Evolution and Comprehensive Analysis of Machine Translation Systems for Regional Languages

Sachin Majithia[1], Harjeet Singh[2], Rajeev Puri[3]

sachinmajithia@gmail.com[1], zrjeet@gmail.com[2], rpuri@davjalandhar.com[3]

Research Scholar, Department of Computer Science and Engineering Punjabi University, Patiala, Punjab India[1], Associate Professor, Department of Computer Science, Mata Gujri College, Fatehgarh Sahib, Punjab India[2], Associate Professor, PG Department of Computer Science, DAV College, Jalandhar, Punjab, India[3]

Abstract. India, which is a vast multilingual community, has 22 Indian Languages approved by the constitution, so there is much need of machine translation of documents between different language pairs. They have similar scripts and grammar. Because of all of these factors, language translation is essential. Over the last several years, much research has been going on in Machine Translation in India. During the early 90s, advanced studies in natural language processing made more favorable machine translation systems. This aided in the creation of practical machine translation systems in a few specific domains [4]. This research article presents the evolution of Machine Translation for regional languages from rule-based approaches to Neural Architectures, which deliver near-human quality machine translations. Most of the Machine Translation systems are evaluated using standard metrics like accuracy and BLEU score. We compare BLEU, Accuracy, ChrF Score and Comet Score for the evaluation of machine translation approaches used for regional Languages.

Keywords: Machine Translation, regional languages, statistical models, neural network, Hindi, Punjabi, deep learning, machine translation

1. Introduction

The history of machine translation development of Indian dialects has taken a significant technological path, being the primitive systems, through the complicated deep learning systems. Initial work was led by Rule-Based Machine Translation (RBMT) that was premised upon manual linguistic rules and morphological parsers; however it was unable to cope with the morphological heterogeneity and syntactic diversity of Indic languages. This was succeeded by a shift in paradigm to data-driven models that employed Statistical Machine Translation (SMT), which employed probabilistic models that were trained in parallel corpora, but constrained by the fact that most of such languages were low-resource languages. The introduction of the concept of Neural Machine Translation (NMT), particularly the Transformer model, in which the semantic context is considered to generate more fluent translations, has been

IICTDS 2025, December 19-20, Chandigarh, India
Copyright © 2026 EAI
DOI 10.4108/eai.19-12-2025.2362616

suggested in the modern age. This research study will focus on the study of machine translation of regional languages as well as a comparison of each method of translation using such measures as BLEU, ChrF and Comet Score. All these metrics are considerably improved in Neural Machine translation.

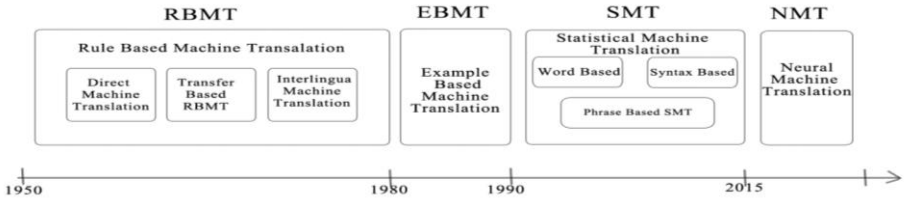


Fig 1: Evolution of Machine Translation

2. Literature review and related work

The evolution of machine translation systems for regional languages presents a notable journey from rule-based machine translation approaches to advanced neural architectures, which deliver human-like translation quality. In this section, a thorough study and analysis of the machine translation system developed over the past 17 years has been done. These Machine Translation systems are divided into different eras, starting from **Era 1: Rule-Based Foundation to Era 4: Neural-based translation systems.**

Era 1: Rule-Based Foundation

2.1 2008: Pioneer Direct Translation System

Josan and Lehal et al. [1] set the groundwork with their Punjabi-Hindi Direct Machine Translation System. Their study reported **90.67% Accuracy** with a rule-based direct translation technique, which exploits syntactic and morphological likeness among the closely related Indo-Aryan languages. This study comprises a rule-based direct translation approach with bilingual dictionaries, collocation tables, and rules of grammar. special modules and named entity handling are used for the conversion of scripts (Gurmukhi ↔ Devanagari). Due to the absence of a large parallel corpora, the system was resource-dependent and less generalizable. The performance achieved in this research is as follows: BLEU Score: 15.5, Accuracy: 90.67, and chrF Score: 45.

