

¹ Vijay Singh
Sen
²Dr.
Shubhnandan
S. Jamwal

Transcending Linguistic Boundaries: A Technical Exploration of The Evolution and Future Trajectory of Corpus-Based Machine Translation



Abstract: - This review paper presents a comprehensive examination of the evolution, current state, and future perspectives of machine translation (MT). The field of MT, encompassing the automatic translation of text from one language to another, has undergone significant transformations from rule-based methods to statistical models, and more recently, to neural network-based approaches. By synthesizing current research and technological progress, this paper aims to provide a detailed understanding of the dynamic landscape of machine translation and its future directions. In today's globalized world, language barriers remain a significant obstacle to accessing information. Relying solely on human translators often fails to meet the growing demand for translation services, which is why Machine Translation (MT) tools are becoming increasingly popular. These tools have the potential to address this issue effectively. As a result, research in MT is expanding rapidly, and new paradigms are continually being developed. This paper presents a systematic literature review to identify the most widely used MT systems, their architectures, the quality assessment methods used to evaluate them, and which systems deliver the best results.

Keywords: machine translation, neural machine translation, statistical machine translation, rule-based translation, low-resource languages.

1. Introduction:

Machine translation (MT), the automated process of converting text from one language to another, has evolved remarkably since its inception. The journey of MT began in the 1950s with rule-based systems that relied heavily on linguistic rules and handcrafted dictionaries. These early systems, though innovative for their time, faced significant limitations due to their inability to scale and handle the complexities of human languages [1]. The advent of statistical machine translation (SMT) in the late 1980s marked a paradigm shift in the field.

SMT leveraged large corpora of bilingual text to generate translations based on probabilistic models, significantly improving translation quality by utilizing real-world data. However, SMT systems were still hampered by issues such as handling idiomatic expressions and achieving fluent, contextually accurate translations [2]. The introduction of neural machine translation (NMT) in the mid-2010s brought about another revolutionary change. NMT systems use deep learning techniques, particularly encoder-decoder architectures with attention mechanisms, to model entire sentences in a more holistic manner. This approach has led to substantial improvements in translation accuracy and fluency, making NMT the dominant paradigm in contemporary MT research and applications. Despite these advancements, several challenges remain unresolved. Low-resource languages, which lack extensive parallel corpora, still pose significant hurdles for MT systems. Additionally, ensuring contextual and cultural accuracy in translations is an ongoing challenge. The need for domain adaptation and the handling of specialized jargon further complicate the development of robust MT systems [3]. Recent advancements and future trends in MT include the development of multilingual models that can handle multiple languages within a single framework, unsupervised and semi-supervised learning techniques that reduce the dependency on large parallel corpora, and the ethical considerations surrounding AI-driven translation technologies. These emerging trends highlight the dynamic and rapidly evolving nature of the MT field.

2. Literature Review:

This section primarily deals with research published in the area of machine translation over a period of time. Many machine translation systems have been developed for different pair of languages across the world. Most popular research in the development of MT includes the one developed by R. A. Dam & A. Guessoum [4] where an English-to-Arabic Machine Translation System employing Bilingual Corpus has been developed. They introduced a transfer-based method using Artificial Neural Networks for machine translation that was trained to learn correspondence between source and target language structures. They claimed to get encouraging experimental results. Other MTS for Hindi-Punjabi using hybrid MT approach has been developed by V. Goyal & G. S. Lehal [5]. The key stages of MT system developed are pre-processing, translation engine and post processing. The accuracy of the system has been evaluated using accuracy test, intelligibility test and BLEU score. They claimed that hybrid system performs better than the constituent systems. Other interlingua based RBMT system that translated Sanskrit to English language was developed by H. S. Sreedeepta & S. M. Idicula [6]

¹ Department of Computer Science and IT, University of Jammu

² Department of Computer Science and IT, University of Jammu senvju@gmail.com, jamwalsnj@gmail.com

Copyright © JES 2024 on-line : journal.esrgroups.org

in Paninian framework. The proposed system uses lexical functional grammar at different levels of analysis phase. In this system Sanskrit text was first converted to an intermediate notation called Interlingua. This notation was mapped to the target language, and translated output was generated in English language.

