	5.23	0.51	0.14	85.32	3.87	0.46	0.10	93.15
	MR-EN	9.02	0.57	0.22	78.90	4.99	0.49	0.14	86.17
PA	EN-PA	5.88	0.56	0.20	82.15	4.48	0.52	0.16	85.10
	PA-EN	12.48	0.61	0.26	77.88	7.88	0.55	0.15	86.97
ML	EN-ML	3.88	0.28	0.08	98.08	2.18	0.23	0.06	122.26
	ML-EN	4.91	0.43	0.16	85.77	1.94	0.28	0.08	92.43
OR	EN-OR	4.36	0.55	0.15	83.36	3.59	0.57	0.10	89.64
	OR-EN	7.46	0.53	0.20	81.21	5.20	0.50	0.15	85.64
SD	EN-SD	1.30	0.39	0.10	88.67	1.65	0.43	0.11	88.57
	SD-EN	2.46	0.36	0.10	88.15	2.48	0.38	0.10	89.88
UR	EN-UR	10.25	0.63	0.28	74.75	7.02	0.56	0.22	80.56
	UR-EN	12.04	0.62	0.25	75.91	7.69	0.55	0.18	86.01
HI	EN-HI	14.32	0.65	0.32	71.21	9.94	0.59	0.24	78.98
	HI-EN	18.51	0.65	0.31	71.30	9.41	0.56	0.21	82.82
GU	EN-GU	8.15	0.53	0.19	81.04	6.63	0.54	0.15	87.88
	GU-EN	11.19	0.58	0.25	76.28	6.01	0.52	0.16	84.70
KN	EN-KN	5.69	0.41	0.12	90.67	3.71	0.36	0.08	107.11
	KN-EN	7.79	0.51	0.20	81.42	3.78	0.41	0.11	88.74
TA	EN-TA	2.76	0.19	0.05	109.37	1.90	0.18	0.04	145.00
	TA-EN	2.81	0.31	0.11	90.10	1.64	0.20	0.06	94.04
TE	EN-TE	6.18	0.44	0.13	89.43	3.68	0.39	0.10	104.18
	TE-EN	8.72	0.53	0.22	79.84	4.75	0.45	0.14	87.84
SI	EN-SI	4.53	0.36	0.11	107.68	2.72	0.33	0.07	105.30
	SI-EN	7.02	0.48	0.19	82.61	3.48	0.38	0.11	89.18
NE	EN-NE	6.53	0.59	0.18	81.88	4.52	0.56	0.14	90.07
	NE-EN	9.32	0.56	0.23	77.91	6.01	0.50	0.16	84.39

Table 5. Percentile threshold of corpora

%	AS	BN	GU	HI	KN	ML	MR	NE	OR	PA	SD	SI	TA	TE	UR
40	7	7	8	13	5	4	7	4	7	10	11	5	4	6	9
80	16	17	14	28	10	8	14	11	14	21	23	10	9	10	23

Using msd-bidirectional reordering (with and without fine-tuning) the BLEU scores lie between 1.30 and 18.51. TER ranges between 71.21 and 109.37. RIBES scores lie between 0.18 and 0.65, whereas METEOR scores lie between 0.04 and 0.32. For most of the ILs, it is observed that msd-bidirectional reordering performs better than distance reordering. HI, UR, PA, and BN languages have qualitative and less noisy datasets compared to other languages and hence performed better than other languages as shown in Table 4. Additionally, HI, UR, and BN corpora are large in size which contributes to better models. HI performs the best among all languages with and without fine-tuning in all three metrics in both directions in both reordering models, except in IL-EN models with distance reordering and fine-tuning, where PA-EN performs better than the rest. Using both reordering methods, it is observed that ML, SD, and TA perform worse than the other languages, while AS and SI also produce low scores in several cases.

TA and ML are highly agglutinative languages, exhibiting great lexical and morphological diversity. Literature shows that word tokenization does not perform very well in the case of agglutinative languages due to its large and sparse vocabulary. In such cases, techniques such as subword tokenization might be useful.