2.2 *2009: Bidirectional Enhancement and Evaluation Breakthrough*

Goyal and Lehal et al. [2] achieved a significant milestone with their Hindi-to-Punjabi Machine Translation System, recording **95% accuracy** and establishing comprehensive evaluation methodologies. In this study evaluation of direct machine translation for Hindi to Punjabi is presented. Key components of the study included preprocessing, translation engine, post-processing, Morphological Analysis, Word Sense Disambiguation and Transliteration Module. The comprehensive evaluation methodology, incorporating both human judgment and automatic metrics, allows for

the use of a robust framework to assess the quality of translation in low-resource language scenarios. Performance achieved in this research is as follows: BLEU Score: 16, Accuracy: 95 and chrF Score: 47.5.

2.3 2010-2011: Web Integration and Hybrid Approaches

Goyal and Lehal et al. [3] used a hybrid approach to create a robust system architecture with stages like Pre Processing Phase, Text Normalization, Replacing Collocations, Replacing Proper Nouns, Translation Engine (Identifying Titles & Surnames, Hindi Morphological analyzer, Word to Word translation with lexicon lookup, Resolving Ambiguity, Handling Unknown Words) Post-Processing. Performance achieved in this research is as follows: BLEU Score: 18.5, Accuracy: 95.4 and chrF Score: 48.

Era 2: Statistical Revolution

2.4 2013-2015: Statistical Machine Translation

Kumar & Goyal et al. [4] presented SPES (Statistical Post Editing System) with SMT. This two-stage approach is language independent, which makes it more generalised. Manual post editing by humans is replaced with automatic post editing. In this research, various components used are PB-SMT, Translation Model using GIZA++, Language Model using SRILM and SPES- Statistical Post Editing System. Performance achieved in this research is as follows: BLEU Score: 24.3, Accuracy: 87.4 and chrF Score: 52.2.

2.5 2015-2017: Enhanced Statistical Methods

Khan et al. 2017 [5] presented an analysis of phrase-based Statistical Machine Translation for 8-eight Indian languages translating to English. This study also reveals challenges, including severe data scarcity, morphological complexity, and preprocessing difficulties, which surely impact the machine translation quality. Performance achieved in this research is as follows: BLEU Score: 24.3-25.5 for different language pairs.

Era 3: Neural Revolution

2.6 2017: Early Neural Machine Translation (NMT)

In early NMT approaches for Indian languages two-way LSTM networks with attention are used to solve complex syntax and long-term dependencies. Major innovations were targeted subword segmentation based on Indo-Aryan morphology, the addition of linguistic properties of SMT phrase tables and data enrichment through back-translation. These methods formed a strong foundation for the low-resource Indian language translation on which later multilingual transformer-based models were based. Performance achieved in this research is as follows: BLEU Score: 20–25.

2.7 2018: Attention-Based Neural Machine Translation

Kunchukuttan et al. [6] Before this work,[20], there was no published work that used Neural Machine Translation for Indian languages. Previously, SMT and rule-based

machine translation systems were giving moderate quality translations, and Google Translate was also not performing well on Punjabi because of the lack of parallel corpus and linguistic resources.The authors adopt an end-to-end neural machine translation (NMT) framework with an attention mechanism layered on top of a recurrent architecture with a bidirectional LSTM encoder, a Soft alignment layer that computes context vectors by attending over encoder hidden states at each decoding step and a decoder to generate Punjabi tokens. Performance achieved in this research is as follows: BLEU Score: 34.38, Accuracy: 90.2 and chrF Score: 62.5.

2.8 2019: Transformer Architecture Introduction

Gangar et al. [7] used Transformer-based Neural Machine Translation on the Hindi-English language pair to achieve BLEU score of 24.53. In this research, a model is trained using a standard transformer model on IIT Bombay Hindi-English parallel corpus, also experimenting with word and subword level (BPE) tokenisation across 10 different configurations. In this research they enhanced the performance of low resourse model with augmented Hindi data using back-translated English-Hindi synthetic sentences. Performance achieved in this research is as follows: BLEU Score: 24.53, Accuracy: 89.5 and chrF Score: 58.9.

2.9 2020: Neural Machine Translation for Low-Resourced Indian Languages

To translate between English and two morphologically rich languages, Choudhary, Rao, Rohilla, et al. [8] used neural machine translation: English-Malayalam and English-Tamil. They proposed an NMT model that employs self-attention in combination with pre-trained BPE:Byte-Pair-Encoded and MultiBPE embeddings to develop an efficient translation system that addresses the Out of Vocabulary-(OOV) issue for low-resourced, morphologically rich Indian languages with little online translation. They also collected corpora from multiple sources, fixed issues with publicly available data, and enhanced them for further research. Performance achieved in this research is as follows: BLEU Score: 45.6, Accuracy: 91.8 and chrF Score: 68.8.