M. A. Ayu & T. Mantoro [7], developed an EBMT based Bahasa Indonesia to English machine translation system. The system was developed by employing Moses system. Experiments in translating Bahasa Indonesia to English by tuning the parameters in Moses decoder like manipulating the weight on translation model, language model, distortion (re-ordering) and word penalty is done in order to increase the quality of the translation.

A. H. Imam, et al. [8] developed an English-Bangla bilingual corpus and studied the impact on Statistical Machine Translation systems for the given language pair. Other SMT based translation system developed by B. N. V. N. Raju and M. S. V. S. B. Raju [9] translated Telugu into English language. The decoder in SMT used Language Model (LM) and Translation Model to generate the translation. They have proposed that if the source language sentence is given, LM can compute the probability of target language sentence and maximization of the probability of translated text is done by Moses. Also S. Kharb, et al. [10] performed a statistical analysis of translation systems for the evaluation of the efficiency of translation systems based on the lexical, syntactical and semantics differences between source and translated text. They have emphasized on translation between English and Hindi.

U. Singh, et al. [11] developed an Urdu to Punjabi Statistical Machine Translation system. A variety of rules were framed for tokenization and segmentation in pre-processing stage. Hidden Markov Model (HMM) and Viterbi algorithm has been used for training and decoding the model respectively. Authors claimed to have achieved 0.86 BLEU score and more than 85% accuracy. Also E. Warhade & M. R. Devale [12] have developed a Phrase-Based Statistical Machine Translation system for English to Sanskrit translation. Language and Translation model were trained using monolingual Sanskrit corpus and bilingual Sanskrit-English corpus. S. Kumar, et al. [13] have proposed a statistical approach that performs phrase-based machine translation, for the development of Kannada to English translation system and whole of the experimentation has been done using Moses toolkit. Giza tool was used for sentence alignment and training, tuning and testing of the system was done on a parallel corpus of 20000 Kannada-English bible sentences. The system claimed BLEU Score of 0.5.

P. Das, et al. [14] have developed a translation system on corpus-based approach. Language, Decoder, translation and transliteration are the four models that constitute the proposed system. An Assamese-English parallel corpus of 8000 sentences was used in the research. The corpus contained data of travel and tourism in India and the system claimed to have achieved a BLEU score of 0.63.

P. Salunkhe et al. [15], developed hybrid English-Marathi machine translator that translated web pages, agricultural text, medical reports in Marathi and tourism related information. The proposed system consisting of Parallel Multi-Engines, performed statistical and rule based translation and the system claimed to produce state-of-the-art results. For statistical evaluation, Mapper algorithm has been used in rule based Translation. In order to enhance dictionary and incorporate better translation result Marathi wordnet has been implemented. Other hybrid based translation model was developed by M. M. Rahman et al. [16], for translating Bengali to English with the help of N-gram language model. The corpus based method finds the corresponding target language translation of sentence fragments, selecting the best match text from the parallel corpus while the N-gram model rearranges the sentence constituents to get an accurate translation without employing external linguistic rules. Another Hybrid based machine translation system combining RBMT and EBMT, was developed by P. Seresantakul and P. Unlee [17], presented the development of the Thai language and Isarn dialect bilingual parallel corpus, that included word segmentation, translation and word alignment, parts of speech (POS) tagging, and the parallel corpus design and construction. Sentences in Thai are segmented into a sequence of words by applying a Conditional Random Field (CRF) approach. They used the rule and example based Thai-Isarn machine translation system as a tool to generate the corresponding Isarn dialect. Hidden Markov Modeling (HMM) is employed for tagging each word. The source and target sentences with their POS tags are validated by experts in both Thai language and Isarn dialects.

Y. M. ShweSin, et al. [18] used large scale parallel corpus for Myanmar to English Translation system based on Neural Machine Translation. Experimentation on word-level model and character-level model was based on neural method for Myanmar to English translation. The evaluation results showed that neural machine translation models improved the performance of Myanmar to English translation. Also K. Revanuru, et al. [19] applied NMT techniques to create a system with multiple models. The model is applied for six Indian language pairs and performance evaluation of NMT models is done using automatic evaluation metrics such as UNK Count, METEOR, F-Measure, and BLEU. They claim that NMT techniques are very effective for machine translations of Indian language pairs. Other corpus-based NMT translation system was developed by M. Singh, et al. [20] that takes Bhagwad Gita text as input and performs Sanskrit to Hindi translation. Deep neural network has been used for training the model and after data analysis auto-tuning was performed. Authors claimed that better BLEU Score and Word Error Rate have been achieved.