The inconsistencies among RIBES, METEOR, TER, and BLEU are observed because they concentrate on different aspects of language quality. BLEU is based on the n-gram, which means if the machine-generated output matches the reference text, then it shares more n-gram with the reference text. RIBES scores focus on machine-generated output with good phrase structure and word order. METEOR is intended to detect semantic similarities between the reference and machine-generated output. TER is evaluated by quantifying the edit operations required for a hypothesis to match a reference translation. It is noticed from the result that TER gives inconsistent results as compared to the remaining three metrics. In various cases, the other three metrics indicate that msd bidirectional reordering is better as compared to distance-based reordering, but TER says otherwise. This is because TER only considers a number of edits and does not take into consideration other factors such as location and size of continuous matchings.

It is also noticed that even though the IL-EN and EN-IL systems are trained using the same corpus, the former performs better for all the evaluation metrics. This is due to the significant morphological diversity of ILs and the relative difficulty of translating from EN to ILs. As per our observation using both the reordering technique, performance is comparatively less in the case of small sentences in ILs (such as TA and ML); whereas it is better in the case of languages (like HI and BN) with longer sentences. Table 5 describes the percentile threshold of the number of lines in corpora in 40 percent and 80 percent with respect to the number of tokens in the line.

4.1 Comparison with neural machine translation (NMT) system

Using NMT models BLEU score lies between 0.01 and 17.78, whereas the RIBES score ranges 0.07–0.76. Additionally, the METEOR score exhibits a range of 0.02–0.49, and the RIBES score spans from 61.87 to 592.74 as shown in Table 6. When examining the results for NMT models, it is clear that the AS language consistently gets lower scores across all metrics. In contrast, PA, UR, and BN languages consistently outperform all languages on these metrics.

While comparing evaluation metrics for SMT and NMT models as shown in Table 7, it is observed that there are many cases where either of them is performing better. Metrics are inconsistent with one another such as in many cases BLEU scores evaluate SMT models as better, while the other three metrics show that NMT is better.

Table 6. Evaluation metrics for Neural Machine Translation (NMT)

Lang	Language pairs	BLEU	RIBES	METEOR	TER
AS	EN-AS	0.01	0.07	0.02	592.74
	AS-EN	0.43	0.37	0.11	170.82
BN	EN-BN	10.58	0.72	0.36	68.87
	BN-EN	17.29	0.76	0.48	61.19
MR	EN-MR	1.98	0.57	0.21	131.43
	MR-EN	10.17	0.67	0.36	71.56
PA	EN-PA	15.46	0.75	0.41	65.60
	PA-EN	17.78	0.78	0.49	57.87
ML	EN-ML	0.71	0.27	0.09	97.69
	ML-EN	2.73	0.53	0.21	103.20
OR	EN-OR	1.38	0.18	0.15	115.94
	OR-EN	2.63	0.32	0.23	109.15
SD	EN-SD	0.03	0.08	0.02	164.46
	SD-EN	0.23	0.18	0.07	135.26
UR	EN-UR	14.32	0.70	0.39	69.35
	UR-EN	15.66	0.76	0.47	61.22
GU	EN-GU	3.49	0.73	0.16	94.54
	GU-EN	9.38	0.45	0.35	73.9
KN	EN-KN	3.99	0.62	0.22	77.0
	KN-EN	7.95	0.65	0.30	78.60
TA	EN-TA	0.31	0.16	0.07	106.65
	TA-EN	1.92	0.46	0.16	114.98
TE	EN-TE	5.54	0.64	0.24	74.69
	TE-EN	9.77	0.68	0.34	74.36
SI	EN-SI	4.23	0.58	0.23	83.30
	SI-EN	8.92	0.65	0.33	76.84
NE	EN-NE	3.74	0.46	0.21	87.80
	NE-EN	9.46	0.63	0.35	73.42

5. Conclusion and future work

This paper presented the MT work for fifteen ILs to EN and vice versa using SMT with two different reordering techniques. The grow-diag-final-and alignment model distance and msd-bidirectional-fe reordering models are utilized for the experiment. A tailor-made preprocessing approach has been incorporated, and linguistic features of ILs are elaborated in this work.