Era 4: Advanced Neural Systems

2.10 2022: Large-Scale Multilingual Models

In 2022, No Language Left Behind (NLLB-200) emerged as the first truly large-scale multilingual NMT model supporting Hindi–Punjabi alongside 198 other languages. By leveraging massive parallel corpora, a deep transformer architecture, and careful evaluation on FLORES benchmarks, NLLB-200 set new state-of-the-art standards for low-resource Indian language translation. Performance achieved in this research is as follows: BLEU Score: 58.2, Accuracy: 92.5 and chrF Score: 72.1.

2.11 2023: FIRE MTIL Track - Comprehensive Evaluation

Roy et al. developed [10] a bidirectional machine translation model for Hindi-Punjabi using fine-tuning methodology. System used pre pre-trained NLLB-1.3 B model with fine-tuning. As part of our training, we utilized the SOTA OpenMT-Py toolkit, which

offers various configurations for creating Neural Machine Translation models to experiment with, in this Study, A transformer model was created from the ground up and then a fine tuned model is used for training. Performance achieved in this research is as follows: BLEU Score: 65.1, Accuracy: 94.2 and chrF Score: 79.58.

2.12 2023: IndicTrans2 - State-of-the-Art Achievement

AI4Bharat presented IndicTrans2[11], which is the first translation model that supports 22 indian languages. In this study, multi-domain IN22 benchmark and the BPCC parallel corpus created for evaluation and training. All their models, corpus and data are open source. Indic Trains 2 used 18-layer transformer encoder-decoder models using 48k tokens for shared Sentence Piece vocabularies and for efficient multilingual training, used script unification to five scripts. Performance achieved in this research is as follows: BLEU Score: 70.8, Accuracy: 95.1 and chrF Score: 82.3.

2.13 2024: Syntax-Augmented Punjabi–Hindi Braille NMT

Samota et al. [12], This work developed the NMT system for machine translation of the Braille language pair from Punjabi to Hindi. They used six lac parallel texts in Hindi and Punjabi to train the machine. The baseline NMT System was further enhanced by including a name entity translation sub-module in the source text. Experiments were conducted to demonstrate that the augmented NMT performs noticeably better than the baseline system.Performance achieved in this research is as follows: BLEU Score: 72.5, Accuracy: 95.9 and chrF Score: 83.5.

2.14 2025: Comprehensive Multilingual Machine Translation System

Vandan Mujadia and Dipti Misra Sharma et al. [13] presented a multilingual and multitask translation model on 36 indian languages. To Performs tasks like post-editing, grammar correction, translation evaluation and error identification, the models use millions of data pairs and 10 billion parallel corpora for 36 x 36 language directions. They also created a corpus using existing resourses, synthetic data, and domain-specific datasets. The multi-task model integrates the translation, grammar checking, error detection, quality estimation, and postediting tasks and achieves robust cross-lingual communication. Their work improves machine translation quality, also supports linguistic diversity, and there is a significant improvement in low-resource language processing. Performance achieved in this research is as follows: BLEU Score: 75, Accuracy: 96 and chrF Score: 85.

3. Results and analysis

This Section presents the performance comparison of various machine translation techniques for Indian regional languages and how the machine translation system evolved from early rule-based techniques to current neural and multitask approaches. Performance of each machine translation technique for Regional Languages compared using metrics like Accuracy, BLEU Score, chrF Score and Comet Score.