T. M. Oo, et al. [21] developed a Neural Machine Translation between Myanmar (Burmese) and Dawei (Tavoyan) language. They first developed Myanmar-Dawei parallel corpus which was implemented using two prominent

neural machine translation systems, RNN and Transformer with syllable segmentation. Authors claimed that Long Short Term

Memory (LSTM) with RNN architecture works best for Dawei-Myanmar and MyanmarDawei neural machine translation. Also H. Sun, et al. [22] proposed an Unsupervised

Neural Machine Translation (UNMT) structure with cross-lingual language model (CMLM) representation agreement to capture the communication between UNMT and unsupervised bilingual word embedding (UBWE)/CMLM during training of UNMT. Authors claimed that experimental results demonstrated that the proposed UNMT models improved significantly over the corresponding state-of-the-art UNMT baselines. Another NMT based system was developed by S. Saini & V. Sahula [23] and the system claimed to have better results than conventional translation techniques. Authors reported that NMT can be trained on small amount of data size and can exhibit satisfactory translation results. Another NMT system was developed by S. Singh, et al. [24], based on neural machine translation for English-Punjabi language pair. BLEU evaluation metric was used to evaluate the proposed system which claimed to provide better result than Google Translate.

E. Greenstein & P. Daniel [25] developed a machine translation system that performed translation from Japanese to English language using RNN (Recurrent Neural Networks) based approach. Two important units of the proposed model were Encoder and GRU (Gated Hidden Unit), which in turn were bidirectional recurrent neural networks. GRU acted as the activation function of the hidden layers in the network and the same recurrent neural network and gated hidden unit were used as decoder of the model. The model has been trained on 150000 corpora and the system claimed BLEU score of 0.73.

R. Narayan, et al. [26] proposed Quantum Neural Machine Translation model that performed English to Hindi language translation using quantum recurrent neural network based approach. In this model a rule based parts of speech tagger and quantum neural network based parts of speech tagger was used that increased the accuracy of the proposed model. Proposed system claimed to attain 0.9814 BLEU score and 98.26% accuracy when human language expert evaluation was done. B. Mishra, et al. [27] developed an English to Hindi neural machine translation system that implemented both feed forward and back propagation artificial neural network approaches. The whole translation process of the proposed model contained nine modules wherein neural networks were used to create the knowledge base and bilingual dictionary mapping. Proposed system has been implemented in java and Matlab and a BLEU score of 0.604 has been claimed by the system. Also, M. Johnson, et al. [28] developed a multilingual Neural Machine Translation (NMT) system. The proposed cross lingual translation model has encoder, decoder model implemented with attention mechanism. A single model performs translation for different language pairs without having any major change in the basic structure of proposed model. System claimed that multilingual model attained comparable performance for English to French translation and gave better than the existing translation quality for French→English, German→English and English→German language pairs.

3. Approaches for Machine Translation:

Language translation, the art of conveying meaning from one language to another, has undergone a remarkable evolution over the years. As our world becomes increasingly interconnected, the demand for accurate and efficient translation methods has never been higher. In this exploration of language translation techniques, we delve into the diverse approaches used to bridge linguistic divides and facilitate effective communication across cultures. From traditional methods rooted in linguistic rules to modern advancements driven by data and artificial intelligence, this journey unveils the intricacies and innovations shaping the field of translation today. In our discussion, we will explore a spectrum of translation methods, ranging from rule-based approaches reliant on linguistic principles to data-driven techniques harnessing the power of machine learning. Each method offers unique strengths and considerations, shaping how languages are transformed and interpreted in the global exchange of ideas and information. Some of the prominent approaches of machine translation are:

- RBMT (Rule Based Machine Translation)
- EBMT (Example Based Machine Translation)
- SMT (Statistical Machine Translation)
- NMT (Neural Machine Translation)
-