Table 7. Translation evaluation metrics by language (right-tick indicates better performance of SMT)

Lang	Lang pairs	BLEU	RIBES	METEOR	TER
AS	EN-AS	✓	✓	✓	✓
	AS-EN	✓	✓	✓	✓
BN	EN-BN	✗	✗	✗	✗
	BN-EN	✗	✗	✗	✗
MR	EN-MR	✓	✗	✗	✓
	MR-EN	✗	✗	✗	✗
PA	EN-PA	✗	✗	✗	✗
	PA-EN	✗	✗	✗	✗
ML	EN-ML	✓	✓	✗	✓
	ML-EN	✓	✗	✗	✓
OR	EN-OR	✓	✓	✓	✓
	OR-EN	✓	✓	✗	✓
SD	EN-SD	✓	✓	✓	✓
	SD-EN	✓	✓	✓	✓
UR	EN-UR	✗	✗	✗	✗
	UR-EN	✗	✗	✗	✗
GU	EN-GU	✓	✓	✓	✓
	GU-EN	✓	✗	✗	✗
KN	EN-KN	✓	✗	✗	✗
	KN-EN	✗	✗	✗	✗
TA	EN-TA	✓	✓	✗	✗
	TA-EN	✓	✗	✗	✓
TE	EN-TE	✓	✗	✗	✗
	TE-EN	✗	✗	✗	✗
SI	EN-SI	✓	✗	✗	✗
	SI-EN	✗	✗	✗	✗
NE	EN-NE	✓	✓	✗	✓
	NE-EN	✗	✗	✗	✗

For checking the quality of translation, different evaluation metrics such as BLEU, RIBES, TER, and METEOR are utilized. From the result, it is observed that the proposed SMT model using the msd-bidirectional-fe reordering technique is better than distance reordering for ILs. However, due to the scarcity and quality of parallel corpus, the metrics obtained are quite low. It has been observed that the translations of the languages are not sufficiently accurate. Measures of validating

corpus quality shall be explored in order to observe the corpus quality and remove inaccurate lines. Dravidian languages are in general agglutinative languages (words are made up of morphemes, with each morpheme contributing to the meaning of the word). The paper also includes a comparison between SMT and NMT. The results show that, across various languages, SMT performs better in some cases, while NMT outperforms in others. In the future, techniques to infer translations from the breakdown of words in these languages shall also be explored. Interestingly, in some of the ILs, fine-tuning schemes are hampering the quality. The causes of this phenomenon shall be analyzed and mitigated via techniques such as noise reduction, corpus cleaning, and fine-tuning schemes for those languages to ensure better quality. Other techniques, such as hybridized SMT-NMT models, and the usage of other alignment, agglutination, and reordering models will be studied for further course of research.

Acknowledgment. This research is partially funded by Meity (Ministry of Electronics and Information Technology, Government of India) for project sanction no. 13 (12)/2020-CC BT dated 24.04.2020.