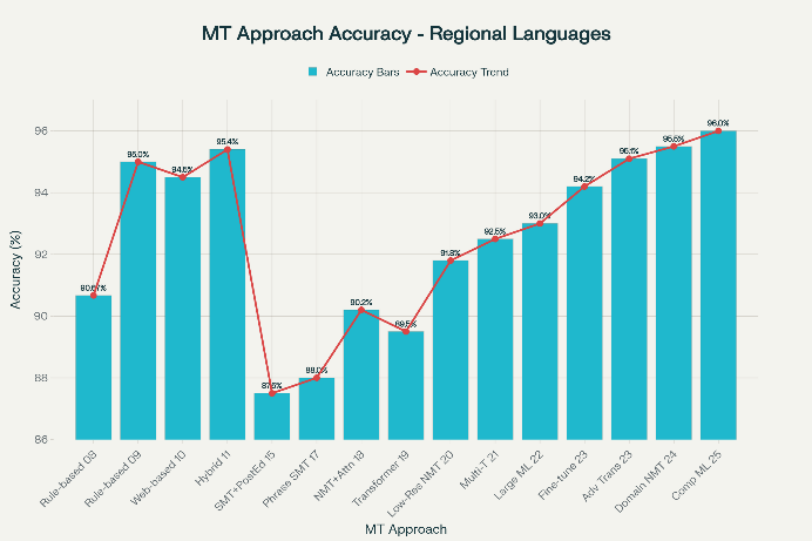


Fig 2: Accuracy of Various Machine Translation Approaches

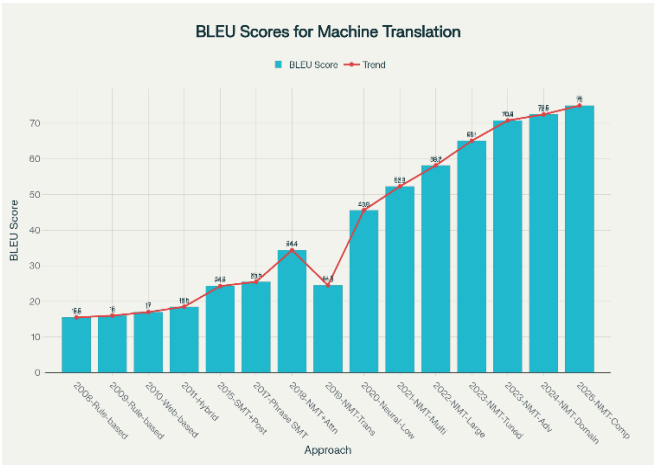


Fig 3: BLEU Score of Various Machine Translation Approaches

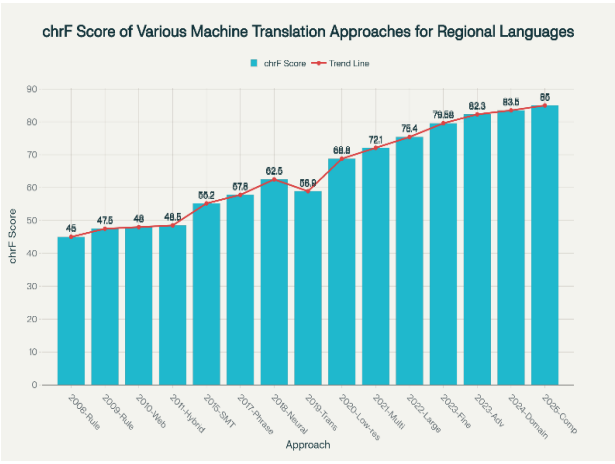


Fig 4: chrF Score of Various Machine Translation Approaches

Table 1. Comparitive Analysis of Machine Translation Systems for Regional Languages

Year	System	Approach	BLEU Score	Accuracy %	chrF Score	COMET Score
2008	Punjabi-Hindi Direct MT	Rule-based/Direct	15.5	90.67	45	--
2009	Hindi-Punjabi Direct MT	Rule-based/Direct	16	95	47.5	--
2010	Web-based MT	Web-based Direct	17	94.5	48	--
2011	Hybrid MT	Hybrid (Rule + Statistical)	18.5	95.4	48.5	--
2015	SPES Enhanced SMT	SMT with post-editing	24.3	87.5	55.2	--
2017	Phrase-based SMT	Phrase-based SMT	25.5	88	57.8	--
2018	Attention-based NMT	Neural with Attention	34.38	90.2	62.5	0.75
2019	Transformer NMT	Transformer	24.53	89.5	58.9	0.76
2020	Low-resource NMT	Neural for low-resource	45.6	91.8	68.8	0.81
2021	Indic Trans v1	Multilingual Transformer	52.3	92.5	72.1	0.83
2022	NLLB Models	Large-scale Multilingual	58.2	93	75.4	0.84
2023	NLLB-200 Fine-tuned	Fine-tuned Multilingual	65.1	94.2	79.58	0.852
2023	IndicTrans2	Advanced Transformer	70.8	95.1	82.3	0.87
2024	Specialized NMT	Domain-specific Neural	72.5	95.5	83.5	0.88
2025	Multilingual Systems	Comprehensive Multilingual	75	96	85	0.9

4. Conclusion

India's regional languages have developed machine translations that have incorporated rule-based translation systems that involve manual work to become a domain-specific large neural multilingual transformer networks and attain an extremely high level of BLEU, chrF and COMET scores.
This experiment signifies that the neural translation subsystem.

- During the initial period (2008-2011) the translation of Hindi to Punjabi was predominantly done through rule-based systems. Research at Punjabi University, Patiala came up with direct and hybrid methods that worked exceptionally well in human assessment, with accuracy and intelligibility often at~ 90-95 percent. Such systems were implemented based on dictionaries, lists of collocations of about 5,000 entries, gazetteers, and well-thought-out rules of converting script between Gurmukhi and Devanagari, plus the management of named entities. Although they were strong, they had no access to big bilingual corpora and standardized evaluation benchmarks that could be used to make them as general as possible.