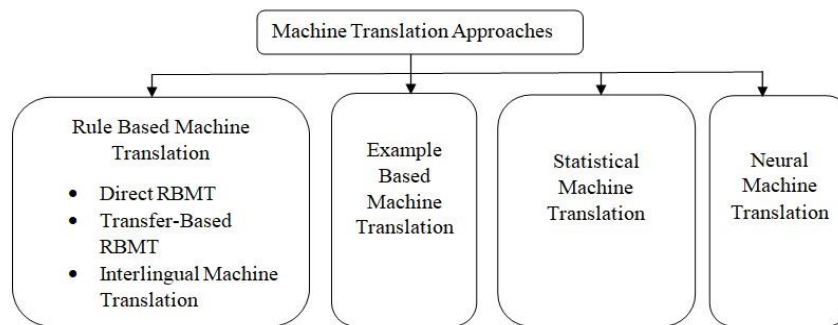


Figure 1: Classification of Machine Translation Methods

3.1. Rule Based Machine Translation (RBMT): For each language pair, rule-based machine translation depends on innumerable built-in linguistic rules and bilingual dictionaries. RBMT parses text and produces a transitional representation from which target language text is generated. This process requires detailed syntactic and semantic knowledge of lexicons along with broad set of rules for the language pair [29]. Some of the RBMT based MT systems are Systran, Eurotra, Apertium and GramTrans.

Rule-based machine translation systems are of three types:

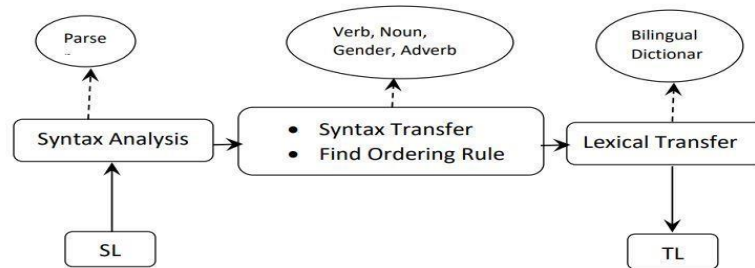


Figure 2: Basic Architecture for Rule Based Machine Translation system

3.1.1. Direct RBMT Systems: These include dictionary based machine translation that map input and output employing basic rules. Some of the MT systems that followed Direct approach are: Punjabi to Hindi MT system developed at Punjabi

University [30] attained an accuracy of 90.67%. Other MT system ‘Anusaaraka’ developed by IIT Kanpur and IIIT Hyderabad translated many Indian Languages like Punjabi, Kannada, Telugu, Marathi and Bengali to Hindi language [31]. Another Hindi-to-Punjabi Translation System developed at Punjabi University, Patiala [32] claimed accuracy of 87.60%.

3.1.2. Transfer based RBMT System: The Transfer Based machine translation also named as linguistic knowledge or indirect translation uses morphological, syntactic and semantic features of language pair. In comparison to the direct MT systems, the transfer based MT splits the translation into three steps: the assessment of source language text in order to identify its grammatical structure, resulting structure is transferred to the grammatical structure of target language text, and finally, the generation of target language text. MANTRA developed by Hemant Darbari et al. [33] translated English to Indian languages (Hindi, Bengali, Telugu and Gujarati) is a rule based machine translation system. Another transfer based system is English to Bangla Machine Translation System developed by Md. Golam et al. [34].

3.1.3. Interlingua based RBMT Systems: Machine Translation systems wherein source language text is transformed into an intermediate language-independent representation (interlingua), which is then converted into target language text. Some of the MT systems that followed Transferred based RBMT approach are: ANGLABHARTI MT system [35] translated text related to public health domain from English to Indian Languages using pseudo-interlingua approach, UNL-based English-Hindi MT System developed by Dungarwal, et al. at IIT Mumbai [36] translated from English and Hindi languages to Indian languages (Hindi, Bengali, Marathi).

3.2. Example Based Machine Translation (EBMT):

EBMT can also be called as translation by analogy, as bilingual corpus having parallel text which acts as the knowledge base at run time. It is based on the assumption that humans translate by first breaking down a sentences into phrases, phrases into sub-phrases or words, then translating these phrases, and eventually by appropriately assembling these fragments to form one complete sentence. Malayalam to English [30], English to Hindi [31] and

Bengali-Assamese Bilingual translation system [32], are some of the machine translation systems developed using Example-based Approach.

Advantages of RBMT include no bilingual text required, Domain-independence, total control (a possible new rule for every situation), reusability (existing rules of languages can be transferred when paired with new languages) but simultaneously more the rules the harder to deal with the system.