References

- Abujar S., Masum A.K.M., Bhattacharya A., Dutta S. and Hossain S.A. (2021). English to Bengali neural machine translation using global attention mechanism. In *Emerging Technologies in Data Mining and Information Security 3*, pp. 359–369.
- Acharya P. and Bal B.K. (2018). A comparative study of SMT and NMT: case study of English-Nepali language pair. *International Journal of Applied Engineering Research*, 90–93.
- Banerjee S. and Lavie A. (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*, pp. 65–72.
- Brown P.F., Cocke J., Della Pietra S.A., Della Pietra V. J., Jelinek F., Lafferty J.D., Mercer R.L. and Roossin P.S. (1990). A statistical approach to machine translation. *Computational Linguistics* 16(2), 79–85.
- Carl M. and Andy W. (2000). Towards a dynamic linkage of example-based and rule-based machine translation. *Machine Translation* 15, 223–257.
- Castilho S., Moorkens J., Gaspari F., Sennrich R., Sosoni V., Georgakopoulou Y., Lohar P., Way A., Miceli Barone A.V. and Gialama M. (2017). A comparative quality evaluation of PBSMT and NMT using professional translators. In *Proceedings of Machine Translation Summit XVI*, pp. 116–131.
- Castilho S., Moorkens J., Gaspari F. and Way A. (2017). Is neural machine translation the new state of the art? *Prague Bulletin of Mathematical Linguistics* 108(1), 109–120.
- Chapelle C.A. and Chung Y.-R. (2010). Natural language processing and machine translation encyclopedia of language and linguistics. *Language Testing* 27(3), 301–315.
- Costa-jussà M.R. (2012). Study and comparison of rule-based and statistical Catalan-Spanish machine translation systems. *Computing and Informatics* 31(2), 245–270.
- Costa-jussà M.R., Cross J., Çelebi O., Elbayad M., Heafield K., Heffernan K., Kalbassi E., et al. (2003). Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, pp. 160–167.
- Costa-jussà M.R., Cross J., Çelebi O., Elbayad M., Heafield K., Heffernan K., Kalbassi E., et al. (2022). No language left behind: scaling human-centered machine translation. arXiv preprint arXiv:2207.04672, pp. 2214–2218.
- Das S.B., Biradar A., Mishra T.K. and Patra B.K. (2022). NIT Rourkela Machine Translation (MT) system submission to WAT 2022 for MultiIndicMT: an Indic language multilingual shared task. In *Proceedings of the 9th Workshop on Asian Translation*, pp. 73–77.
- Das S.B., Biradar A., Mishra T.K. and Patra B.K. (2022). Improving multilingual neural machine translation system for Indic languages. *ACM Transactions on Asian and Low-Resource Language Information Processing* 22(6), 1–23.
- Das S.B., Panda D., Mishra T.K., Patra B.K. and Ekbal A. (2024). Multilingual neural machine translation system for Indic to Indic languages. *ACM Transactions on Asian and Low-Resource Language Information Processing* 23(5), 1–32.
- Dasgupta S., Wasif A. and Azam S. (2004). An optimal way of machine translation from English to Bengali. An optimal way of machine translation from English to Bengali. In *Proceeding of International Conference on Computer and Information*, pp. 1–6.
- Dorr B.J., Hovy E.H. and Levin L.S. (2004). Natural language processing and machine translation encyclopedia of language and linguistics, (ell2). machine translation: Interlingual methods. In *Proceeding of International Conference of the World Congress on Engineering*, pp. 1–20.