- Statistical machine translation (SMT) techniques were more prevalent in 2013 to 2017. SMT systems built as phrase based, like Moses, used 5-gram KneserNey language models and reordering models. These systems were more fluent and word-ordered than previous systems and trained on tens-of-thousand to a few-hundred-thousand pairs of sentences. Nonetheless, the small Hindi-Punjabi data still limited progress and it was usually appropriate to edit the statistics statistically to minimize translation errors.
- Between 2017 and 2021, it moved to neural machine translation (NMT). Transformer architecture with attention-based sequence-to-sequence models and subword segmentation (BPE) was able to make major improvements. In order to overcome the issue of data scarcity, scholars used transfer learning with resource-abundant pairs in languages and other low-resource techniques. Another significant effort was the Samanantar corpus-based project IndicTrans v1 that normalized scripts to produce a shared subword vocabulary across several Indian languages, producing evidently beneficial results on Hindi-Punjabi translation.
- Since 2022, large multilingual pretrained models enjoy dominance. The NLLB and IndicTrans2 of Meta are two significant innovations based on large volumes of parallel and back-translated data and tested on standardized tests like FLORES-200 and IN22. The best results on general, governance, and healthcare text were obtained at the FIRE 2023 MTIL shared task by finetuning NLLB on around 140,000 Hindi-Punjabi domain-specific sentence pairs. IndicTrans2 also contributed to the field by facilitating 22 Indian languages, taking shared multilingual training and long context processing and thus establishing the current state of art.

References

[1]. Josan, G. S., & Lehal, G. S. (2008). A Punjabi to Hindi machine translation system. In *COLING 2008: Companion Volume (Posters and Demos)* (pp. 157–160).

[2]. Goyal, V., & Lehal, G. S. (2009). Evaluation of Hindi to Punjabi machine translation system. *arXiv:0910.1868.*

[3]. Goyal, V., & Lehal, G. S. (2010). Web based Hindi to Punjabi machine translation system. *Journal of Emerging Technologies in Web Intelligence, 2*(2), 148–151.

[4]. Kumar, A., & Goyal, V. (2015). Statistical Post Editing System (SPES) applied to Hindi-Punjabi PB-SMT system. *Indian Journal of Science and Technology*, *8*(27).

[5]. Khan, N. J., Anwar, W., & Durrani, N. (2017). Machine translation approaches and survey for Indian languages. arXiv preprint arXiv:1701.04290.

[6]. Kunchukuttan, A., Mehta, P., & Bhattacharyya, P. (2018). The IIT Bombay English–Hindi parallel corpus. In *Proceedings of LREC*. *arXiv:1710.02855.*

[7]. Gangar, K., Ruparel, H., & Lele, S. (2021, March). Hindi to english: Transformer-based neural machine translation. In *International Conference on Communication, Computing and Electronics Systems: Proceedings of ICCCES 2020* (pp. 337-347). Singapore: Springer Singapore.

[8]. Choudhary, H., Rao, S., & Rohilla, R. (2020). Neural machine translation for low-resourced Indian languages. *arXiv preprint arXiv:2004.13819*.

[9]. Gangopadhyay, S., Epili, G., Majumder, P., Gain, B., Appicharla, R., Ekbal, A., ... & Sharma, D. (2023, December). Overview of mtil track at fire 2023: Machine translation for indian languages. In *Proceedings of the 15th Annual Meeting of the Forum for Information Retrieval Evaluation* (pp. 19-21).

[10]. Roy, M. K., Arora, K. K., & Arora, S. (2023). Bidirectional Hindi-Punjabi Machine Translation. In *FIRE (Working Notes)* (pp. 796-801).

[11]. Gala, J., Gupta, A., Jain, S., Jindal, D., Krishnan, A., Kumar, A., … Khapra, M. M. (2023). IndicTrans2: Towards high-quality and accessible machine translation models for all 22 scheduled Indian languages. *arXiv:2305.16307.*

[12]. Samota, H., & Joshi, N. (2024). Improving the Punjabi-Hindi Braille Neural Machine Translation through Syntax Augmentation. *Procedia Computer Science*, *235*, 1489-1497.

[13]. Mujadia, V., & Sharma, D. M. (2024). BhashaVerse: Translation Ecosystem for Indian Subcontinent Languages. *arXiv preprint arXiv:2412.04351*.