3.3. Statistical Machine Translation (SMT):

The statistical machine translation approach including the ideas of applying Shannon's information theory was first introduced by Warren Weaver in 1949. In late 1980s and early 1990s, statistical machine translation approach was re-introduced by researchers at the Thomas J. Watson Research Center of IBM and has led to the important revival of corpus based approach in machine translation. SMT approach is centered on the assumption that every sentence in source language has possible translation into target language. In SMT system learning algorithms are applied to a significant set of parallel corpus. The accuracy of SMT system relies greatly on two important factors that are quantity and quality of the parallel corpora. SMT model gives better results for closely related languages in specific domains or for language-pair having huge amount of corpus. Conditional probability is used to find the appropriate translation in SMT. Say, t is target language and s is source language, according to Bayes' theorem $P(t|s)$ can be represented as:

$$P(t|s) = \frac{P(s|t)P(t)}{P(s)}$$

The statistical machine translation involves three main components Language model, Translation model and Decoder [37, 38]. The job of Language model is to perform the computation of the probability of the target language sentence ' t ' as $p(t)$ using conditional probability. For a given target sentence ' t ' computation of the probability of source sentence ' s ' i.e $P(t|s)$ is determined by the translation model and finally Decoder maximizes the product $P(s|t) * P(t)$.

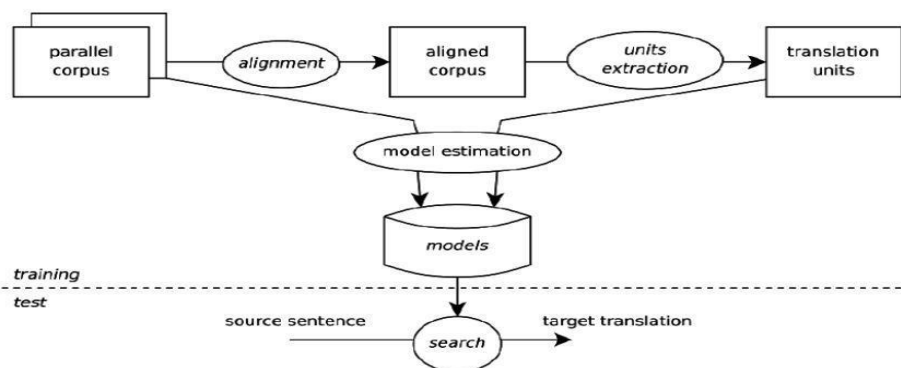


Figure 3: Basic Architecture for Statistical Machine Translation system

English to Sanskrit [39], Assamese to English [11], English to kannada [10] and Urdu to Punjabi [8] are some of the machine translation systems developed using statistical machine translation approach.

3.4. Hybrid Approach:

Hybrid approach involves the application of two or more approaches in developing a machine translation system. Although most of the system developed using hybrid approach apply SMT and rule based approaches. English to Telugu Machine Translation System [40], English to Indian Languages involving Malayalam, Tamil and Hindi [41] and Bengali to Hindi Machine Translation System [42][43] are some hybrid based machine translation systems.

3.5. Neural Machine Translation (NMT):

With the advent of deep learning, Neural Machine Translation (NMT) has become a robust approach for performing machine translation in recent years. NMT approach uses neural network to compute the probability of a sequence of words so that a single integrated model can give the appropriate translation results. NMT system consists of two main components, Encoder and Decoder. Encoder takes the source language input text and performs vectorization i.e assigning number to text data and representation of data in matrix form so that computation on data can be performed. Decoder then interprets the vectorized text which is the representation of data in matrix form and translates it to the target language text. Encoder and Decoder are themselves neural networks which can be either feed-forward neural network (FFNN), neural network using back propagation or

recurrent neural network (RNN). Many researchers prefer RNN over basic FFNN in developing machine translation systems because of its state-of-the-art translation. Neural Machine Translation for English to Hindi [20], Attention based English to Punjabi neural machine translation [21], Neural Machine Translation of Indian Languages [16] and Deep Neural Network based Sanskrit to Hindi translation system [17] are some of the recently developed NMT based translation systems. NMT has many advantages over SMT [37][38].