- Girgzdis V., Kale M., Vaicekauskis M., Zarina I. and Skadina I. Tracing mistakes and finding gaps in automatic word alignments for Latvian-English translation. In *Baltic HLT*, pp. 87–94.
- Hearne M. and Way A. (2011). Statistical machine translation: a guide for linguists and translators. *Language and Linguistics Compass* 5(5), 205–226.
- Hutchins W.J. (2005). Towards a definition of example-based machine translation. In *Workshop on Example-Based Machine Translation*, pp. 223–257.
- Islam Z., Tiedemann J. and Eisele A. (2010). English to Bangla phrase-based machine translation. In *Proceedings of the 14th Annual Conference of the European Association for Machine Translation*, pp. 1–8.
- Kaur J. and Josan G.S. (2011). Statistical approach to transliteration from English to Punjabi. *International Journal on Computer Science and Engineering* 3(4), 1518–1527.
- Khan N.J., Anwar W. and Durrani N. (2017). Machine translation approaches and survey for Indian languages. arXiv preprint arXiv: 1701.04290.
- Koehn P., Hoang H., Birch A., Callison-Burch C., Federico M., Bertoldi N., Cowan B., Shen W., Moran C., Zens R., Dyer C., Bojar O., Constantin A. and Herbst E. (2018). Moses: open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pp. 2255–2266.
- Koehn P., Hoang H., Birch A., Callison-Burch C., Federico M., Bertoldi N., Cowan B., Shen W., Moran C., Zens R., Dyer C., Bojar O., Constantin A. and Herbst E. (2007). Moses: open source toolkit for statistical machine translation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 177–180.
- Kumawat S. and Chandra N. (2014). Distance-based reordering in English to Hindi statistical machine translation. *International Journal of Computer Applications* 124(10), 27–40.
- Laskar S.R., Manna R., Pakray P. and Bandyopadhyay S. (2022). Investigation of multilingual neural machine translation for Indian languages. In *Proceedings of the 9th Workshop on Asian Translation*, pp. 78–81.
- Li Z., Callison-Burch C., Dyer C., Khudanpur S., Schwartz L., Thornton W., Weese J. and Zaidan O. (2009). Joshua: an open source toolkit for parsing-based machine translation. In *Proceedings of the Fourth Workshop on Statistical Machine Translation*, pp. 135–139.
- Lita L.V., Ittycheriah A., Roukos S. and Kambhatla N. (2003). Development of English-to-Bengali neural machine translation systems. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pp. 152–159.
- Lita L.V., Ittycheriah A., Roukos S. and Kambhatla N. (2003). Trucasing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, pp. 152–159.
- Liu J. (2020). Comparing and analyzing cohesive devices of SMT and NMT from Chinese to English: a diachronic approach. *Open Journal of Modern Linguistics* 10(06), 765–772.
- Lohar P., Popović M., Alfi H. and Way A. (2019). A systematic comparison between SMT and NMT on translating user-generated content. In *Proceeding of International Conference on Computational Linguistics and Intelligent Text Processing*, pp. 7–13.
- Lopez A. (2008). Statistical machine translation. *ACM Computing Surveys (CSUR)* 40(3), 1–49.
- Mahata S.K., Mandal S., Das D. and Bandyopadhyay S. (2018). SMT vs NMT: a comparison over Hindi and Bengali simple sentences. arXiv preprint arXiv:1812.04898.
- Marton Y., Callison-Burch. C. and Resnik P. (2009). Improved statistical machine translation using monolingually-derived paraphrases. *Association for Computational Linguistics*, pp. 381–390.
- Nakazawa T., Nakayama H., Ding C., Dabre R., Higashiyama S., Mino H., Goto I., Pa W.P., Kunchukuttan A., Parida S., Bojar O., Chu C., Eriguchi A., Abe K., Oda Y. and Kurohashi S. (2022). Overview of the 8th workshop on Asian translation. In *Proceedings of the 8th Workshop on Asian Translation (WAT2021)*, pp. 1–45.
- Nalluri A. and Kommaluri V. (2011). Statistical Machine Translation using Joshua: an approach to build “enTel” system. In *Parsing in Indian Languages* 11(5), pp. 1–6.
- Ott M., Edunov S., Baevski A., Fan A., Gross S., Ng N., Grangier D. and Auli M. (2019). fairseq: a fast, extensible toolkit for sequence modeling. arXiv preprint arXiv:1904.
- Papineni K., Roukos S., Ward T. and Zhu W.-J. (2002). Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pp. 311–318.
- Pushpananda R., Weerasinghe R. and Niranjana M. (2014). Sinhala-Tamil machine translation: towards better translation quality. In *Proceedings of the Australasian Language Technology Association Workshop*, pp. 129–133.
- Ramanathan A., Choudhary H., Ghosh A. and Bhattacharyya P. (2009). Case markers and morphology: addressing the crux of the fluency problem in English-Hindi SMT. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pp. 800–808.
- Ramanathan A., Hegde J., Shah R.M., Bhattacharyya P. and Sasikumar M. (2008). Simple syntactic and morphological processing can help English-Hindi statistical machine translation. In *Proceedings of the Third International Joint Conference on Natural Language Processing*, pp. 1–6.
- Ramesh G., Doddapaneni S., Bheemaraj A., Jobanputra M., Raghavan A.K., Sharma A. and Sahoo S. (2022). Samanantar: the largest publicly available parallel corpora collection for 11 Indic languages. *Transactions of the Association for Computational Linguistics* 10, 145–162.