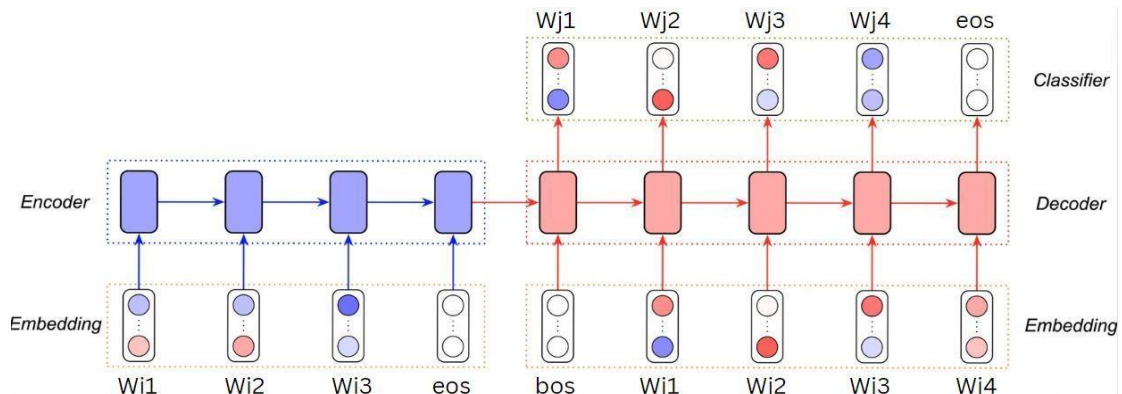


Figure 4: Basic Architecture for Neural Machine Translation system

- The SMT system consists of several components tuned separately. In contrast, The NMT model is a large end-to-end single network that consists of two sub recurrent neural network: the encoder and the decoder.
- While the SMT system needs many features that are accurately defined to do the translation, the NMT model depends on a training corpus to learn the translation task, with less or no feature engineering effort by linguists or engineers.
- In contrast to SMT, NMT can seize potential long-distance dependencies and complicated word alignment information
- The NMT model does not require a large memory space, such as those used by the SMT to store a translation model, a reordering model and a language model.

Conclusion

This review paper aims to provide a comprehensive overview of the evolution, current state, and future perspectives of machine translation. By examining the historical milestones, technical advancements, and persistent challenges, this paper seeks to offer insights into the dynamic landscape of MT and its potential future directions. The continued development of innovative approaches and technologies holds promise for overcoming existing limitations and enhancing the effectiveness and accessibility of machine translation for diverse global audiences.

References:

- [1] Stahlberg, Felix. (2020). Neural Machine Translation: A Review. *Journal of Artificial Intelligence Research*. 69. 343-418. 10.1613/jair.1.12007.
- [2] Lopez, Adam. (2008). Statistical machine translation. *ACM Comput. Surv.* 40. 10.1145/1380584.1380586.
- [3] Mohamed, Yasir & Khanan, Akbar & Bashir, Mohamed & Mohamed, Abdul & Adiel, Mousab & Elsadig, Muawia. (2024). The Impact of Artificial Intelligence on Language Translation: A Review. *IEEE Access*. 12. 25553-25579. 10.1109/ACCESS.2024.3366802.
- [4] R. A. Dam & A. Guessoum, "Building a neural network-based English-to-Arabic transfer module from an unrestricted domain," *International Conference on Machine and Web Intelligence, Algiers*, pp. 94-101, 2010. doi: 10.1109/ICMWI.2010.5648157.
- [5] V. Goyal & G. S. Lehal, "Hindi to Punjabi machine translation system," *In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Systems Demonstrations (HLT '11). Association for Computational Linguistics, USA*, 1-6, 2011.
- [6] H. S. Sreedeepea & S. M. Idicula, "Interlingua based Sanskrit-English machine translation," *International Conference on Circuit, Power and Computing Technologies (ICCPCT), Kollam*, pp. 1-5, 2017. doi:10.1109/ICCPCT.2017.8074251.
- [7] M. A. Ayu & T. Mantoro, "An Example-Based Machine Translation approach for Bahasa Indonesia to English: An experiment using Moses," *IEEE Symposium on Industrial Electronics and Applications, Langkawi*, pp. 570-573, 2011. doi: 10.1109/ISIEA.2011.6108777.
- [8] A. H. Imam, M. R. M. Arman, S. H. Chowdhury & K. Mahmood, "Impact of corpus size and quality on English-Bangla statistical Machine Translation system," *14th International Conference on Computer and Information Technology (ICCIT), Dhaka*, pp. 566-571, 2011. doi: 10.1109/ICCITechn.2011.6164853.

- [9] B. N. V. N. Raju & M. S. V. S. B. Raju, "Statistical Machine Translation System for Indian Languages," *IEEE 6th International Conference on Advanced Computing (IACC), Bhimavaram*, pp. 174-177, 2016. doi: 10.1109/IACC.2016.41.
- [10] S. Kharb, H. Kumar, M. Kumar & A. K. Chaturvedi, "Efficiency of a machine translation system," *International conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore*, pp. 140-148, 2017. doi: 10.1109/ICECA.2017.8203660.
- [11] U. Singh, V. Goyal & G. Singh, "Urdu to Punjabi Machine Translation: An Incremental Training Approach," *International Journal of Advanced Computer Science and Applications*, 2016. doi: 7.10.14569/IJACSA.2016.070428.
- [12] E. Warhade & M. R. Devale, "Design Of Phrase-Based Decoder For English-To-Sanskrit Translation," *Journal of Global Research in Computer Sciences*, 3, 35-38, 2012.
- [13] S. Kumar, B. N. Namitha & R. Nithya, "A comparative study of english to kannada baseline machine translation system with general and bible text corpus", *International Journal of Applied Engineering Research*, vol 10, pp. 30195-30201 2015. doi: 10. 30195-30202.
- [14] P. Das, K. Baruah, A. Hannan, & S. Sarma, "Assamese-English Bilingual Machine Translation," *International Journal on Natural Language Computing*, 3, 2014. doi: 10.5121/ijnlc.2014.3307.
- [15] P. Salunkhe, A. D. Kadam, S. Joshi, S. Patil, D. Thakore & S. Jadhav, "Hybrid machine translation for English to Marathi: A research evaluation in Machine Translation: (Hybrid translator)," *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), Chennai*, pp. 924-931, 2016. doi: 10.1109/ICEEOT.2016.7754822.
- [16] M. M. Rahman, M. F. Kabir & M. N. Huda, "A Corpus Based N-gram Hybrid Approach of Bengali to English Machine Translation," *21st International Conference of Computer and Information Technology (ICCIT), Dhaka, Bangladesh*, pp. 1-6, 2018. doi: 10.1109/ICCITECHN.2018.8631938.
- [17] P. Seresangtakul & P. Unlee, "Thai-Isarn Dialect Parallel Corpus Construction for Machine Translation," *11th International Conference on Knowledge and Smart Technology (KST), Phuket, Thailand*, pp. 121-125, 2019. doi: 10.1109/KST.2019.8687534.
- [18] Y. M. ShweSin, K. M. Soe & K. Y. Htwe, "Large Scale Myanmar to English Neural Machine Translation System," *IEEE 7th Global Conference on Consumer Electronics (GCCE), Nara*, pp. 464-465, 2018. doi: 10.1109/GCCE.2018.8574614.
- [19] K. Revanuru, K. Turlapaty & S. Rao, "Neural Machine Translation of Indian Languages," *In Proceedings of the 10th Annual ACM India Compute Conference (Compute '17). Association for Computing Machinery, New York, NY, USA*, 11-20, 2017. doi: 10.1145/3140107.3140111.
- [20] M. Singh, R. Kumar, & I. Chana, "Corpus based Machine Translation System with Deep Neural Network for Sanskrit to Hindi Translation," *Procedia Computer Science*, 167, 25342544, 2020.
- [21] T. M. Oo, Y. K. Thu, K. M. Soe & T. Supnithi, "Neural Machine Translation between Myanmar (Burmese) and Dawei (Tavoyan)," *IEEE Conference on Computer Applications (ICCA), Yangon, Myanmar*, pp. 1-7, 2020. doi: 10.1109/ICCA49400.2020.9022813. H. Sun, R. Wang, K. Chen, M. Utiyama, E. Sumita & T. Zhao, "Unsupervised Neural Machine Translation With Cross-Lingual Language Representation Agreement," *IEEE/ACM Trans. Audio, Speech and Lang. Proc.* 28, 1170-1182, 2020. doi: 10.1109/TASLP.2020.2982282.
- [22] S. Saini & V. Sahula, "Neural Machine Translation for English to Hindi," *Fourth International Conference on Information Retrieval and Knowledge Management (CAMP) IEEE*, 1-6, 2018. doi: 10.1109/INFRKM.2018.8464781.
- [23] S. Singh, M. Kumar & K. Soman, "Attention based English to Punjabi neural machine translation" *Journal of Intelligent & Fuzzy Systems*, 34, 1551-1559, 2018. doi: 10.3233/JIFS169450.
- [24] E. Greenstein & Daniel Penner, "Japanese-to-English Machine Translation Using Recurrent Neural Networks," 2015.
- [25] R. Narayan, S. Chakraverty & V. P. Singh, "Quantum neural network based machine translator for English to Hindi" *Applied Soft Computing Journal*, pp. 1-16, 2015.
- [26] S. Shahawaz & R. B. Mishra, "A Neural Network based Approach for English to Hindi Machine Translation" *International Journal of Computer Applications*, pp. 50-56, 2012.
- [27] Melvin Johnson, Mike Schuster, Quoc Le & Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, G.S. Corrado, Macduff Hughes & Jeffrey Dean, "Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation," *Transactions of the Association for Computational Linguistics*, 2016.
- [28] Farhat, A., & Al-Taani, A. (2015). A rule-based English to Arabic machine translation approach. In international Arab conference on information technology (ACIT'2015).
- [29] G. S. Joshan & G. S. Lehal, "Punjabi to Hindi machine translation system," *In Proceedings of the 22nd International Conference on Computational Linguistics, Manchester, UK*, pp. 157160. 2010.
- [30] A. Bharati, V. Chaitanya, A. Kulkarni & R. Sangal, "Anusaaraka: Machine Translation in Stages," 2003. doi: arxiv preprint cs/0306130.
- [31] V. Goyal & G. S. Lehal, "Hindi to Punjabi Machine Translation System," *Information Systems for Indian Languages. ICISIL Communications in Computer and Information Science*, vol 139. Springer, Berlin, Heidelberg, 2011. doi: 10.1007/978-3-642-19403-0_40.
- [32] H. Darbari, A. Kumar, A. Dasgupta & K. S. Mishra, "Complexity of Language in Rajya Sabha Domain and MANTRA Approach," *Proceedings of ICON-2010: 8th international conference on natural language processing*, 2010. Golam Md, Alam, Md Golam Rabiul Islam & Md Nowrin Islam, "A New Approach To Develop An English To Bangla Machine Translation System," *Daffodil International University Journal of Science and Technology*, 2011.
- [33] R. M. K. Sinha, K. Sivaraman, A. Agrawal, R. Jain, R. Srivastava & A. Jain, "Anglabharti: a multilingual machine aided translation project on translation from English to Indian languages," *IEEE International Conference on Systems, Man*