- Ramesh A., Parthasarathy V.B., Haque R. and Way A.** (2020). Comparing statistical and neural machine translation performance on Hindi-to-Tamil and English-to-Tamil. *Digital* 1(2), 86–102.
- Shiva Kumar K.M., Namitha B.N. and Nithya R.** (2015). A comparative study of english to kannada baseline machine translation system with general and bible text corpus. *International Journal of Applied Engineering Research* 10(12), 30195–30201.
- Snover M., Dorr B., Schwartz R., Micciulla L. and Makhoul J.** (2006). A study of translation edit rate with targeted human annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*, pp. 223–231.
- Somers H.** (2011). Machine translation: history, development, and limitations. In *The Oxford Handbook of Translation Studies*.
- Srivastava N., Hinton G., Krizhevsky A., Sutskever I. and Salakhutdinov R.** (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research* 15, 1929–1958.
- Stahlberg F.** (2020). The roles of language models and hierarchical models in neural sequence-to-sequence prediction. *European Association for Machine Translation*, 5–6.
- Stasimioti M.** (2020). Machine translation quality: a comparative evaluation of SMT, NMT, and tailored-NMT outputs. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, pp. 1–49.
- Steinberger R., Eisele A., Kloczek S., Pilos S. and Schlüter P.** (2013). DGT-TM: a freely available translation memory in 22 languages. arXiv preprint arXiv:1309.5226.
- Tiedemann J.** (2012). Parallel data, tools and interfaces in OPUS. In *Lrec*, pp. 2214–2218.
- Toral A. and Sánchez-Cartagena V.M.** (2017). A multifaceted evaluation of neural versus phrase-based machine translation for 9 language directions. *Association for Computational Linguistics* 1, pp. 1063–1073.
- Tripathi S. and Sarkhel J.K.** (2010). Approaches to machine translation. Council Of Scientific And Industrial Research–National Institute Of Science Communication and Policy Research Council Of Scientific And Industrial Research–National Institute Of Science Communication and Policy Research, pp. 388–393.
- Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł. and Polosukhin I.** (2017). Attention is all you need. In *Advances in neural information processing systems* 30, 1–11.
- Wang X., Lu Z., Tu Z., Li H., Xiong D. and Zhang M.** (2017). Neural machine translation advised by statistical machine translation. In *Thirty-First AAAI Conference on Artificial Intelligence* 31(1), pp. 1–7.
- Wang X., Tu Z. and Zhang M.** (2018). Incorporating statistical machine translation word knowledge into neural machine translation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 26(12), 2255–2266.
- What are the top 200 Most Spoken Languages (2022). *Ethnologue: Languages of the World*. Available at <https://www.ethnologue.com/guides/ethnologue200>.
- Wolk K. and Koržinek D.** (2016). Comparison and Adaptation of Automatic Evaluation Metrics for Quality Assessment of Re-Speaking. arXiv preprint arXiv:1601.02789.
- Zbib R., Matsoukas S., Schwartz R. and Makhoul J.** (2010). Decision trees for lexical smoothing in statistical machine translation. *Association for Computational Linguistics*, pp. 428–437.
- Zens R., Och F.J. and Ney H.** (2002). Phrase-based statistical machine translation. In *Annual Conference on Artificial Intelligence*, pp. 18–32.
- Zhou L., Hu W., Zhang J. and Zong C.** (2017). Neural system combination for machine translation, pp. 378–384. arXiv preprint arXiv:1704.06393.
- Ziemski M., Junczys-Dowmunt M. and Pouliquen B.** (2016). The United Nations parallel corpus v1.0. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, pp. 3530–3534.