- and Cybernetics, Intelligent Systems for the 21st Century, Vancouver, Canada, pp. 1609-1614 vol.2, 1995. doi: 10.1109/ICSMC.1995.538002.
- [34] Dungarwal, P., Chatterjee, R., Mishra, A., Kunchukuttan, A., Shah, R.M., & Bhattacharyya, P. (2014). The IIT Bombay Hindi-English Translation System at WMT 2014. WMT@ACL.
- [35] Wang X, Tu Z, Zhang M (2018) Incorporating Statistical Machine Translation Word Knowledge Into Neural Machine Translation. *IEEE/ACM Transact on Audio, Speech, Lang Process* 26(12):2255–2266
- [36] 8 -Yang Z, Chen W, Wang F, Xu B (2018) Generative adversarial training for neural machine translation. *Neurocomputing* 321:146–155
- [37] S. R. Warhade, S. H. Patil & P. R. Devale, “English-to-Sanskrit statistical machine translation with ubiquitous application,” *International Journal of Computer Applications*, 51(1), 2012.
- [38] K. Lingam, E. Ramalakshmi & S. Inturi, “English to Telugu Rule based Machine Translation System: A Hybrid Approach” *International Journal of Computer Applications*, 101(2), pp. 19-24. 2014.
- [39] Sangeetha, S. Jothilakshmi & R. Kumar, “An Efficient Machine Translation System for English to Indian Languages Using Hybrid Mechanism,” *International Journal of Engineering and Technology*, 6(4), pp. 1909-1919, 2014.
- [40] S. Chatterji, D. Roy, S. Sarkar & A. Basu, “A Hybrid Approach for Bengali to Hindi Machine Translation,” *7th International Conference on Natural Language Processing*, 2009.
- [41] Sangeetha, S. Jothilakshmi & R. Kumar, “An Efficient Machine Translation System for English to Indian Languages Using Hybrid Mechanism,” *International Journal of Engineering and Technology*, 6(4), pp. 1909-1919, 2014.